Pergamon

PII: S0042-6989(96)00247-7

Vision Res., Vol. 37, No. 17, pp. 2455–2464, 1997

Printed in Great Britain
0042-6989/97 $17.00 + 0.00

A Subspace Reverse-correlation Technique for the Study of Visual Neurons

DARIO L. RINGACH,*‡ GUILLERMO SAPIRO,† ROBERT SHAPLEY*

Received 13 May 1996; in revised form 8 August 1996

A new discrete-time reverse-correlation scheme for the study of visual neurons is proposed. The visual stimulus is generated by drawing with uniform probability, at each refresh time, an image from a finite set S of orthonormal images. We show that if the neuron can be modeled as a spatio-temporal linear filter followed by a static nonlinearity, the cross-correlation between the input image sequence and the cell's spike train output gives the projection of the receptive field onto the subspace spanned by S. The technique has been applied to the analysis of simple cells in the primary visual cortex of cats and macaque monkeys. Experimental results are presented where S spans a subspace of spatially low-pass signals. Advantages of the proposed scheme over standard white-noise techniques include improved signal to noise ratios, increased spatial resolution, and the possibility to restrict the study to particular subspaces of interest. © 1997 Elsevier Science Ltd

Reverse-correlation Simple cell Macaque monkey Cat

INTRODUCTION

Some classes of neurons in the auditory and visual system can be modeled as a cascade of a linear filter, a static nonlinearity, and a spike generating mechanism, as illustrated in Fig. 1(A) (Weiss, 1966; Enroth-Cugell & Robson, 1966; de Boer & Kuyper, 1968; DeValois *et al.*, 1978; Movshon *et al.*, 1978; Andrews & Pollen, 1979; Segal & Outerbridge, 1982; Tolhurst & Dean, 1990). If the spike generating mechanism is assumed to be an inhomogeneous Poisson process, the model can be further simplified by having a continuous output variable represent the instantaneous rate of the Poisson process. This kind of model, shown in Fig. 1(B), is called a linear–nonlinear (LN) cascade system (Marmarelis & Marmarelis, 1978; Hunter & Korenberg, 1986; Korenberg & Hunter, 1990).

One possible way to identify a single-input single-output LN system consists of using a Gaussian white-noise (GWN) input and cross-correlating the output with the input signal. The result of this computation is proportional to the impulse response of the front-end linear filter (Bussgang, 1952; Price, 1958; Lee & Schetzen, 1965; Papoulis, 1984; Hunter & Korenberg, 1986; Eggermont *et al.*, 1983).§

The input to the visual system is a function of both time *and* space. The neural responses depend on the recent past of spatio-temporal luminance values in their receptive fields. The GWN method can be generalized to this case by covering the receptive field of the cell with an $M \times M$ square grid [as shown in Fig. 1(C)] and modulating the luminance value at each "pixel" by independent GWN processes. The input to the system is a vector $\mathbf{x}(t)$ of dimension M^2. The impulse response of the front-end linear filter of the LN cascade is now a function of both space and time, and can be identified by cross-correlating the scalar output $y(t)$ and the vector-valued input $\mathbf{x}(t)$.

In theory, there is nothing wrong with this approach. In practice, however, one is faced with the following dilemma. A large value for M (corresponding to a small pixel size) is desirable to achieve high spatial resolution. As M is increased, the spatio-temporal spectrum of the input becomes more uniformly distributed over the Fourier plane. This causes a decrease in the stimulus power that falls within the spatio-temporal integration area of the cell under study. If we assume a flat noise spectral power distribution, the signal-to-noise ratio will diminish as well. Thus, cells are expected to respond poorly to fine-grain spatio-temporal white-noise and long experiments are required to collect sufficient data for an acceptable reconstruction of the receptive field. Previous studies have been limited, therefore, to relatively coarse coverings of the receptive field (about 16×16 pixels) (Jones & Palmer, 1987; Reid & Shapley, 1992; Jacobson

*Center for Neural Science, New York University, 4 Washington Place, New York, NY 10003, U.S.A.
†Hewlett-Packard Laboratories, Palo Alto, CA 94304, U.S.A.
‡To whom all correspondence should be addressed [*Tel*: 01-212-998-7614; *Fax*: 01-212-995-4011; *Email*: dario@cns.nyu.edu].
§When the output of the system $y(t)$ is a spike train with arrival times $\{t_1, t_2, \ldots, t_n\}$ the empirical cross-correlation function is determined by $R_{xy}(\tau) = 1/T \sum_{k=1}^{T} x(t_k - \tau)$, where $x(t)$ is the input to the system. This can be interpreted as the expected value of the input τ sec before a spike occurred. The calculation is also referred to as the *reverse-correlation* between the input and the output spike train (de Boer & Kuyper, 1968; Eggermont *et al.*, 1983).

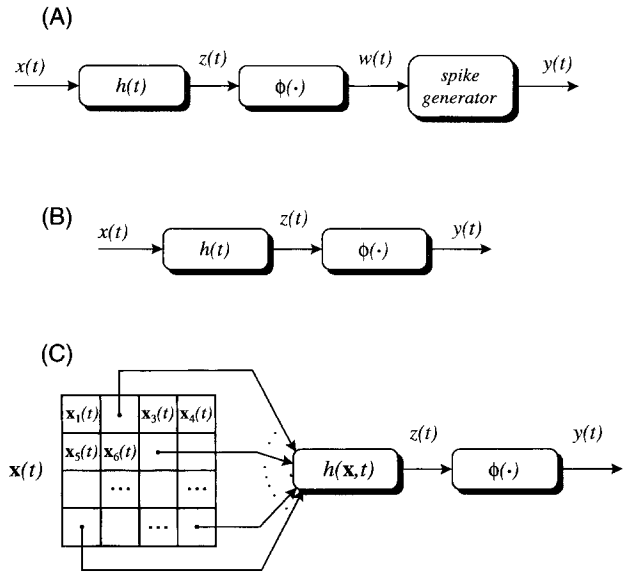

FIGURE 1. (A) A spiking linear–nonlinear (LN) model. The system is a cascade of a linear filter with impulse response $h(t)$, a static (memoryless) nonlinearity $\phi(\cdot)$, and a spike generator. (B) The standard linear–nonlinear (LN) cascade model. $h(t)$ is the transfer function of the front-end linear system and $\phi(\cdot)$ is a static (memoryless) nonlinearity. The output represents the instantaneous rate of firing of a Poisson process. (C) A spatio-temporal LN model. The input to the system is an array $M \times M$ of pixels ($M = 4$ in this example). The filter $h(\mathbf{x},t)$ is the transfer function of the spatio-temporal linear filter at the front-end, and $\phi(\cdot)$ is a static nonlinearity.

et al., 1993). The basic idea of our method is to use *a priori* knowledge about the spatial frequency tuning of the cell to generate a stimulus sequence whose spatial power distribution is shaped to match that of the neuron. In other words, we attempt to place the input signal power in those regions of the Fourier plane where the cell is most sensitive. This is achieved by restricting the stimulus ensemble to particular low-dimensional signal subspaces.

Similar ideas have been proposed in the time domain, such as the expansion of the first order kernel of a single-input single-output system into a small set of Laguerre orthogonal polynomials (Wiener, 1958; Alshebeili *et al.*, 1992; Marmarelis, 1993). These functions are a "natural" basis set to use because a small number of Laguerre polynomials are often enough to capture the shape of the impulse response in biological systems. In the temporal frequency domain, one can similarly reduce the dimension of the input space by using a stimulus signal that is the sum of a small number of sinusoids (Victor, 1979; Victor & Knight, 1979; Victor & Shapley, 1980).

In the space domain, researchers have implicitly been using related methods for a long time. For example, if it is known that the receptive field has radial symmetry, such as in retinal ganglion cells, it is useful to divide the visual space into a small number of concentric annuli, or perhaps just a central disk and a surrounding annulus, and take the luminance as a function of time in those areas as inputs to the system (Marmarelis & McCann, 1973; Marmarelis & Naka, 1974; Marmarelis & Marmarelis, 1978). In this way, the difference of the outer diameter of consecutive annuli and the total number of inputs to the system can be small. This means that high spatial resolution and a small input space can both be achieved if one carefully chooses the stimulus, using additional knowledge about the system.

Other examples in the space domain include methods developed to study simple cells in the primary visual cortex of cats and monkeys, which frequently show separable spatial receptive fields. One can take advantage of this knowledge to limit the study to a single dimension in space by using a one-dimensional white noise stimulus, such as random bars, oriented along the preferred direction of spatial integration (Citron & Emerson, 1983). Spatial frequency domain methods have also been suggested: Brodie *et al.* (1978) employed sinewave gratings of a fixed spatial frequency modulated in time by a sum of a small number of sinusoids to study the response of the Limulus retina.

In this work we propose a general framework that allows the experimenter to incorporate *a priori* knowledge about the spatial properties of a cell and to restrict the study of the system to particular dimensions (or spaces) of interest. This effectively reduces the dimension of the input space and yields higher signal to noise ratios. In the next section we present the main theoretical results. Next, computer simulations are presented to evaluate the performance of the algorithm for a number of different static nonlinearities and to compare the new scheme to the standard GWN technique. Finally, we report experimental data obtained using a subspace of band-limited signals.

THE METHOD

The theoretical results below yield a method to recover, under certain assumptions, the projection of the linear front-end filter in a LN model onto an arbitrary vector subspace.

The formulation is discrete both in time and space. Let us assume we have a system composed of a linear spatio-temporal filter followed by a static (memoryless) nonlinearity, as illustrated in Fig. 1(C). We cover the receptive field with a grid of $M \times M$ pixels. The input of the system is represented by a sequence of vectors $\mathbf{x}(n)$ of dimension M^2, where $n = \ldots, -2, -1, 0, 1, 2, \ldots$ denotes discrete time steps. The i-th component of $\mathbf{x}(n)$ represents the luminance of the i-th pixel in the image at time n. The receptive field of the cell can be characterized by vectors $\mathbf{h}(k)$ of dimension M^2, $k = 0, 1, 2, \ldots$ The i-th component of $\mathbf{h}(k)$ represents the impulse response of the i-th pixel in the $M \times M$ array at the k-th discrete time step. The output of the LN system is given by

$$y(n) = \phi\left(\sum_{k=0}^{\infty} \mathbf{h}(k) \cdot \mathbf{x}(n-k)\right), \qquad (1)$$

where $\mathbf{h}(k) \cdot \mathbf{x}(n-k)$ represents the standard inner product between $\mathbf{h}(k)$ and $\mathbf{x}(n-k)$, and $\phi(\cdot)$ is a static nonlinearity.

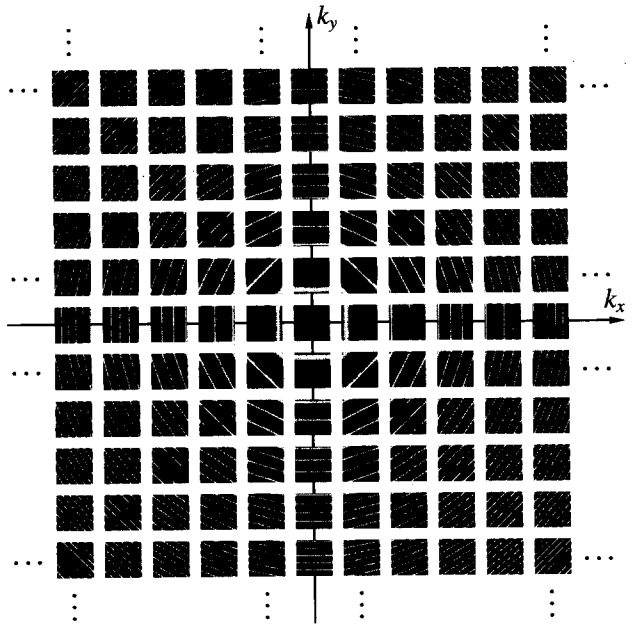

FIGURE 2. Elements of the two-dimensional Hartley basis function set on a 32×32 square grid. Only a few elements are shown. The origin is centered at (0,0). Negative values of the wave-numbers k_x and k_y should be interpreted according to the toroidal cyclicity $H(k_x, k_y) = H(k_x \pm M, k_y \pm M)$.

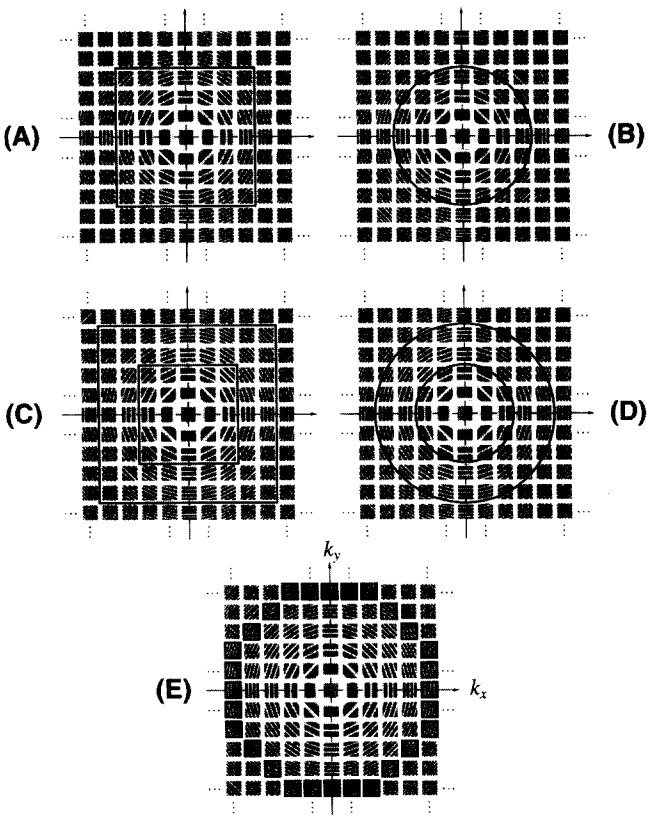

FIGURE 3. Subspaces obtained as subsets of the Hartley basis functions. (A, B) Subspaces of low-pass signals, S_Ω. The images whose centers lie *inside* the outlined regions define the basis set. (C, D) Subspaces of band-limited signals, $S_{(\omega,\Omega)}$. The images whose centers lie *between* the outlined regions define the basis set. (E) An "orientation" subspace defined by a digital circle of radius Ω. The images outlined by a square form the basis set for this subspace.

We first develop the theory when $\phi(\cdot)$ is a half rectifier; $\phi(x) = x$ if $x \geq 0$ and $\phi(x) = 0$ otherwise.

Let us pick an arbitrary finite set $S \equiv \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_Q\}$ of Q orthonormal images. Each image is represented as a vector of dimension M^2. Let us denote by $\hat{S} \equiv \{\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_2, \ldots, \mathbf{e}_Q, -\mathbf{e}_Q\}$ the set containing the elements in S *and* their negatives. The stimulus sequence is generated by drawing with uniform probability, at each discrete time step n, an element from $\hat{S}$. We denote by $R_{\mathbf{xy}}(j) \equiv E\{\mathbf{x}(n-j)y(n)\}$ the cross-correlation between the scalar output $y(n)$ and the vector input $\mathbf{x}(n)$. In the Appendix, we prove the following:

Theorem 1 *Given the above conditions,* $R_{\mathbf{xy}}(j) = \frac{1}{2Q} P_S \mathbf{h}(j)$, *for* $j = 0, 1, \ldots$; *where* $P_S \mathbf{h}(j) \equiv \Sigma_{e \in S} (\mathbf{h}(j) \cdot \mathbf{e})\, \mathbf{e}$ *denotes the projection of* $\mathbf{h}(j)$ *onto the subspace spanned by the vectors in S.*

A reduction of the dimensionality of the input space is achieved by selecting subspaces S for which the number of images $Q \ll M^2$. Note that M^2 is the dimension of the input space in the standard GWN approach on a $M \times M$ grid. We will see below that this reduction in the dimension of the input space can lead to higher signal to noise ratios and faster convergence rates than the standard GWN stimulation.

The result of Theorem 1 can be extended to other types of nonlinearities. Specifically, in the Appendix we also prove that:

Theorem 2 *A first order approximation to the cross-correlation between the scalar output* $y(n)$ *and the vector input* $\mathbf{x}(n)$ *is given by* $R_{\mathbf{xy}}(j) \approx \frac{C_\phi}{Q} P_S \mathbf{h}(j)$, *for* $j = 0, 1, \ldots$; *where* C_ϕ *is a scalar that depends on the nonlinearity.*

In the next section we show, by means of computer simulations, that this approximation is satisfactory for the kind of nonlinearities we can expect in biological systems. These results imply that one can obtain information about the front-end linear filter in a LN system without being concerned about the identification of the static nonlinearity. This is a clear advantage over traditional methods using drifting sinusoidal gratings.

To apply the method one must first define a subspace of signals. We do this by including in the set S a small number of elements taken from a complete orthonormal basis set of R^{M^2}. One intuitive basis to consider is the set of complete two-dimensional Hartley functions $H(k_x, k_y)$ on an $M \times M$ square grid (Bracewell, 1986):

$$H(k_x, k_y) = \text{cas}\left(\frac{2\pi(k_x l + k_y m)}{M}\right) \forall 0 \leq l,\ m \leq (M-1) \tag{2}$$

where $\text{cas}\ \alpha \equiv \sin\alpha + \cos\alpha$, and $0 \leq k_x,\ k_y \leq (M-1)$. Note that $H(k_x, k_y) \equiv H(k_x \pm M, k_y \pm M)$. Figure 2 shows some of the elements of the two-dimensional

Hartley basis set arranged with $H(0,0)$ centered at the origin. Each element in the set is a sinewave grating of a particular spatial frequency, orientation, and spatial phase. Note that the pair of images $H(k_x,k_y)$ and $H(M-k_x, M-k_y)$ have the same orientation and spatial frequency, but they are 90 deg out of spatial phase. The Hartley basis set is in some sense a "natural" choice for our purposes because we know simple cells in the visual cortex tend to be selective for orientation and spatial frequency, and their spectral structure is compact in such a space (Jones *et al.*, 1987; DeValois & DeValois, 1988). Another advantage of using Hartley basis functions is the existence of a fast Hartley transform (Bracewell, 1984, 1986) that provides a way to generate efficiently all the images required for the stimulus in a time proportional to $QM \log M$.

We now define some useful subspaces using the Hartley basis functions. Given a maximum (integer) wavenumber Ω, we define the space of low-pass signals, $S_\Omega \equiv \text{span}\ \{H(k_x,k_y) \mid \max(|k_x|, |k_y|) \leq \Omega\}$. This condition restricts the basis functions to a square centered at the origin in the (k_x, k_y)-plane with side $2\Omega + 1$ [see Fig. 3(A)]. A similar subspace can be obtained by using the elements inside a disk centered at the origin; $S_\Omega \equiv \text{span}\ \{H(k_x, k_y) \mid k_x^2 + k_y^2 \leq \Omega^2\}$ [Fig. 3(B)]. It is possible to quickly obtain a good estimate of the maximal spatial frequency that a cell responds to by measuring the spatial frequency tuning curve of the cell at its optimal orientation and temporal frequency. This estimate can then be used to select a low-pass space that will allow a complete reconstruction of the neuron's impulse response. A further reduction of the input space can be achieved if we also know that the cell does not respond below a critical spatial frequency, ω. In this case, analogous spaces of band-limited signals, $S_{(\omega,\Omega)}$, can be defined [see Fig. 3(C, D)].

The low-pass and band-limited signal sub-spaces allow one to easily control, or "shape", the spectral content of the input images. These spaces can be used to define effective stimuli for the cell. This is a clear advantage over the standard method of GWN on a square grid, whose spatial spectrum is *always* a sinc shaped function of spatial frequency automatically determined by the pixel size. In contrast, our method provides independent control of the pixel size and the spectral content of the images, allowing us to achieve high spatial resolution and high signal-to-noise ratios simultaneously. For instance, in the experiments reported below, the pixel size was in the order of 1.8 min of arc, while the number of Hartley images (the dimension of the input space) was less than 300. In the next section we present computer simulations that compare the performance of the standard GWN method to that of the subspace reverse-correlation method.

It is important to realize that some properties of the cell can be studied without requiring a full reconstruction of the RF. For example, the orientation tuning of a cell at a single spatial frequency can be investigated using a small subspace of Hartley images that lie on a (digital) circle centered at the origin, as depicted in Fig. 3(E). These images are sinewave gratings of very similar (but not identical) spatial frequencies at different orientations and spatial phases in quadrature.* This "orientation subspace" provides a framework to apply the reverse-correlation technique in the orientation domain, and is currently being used to study the neural circuitry underlying the orientation selectivity of cortical cells (Ringach *et al.*, 1997).

A consequence of the above results is that the measured projection of the impulse response $P_S\mathbf{h}(j)$ depends only on the subspace selected, *not* on the particular choice of the basis set. Therefore, the estimate of $P_S\mathbf{h}(j)$ should be the same with two basis sets, $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_Q\}$ and $S' = \{\mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_Q\}$, for which span S = span S'. Given a basis set S one can generate another basis set S' spanning the same subspace by a applying a "random rotation" (an orthogonal transformation) to S. This provides a simple (necessary) test to verify if a system is a LN cascade: the result of two reverse-correlation experiments using basis sets related by an orthogonal transformation should be equal. We are now applying this test to the responses of simple cortical cells.

COMPUTER SIMULATIONS

Two sets of computer simulations are presented. First, we examine the error introduced by the first order approximation in Theorem 2, when the nonlinearity ϕ is not a half-rectifier. The possible differences in the results obtained by using different basis sets spanning the same subspace is also studied. The second simulation compares the performance of our method to that of the standard GWN technique.

A typical space–time inseparable Gabor-like receptive field with four non-zero elements $\mathbf{h}(j)$, $j = 0, 1, 2, 3$, was first generated on a 32×32 grid. Each $\mathbf{h}(j)$ represents a time slice of the impulse response of the front-end linear filter to be used in a LN system. A subspace of low-pass signals S_Ω, with $\Omega = 4$, was used in these simulations. The true projection of the receptive field onto this subspace is illustrated in the column of Fig. 4(A). These are the spatial profiles we expect the algorithm to recover. Three different types of nonlinearities were used; a half-rectifier with zero threshold (for which Theorem 1 holds), a smooth sigmoidal nonlinearity, and a hard-step nonlinearity. In addition, for each nonlinearity simulations were done using 2D Hartley images and a new set obtained by a random rotation of the Hartley basis set. Figure 4(H) shows a segment of the stimulation sequence when the Hartley basis elements were used; Fig. 4(I) shows a segment of the stimulation sequence for a randomly rotated basis set.

The results shown in Fig. 4(B–G) were obtained after 5×10^4 discrete time steps. Figure 4(B) shows the estimated projection of the receptive field when a half-rectifier and the Hartley basis set were used. Figure 4(C)

*For digital circles of radius $\Omega \geq 7$ the variability in the spatial frequency of the gratings is less than 4%, and the angular resolution is higher than 12 deg. A value of $\Omega \geq 7$ is used in our experiments.

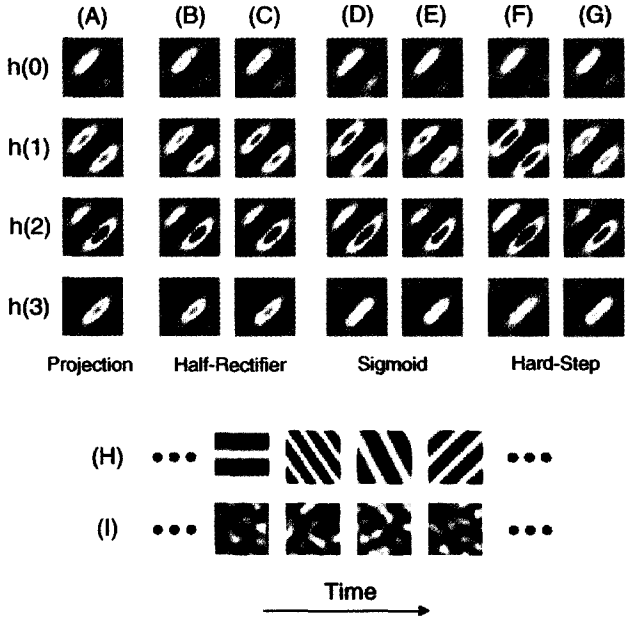

FIGURE 4. (A) The true projection of a simulated Gabor-like receptive field onto the subspace used in the simulations. (B, C) Estimated projections of the front-end linear filter in a LN system with half-rectifier nonlinearity. (B) Obtained with the Hartley basis set; (C) obtained with a randomly rotated basis set. (D, E) Projections estimated in a LN system with a sigmoidal nonlinearity. (D) Using Hartley basis set; (E) using a randomly rotated basis set. (F, G) System with a hard-step nonlinearity. (F) Projections estimated using a Hartley basis set; (E) projections obtained with a randomly rotated basis set. (H) Example of a segment of the stimulus sequence when Hartley basis elements are used. (I) Example of a segment of the stimulus sequence when a randomly rotated basis set is used.

TABLE 1. Normalized mean squared errors (MSE) in the estimation of the projection of the front-end linear filter $\mathbf{h}(j)$ for the six simulation conditions after 5×10^4 and 5×10^5 discrete time steps

	Steps = 50000					
	Half-rectifier		Sigmoid		Hard-step	
$\mathbf{h}(j)$	Hartley	Rotated	Hartley	Rotated	Hartley	Rotated
$\mathbf{h}(0)$	0.0195	0.0163	0.0363	0.0194	0.0965	0.0222
$\mathbf{h}(1)$	0.0045	0.0057	0.0254	0.0055	0.0603	0.0072
$\mathbf{h}(2)$	0.0035	0.0049	0.0230	0.0095	0.0468	0.0126
$\mathbf{h}(3)$	0.0160	0.0206	0.0497	0.0337	0.1231	0.0337
	Steps = 500000					
	Half-rectifier		Sigmoid		Hard-step	
$\mathbf{h}(j)$	Hartley	Rotated	Hartley	Rotated	Hartley	Rotated
$\mathbf{h}(0)$	0.0019	0.0016	0.0138	0.0025	0.0478	0.0020
$\mathbf{h}(1)$	0.0004	0.0004	0.0214	0.0017	0.0650	0.0005
$\mathbf{h}(2)$	0.0003	0.0006	0.0165	0.0028	0.0426	0.0004
$\mathbf{h}(3)$	0.0022	0.0026	0.0173	0.0020	0.0415	0.0021

illustrates the result for a half-rectifier and a randomly rotated basis set. Figure 4(D, E) depicts the outcome when a sigmoidal nonlinearity was used together with Hartley and randomly rotated basis elements, respectively. Similarly, Fig. 4(F, G) shows the result when a hard-step nonlinearity was selected using both sets of basis elements.

We note that the thresholds in the sigmoidal and hard-step nonlinearities were larger than zero. The distribution of $z(n)$ values (the input signal to the static nonlinearity) and the shape of the nonlinearities used in the simulations are shown in Fig. 5. The threshold for the sigmoid and hard-step nonlinearities was set to 50. In the case of the hard-step this meant that ≈85% of the time $z(n)$ was below threshold.

A normalized mean squared error (MSE) was obtained for each of the four time slices that compose the impulse response of the filter. This was done by first normalizing $\mathbf{h}(j)$ and its estimate to have a norm of one and taking the norm of their difference as a measure of the departure from the true projection. Table 1 shows the normalized

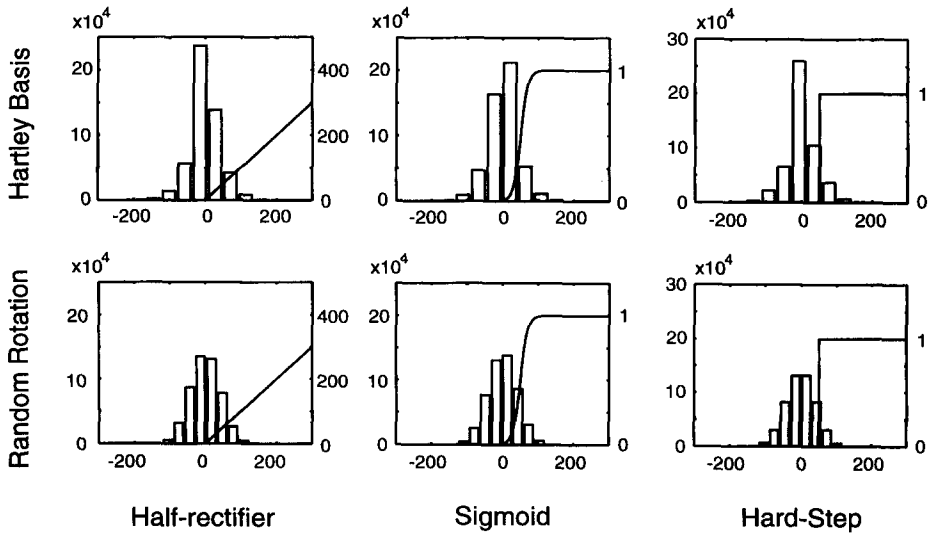

FIGURE 5. Histograms of $z(n)$ values and the shape of the three different nonlinearities used in the simulations of Fig. 4. Note that the sigmoidal and hard-step nonlinearities have thresholds larger than zero. The output value after the nonlinearity is given by the right ordinate.

MSE for each $\mathbf{h}(j)$ in each of the simulation cases after 5×10^4 and 5×10^5 discrete time steps. We see that the errors tend to be smaller using a randomly rotated Hartley basis set. For 5×10^5 steps the maximum normalized MSE is $\approx 5\%$ and occurs for the hard-step nonlinearity, but is typically less than 1%. The MSE decreases as the number of nonzero $\mathbf{h}(j)$ elements increases (see the Appendix). The simulations attempt to replicate the worst case experimental situation: each frame has a duration of 16.6 msec and cells show responses extending in time for about 60–120 msec. This implies a minimum of four nonzero $\mathbf{h}(j)$ elements. From these simulations we conclude that the method is expected to yield good estimates of the impulse response of the LN system, even in the presence of strong static nonlinearities (such as a hard-step).

We now evaluate how the standard GWN technique compares to the subspace reverse-correlation method. For these simulations a hard-step nonlinearity with threshold equal to 20 was used. In addition, independent Gaussian white noise with $\sigma = 40$ was added to $z(n)$ at each time step to simulate noise in the system. Figure 6(C) shows the true projection of the front-end linear filter into the subspace spanned by the Hartley set of images [this is the same as in Fig. 4(A)].

In the GWN case each pixel was modulated by independent Gaussian white noise. The variance of the input signal in the GWN and subspace reverse-correlation cases were equalized. The cross-correlation between the input image sequence and the output gives the first order kernel of the system (Lee & Schetzen, 1965). The result from this calculation is then projected into the subspace spanned by the images in the Hartley set.* This operation is essentially smoothing the first order kernel obtained with GWN by projecting it into a space of low-pass signals. The estimated projected profiles obtained after $N = 1000$, 2000, 4000, 8000, 32000 and 64000 iterations are shown in Fig. 6(A).† It can be seen that results obtained in the GWN method are much noisier than the ones obtained with the reverse-correlation technique, which are depicted in Fig. 6(B). These simulations confirm the claim that the signal to noise is higher, and the rate of convergence faster, in the subspace reverse-correlation scheme than in the standard GWN method.

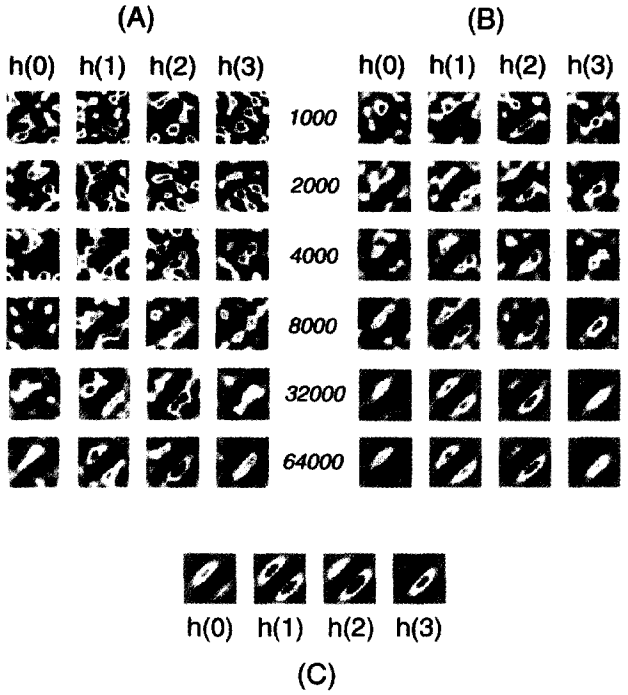

FIGURE 6. Comparison of the performance of the subspace reverse-correlation method and the standard GWN technique. Results after $N = 1000$, 2000, 4000, 8000, 32000, and 64000 discrete time steps are shown in different rows. The columns represent the four estimates of $\mathbf{h}(j)$, for $j = 0$, 1, 2, 3. (A) Projections estimated using the GWN technique, (B) projections estimated using the subspace reverse-correlation technique. The true projection of the simulated system is shown in (C).

EXPERIMENTAL RESULTS

In this section we present experimental results based on the low-pass space of signals, S_Ω, depicted in Fig. 3(A). Extracellular recordings from cells in the primary visual cortex of anesthetized cats and monkeys were performed with methods described elsewhere (Reid *et al.*, 1991; Hawken *et al.*, 1996). The maximum spatial frequency, Ω, at which each cell responded was estimated from the spatial frequency tuning curve of the cell at the optimal orientation and temporal frequency using drifting sinewave gratings. A Silicon Graphics Indigo Elan R4000 was used to generate the elements in the set S_Ω using a fast, two-dimensional Hartley transform. The presentation mode consisted of 30 sec trials, in which the computer displayed a stimulus sequence at a frame refresh frequency of 60 Hz. Either 20 or 30 trials were run on each cell; the total experimental duration was 10 or 15 min. Spikes were discriminated using a dual window discriminator (model DDIS-I, Bak Electronics). A CED-1401 Plus (Cambridge Electronic Design Ltd., Cambridge, U.K.) was used for data acquisition and to time-stamp the spike events with 1 msec precision. Data were stored on disk for off-line analysis.

A reconstruction of the receptive field (RF) was computed by the subspace reverse–correlation technique described above. Each of the columns in Fig. 7 shows the reconstructed profiles of four cortical V1 neurons. The illustration shows time-slices of the spatial profiles of the RFs. Each spatio-temporal RF was independently scaled and translated so that its maximum attainable value is mapped to +1 (red = excitation) and the minimum value mapped to −1 (blue = inhibition). This achieves the maximum dynamic range possible for the pseudo-color map. The scale bar in the last frame of each sequence represents 1 deg of visual angle. In each case, measure-

*If this is not done the results of the GWN calculation are even noisier than the ones presented here.

†To be able to compare these simulations with the experimental results in the next section, note that a 15 min experiment using a frame refresh rate of 60 Hz represents a total of $N = 56000$ frames.

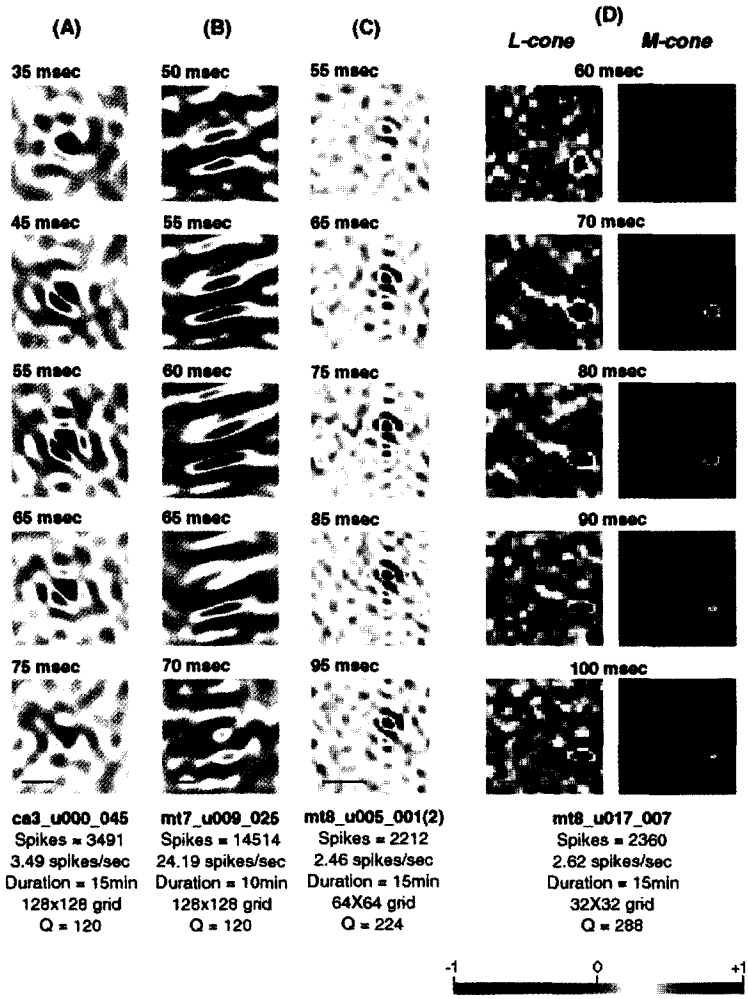

FIGURE 7. Reconstructed spatio-temporal profiles of four cortical V1 neurons using the subspace reverse-correlation method. (A) Simple cell in cat area 17. (B) Simple cell in monkey V1. (C) Unclassified cell from superficial layers of macaque area V1. (D) Simple chromatic cell from area V1 of the monkey. The L-cone and M-cone columns show the reconstructions of the experiment when the stimulus was modulated in the M-cone and L-cone directions by the method of silent substitution.

ment parameters are indicated at the bottom of the graph: the total number of spikes on which the RF reconstruction is based, the mean firing rate of the cell, the total duration of the experiment, the grid size used, and the dimension of the input space (the number of images in the subspace) Q. We find that cells tend to respond with mean firing rates in the range of 1–30 spikes/sec. As expected from the nature of the stimulus, directionally selective cells are usually less responsive than nondirectionally selective cells.

Figure 7(A) shows responses from a simple cell in cat area 17. The cell's estimated spatio-temporal impulse response has two main sub-regions, one inhibitory and one excitatory. However, two additional excitatory regions are clearly seen at $t = 55$–65 msec. Thus, the RF of this cell has a two-dimensional structure. Sun and Bonds (1994) have also reported cells with two-dimensional profiles. In their population, 44% of the cells showed a multi-peaked response.

Figure 7(B) depicts the analysis of a simple cell in macaque V1. The cell exhibits a Gabor-like spatial profile (Marcelja, 1980). In addition, we see that the main inhibitory (central) region develops from two "hot spots" clearly seen at $t = 50$ msec. The same "hot spots" are seen when the response is disappearing at $t = 70$ msec. We conjecture that these "hot spots" may represent direct inputs from the LGN (Reid & Alonso, 1995).

Figure 7(C) shows results from a superficial layer cell from macaque primary visual cortex. We typically find that these cells are very difficult to stimulate with drifting gratings, while they do seem to respond to small colored spots on dark backgrounds. Even though we expect these cells to exhibit strong nonlinearities and to depart from a simple LN system, we feel it is instructive to present the result of the method in this case. The cell seems to have a center-surround organization with a central inhibitory region and a strong excitatory surround. Under close inspection of the response at $t = 85$–95 msec, the surround seems to be composed of separate (punctate) excitatory regions. We emphasize that care should be take in interpreting the result when the system is known to depart from a simple LN configuration. The outcome

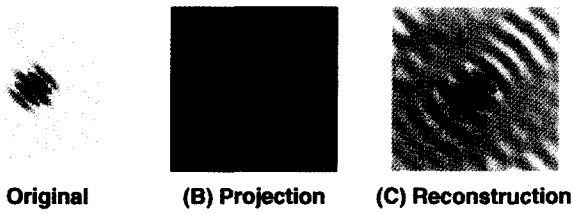

FIGURE 8. (A) A computer simulated spatial profile of a Gabor-like receptive field. (B) The power distribution of the receptive field in the Hartley domain. The highlighted circular region in light blue represents the subspace onto which the projection is taken. (C) The spatial reconstruction of the projected receptive field shows additional ripples not present in the original profile.

of the experiment may only serve as a guide on which to base further experimental study of the cell's properties.

Finally, Fig. 7(D) shows a simple cell that responds preferentially to chromatic stimuli in monkey V1. A detailed map of the spatial distribution of the cone inputs to such a cell can be obtained by applying our method with Hartley basis functions modulated in color so that they can only be seen by a single type of cone photoreceptor (Reid & Shapley, 1992; Estevez & Spekreijse, 1982). This technique is also known as the silent substitution method or as color modulation in the cone directions. The left column in Fig. 7(D) shows the L-cone receptive field and the right column the M-cone input. There is a clear spatial opponency of the two mechanisms (M+ and L−). Interestingly, the L-cone response seems to show some orientation selectivity at $t = 90$ msec, while the M-cone receptive field is more isotropic with a focal excitatory center.

The profiles shown here represent *projections* onto a subspace of signals, and caution should be exercised if one is to interpret them directly in the space domain. Figure 8 illustrates the problem. Here, Fig. 8(A) represents a simulated spatial profile of a Gabor-like RF. The middle image, Fig. 8(B), shows the power distribution of the RF on the Hartley domain (with the origin at the center), and the highlighted blue circular region indicates the subspace onto which we project the RF. Clearly, only a fraction of the filter's energy lies within the selected subspace. The spatial reconstruction of the projected RF onto this subspace is shown in Fig. 8(C). The numerous ripples observed in the reconstruction are a consequence of the projection operation and are not real features of the original RF. The appearance of the ripples can be understood by noticing that using a low-pass subspace is similar to obtaining a truncated Fourier series approximation of the RF. If high spatial frequency components are present in the RF, ripples will appear in any truncated approximation (Gibbs' phenomenon). Note that the information available in Fig. 8(C), however, should allow one to predict the response of the system to any stimulus in which each frame belongs to the subspace.

It is possible to check if a selected subspace is likely to contain most of the front-end filter's energy by inspecting

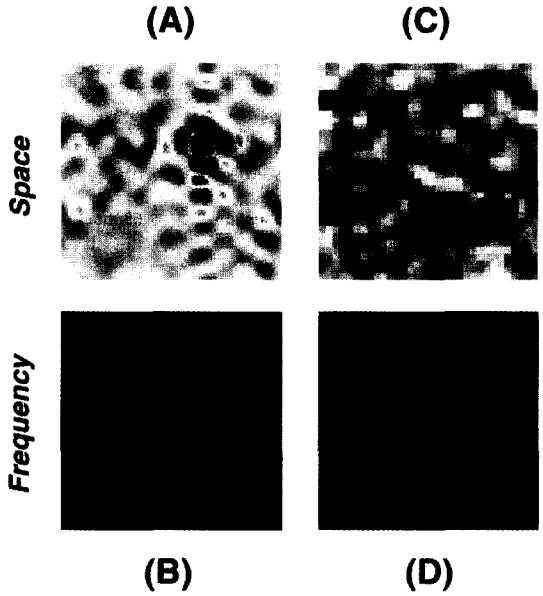

FIGURE 9. A test to verify if a subspace is likely to contain most of the front-end filter's energy after the experiment. (A) The spatial profile of a cell shown in Fig. 4(C). In (B) the power distribution of the receptive field in the Hartley domain is shown. The dashed square indicates the subspace used in the experiment. In this the cell was responding to elements very near the boundary defining the subspace. (C) The spatial profile of the L-cone input of the cell shown in Fig. 4(D). In (D) we can see that the cell was responding weakly to the basis elements near the boundary defining the subspace.

the power distribution of the RF in the Hartley domain after the experiment is done. Figure 9(A) shows a slice of the RF of the cell we presented in Fig. 7(C), and the power distribution in the Hartley domain is depicted in Fig. 9(B). The dashed square indicates the limits of the subspace used for the experiment. Clearly, the power distribution reaches the boundaries of the subspace. Thus, this subspace may be "too small" to allow a full identification of the RF. Many of the secondary ripples seen in Fig. 9(A) are probably not a real feature of the RF. Figure 9(C) shows a slice of the RF of L-cone isolating response presented in Fig. 7(D). It can be seen from the corresponding power distribution, illustrated in Fig. 9(D), that the responses were very small near the boundaries defining the subspace. In this case, one can be confident that most of the energy of the front-end linear filter lies within the selected subspace and, therefore, the reconstruction of the receptive field profile in space reflects real features of the neuron's spatiotemporal impulse response.

CONCLUSIONS

A discrete reverse–correlation technique based on a subspace approach was proposed. This new method allows one to incorporate *a priori* information in the design of the stimulus set to reduce the dimensionality of the input space. Similarly, one can restrict the study to particular spaces of interest, such as the "orientation subspace" described above. The method was applied to

study the receptive field organization of cells in the primary visual cortex of cats and monkeys. High resolution spatial and temporal maps of the RFs were obtained in a reasonable amount of experimental time. A necessary test for LN systems was also suggested. We are currently applying the technique to study the cortical circuitry involved in the generation of orientation tuning of simple cells in monkey V1 (Ringach *et al.*, 1997). The possibility of extending the concept of using a restricted (effective) stimulus set to the computation of "high-order" interactions between the elements in the set is a topic for further research.

REFERENCES

Alshebeili, S. A., Mertzios, B. G. & Venetsanopoulous, A. N. (1992). Efficient implementation of nonlinear Volterra systems via Laguerre orthogonal functions. In Vanderwalle, J., Boite, R., Moonen M. & Oosterlinck A. (Eds), *Signal processing VI: Theories and applications.* Amsterdam: Elsevier.

Andrews, B. W. & Pollen, D. A. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology, 287*, 163–176.

Bracewell, R. N. (1984). The fast Hartley transform. *Proceedings of the IEEE, 728*, 1010–1018.

Bracewell, R. N. (1986). *The Hartley transform*, NY: Oxford University Press.

Brodie, S. E., Knight, B. W. & Ratliff, F. (1978). The response of the *Limulus* retina to moving stimuli: a prediction by Fourier synthesis. *Journal of General Physiology, 72*, 129–166.

Bussgang, J. J. (1952). Crosscorrelation functions of amplitude distorted Gaussian signals. Technical Report 216, MIT Res. Lab. of Electronics.

Citron, M. C. & Emerson, R. C. (1983). White noise analysis of cortical directional selectivity in cat. *Brain Research, 279*, 271–277.

de Boer, E. & Kuyper, P. (1968). Triggered correlation. *IEEE Transactions on Biomedical Engineering, 153*, 169–179.

DeValois, R. L. & DeValois, K. K. (1988). *Spatial vision.* New York: Oxford University Press.

DeValois, R. L., Thorell, L. G. & Albrecht, D. G. (1978). Cortical cells: bar and edge detectors or spatial-frequency analyzers? In Cool, S. J. & Smith, E. L. (Eds), *Frontiers in visual science* (pp. 544–556). New York: Springer.

Eggermont, J. J., Johannesma, P. I. M. & Aertsen, A. M. H. J. (1983). Reverse-correlation methods in auditory research. *Quarterly Review of Biophysics, 16*, 341–414.

Enroth-Cugell, C. & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology, 187*, 517–522.

Estevez, O. & Spekreijse, H. (1982). The "silent substitution" method in visual research. *Vision Research, 226*, 681–691.

Hawken, M. J., Shapley, R. & Grosof, D. H. (1996). Temporal frequency selectivity in macaque visual cortex. *Visual Neuroscience, 133*, 477–492.

Hunter, I. W. & Korenberg, M. J. (1986). The identification of nonlinear biological systems: Wiener and Hamerstein cascade models. *Biological Cybernetics, 55*, 135–144.

Jacobson, L. D., Gaska, J. P., Chen, H. & Pollen, D. A. (1993). Structural testing of multi-input linear–nonlinear cascade models for cells in macaque striate cortex. *Vision Research, 33*, 609–626.

Jones, J. P. & Palmer, L. A. (1987). The two dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*, 1187–1211.

Jones, J. P., Stepnoski, A. & Palmer, L. A. (1987). The two-dimensional spectral structure of simple cells in cat striate cortex. *Journal of Neurophysiology, 58*, 1212–1232.

Korenberg, M. J. & Hunter, I. W. (1990). The identification of nonlinear biological systems: Wiener kernel approaches. *Annals of Biomedical Engineering, 18*, 629–654.

Lee, Y. W. & Schetzen, M. (1965). Measurement of the Wiener kernels of a nonlinear system by cross-correlation. *International Journal of Control, 2*, 237–254.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America, 70*, 1297–1300.

Marmarelis, P. N. & Marmarelis, V. Z. (1978). *Analysis of physiological systems: the white noise approach.* New York: Plenum Press.

Marmarelis, P. Z. & McCann, G. D. (1973). Development and application of white-noise modeling techniques for studies of insect visual nervous system. *Kybernetic, 12*, 74–89.

Marmarelis, P. Z. & Naka, K. I. (1974). Multi-input statistical testing on physiological systems. In *Proceedings of the 27th Annual Conference on Engineering in Medicine and Biology*, Philadelphia.

Marmarelis, V. Z. (1993). Identification of nonlinear biological systems using Laguerre expansions of kernels. *Annals of Biomedical Engineering, 21*, 573–589.

Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology, 283*, 101–120.

Papoulis, A. (1984). *Probability, random variables, and stochastic processes*, 2nd edition). McGraw-Hill.

Price, R. (1958). A useful theorem for nonlinear devices having Gaussian inputs, *IRE, PGIT*, **4**.

Reid, R. C. & Alonso, J. M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature, 378*, 281–284.

Reid, R. C. & Shapley, R. M. (1992). Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature, 365*, 716–718.

Reid, R. C., Soodak, R. E. & Shapley, R. (1991). Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *Journal of Neurophysiology, 662*, 505–529.

Ringach, D. L., Hawken, J. J. & Shapley, R. (1997). The dynamics of orientation tuning in macaque primary visual cortex. *Nature, 387*, 281–284.

Segal, B. N. & Outerbridge, J. S. (1982). Vestibular (semicircular canal) primary neurons in bullfrog: nonlinearity of individual and population response to rotation. *Journal of Neurophysiology, 47*, 545–562.

Sun, M., & Bonds, A. B. (1994). Two-dimensional receptive-field organization in striate cortical neurons of the cat. *Visual Neuroscience*, 11, 703–720.

Tolhurst, D. J. & Dean, A. F. (1990). The effects of contrast on the linearity of the spatial summation of simple cells in the cat's striate cortex. *Experimental Brain Research, 79*, 582–588.

Victor, J. D. (1979). Nonlinear system analysis: comparison of white noise and sum of sinusoids in a biological system. *Proceedings of the National Academy of Sciences USA, 76*, 996–998.

Victor, J. D. & Knight, B. W. (1979). Nonlinear analysis with an arbitrary stimulus ensemble, *Q. Appl. Math., 37*, 113-136.

Victor, J. D. & Shapley, R. (1980). A method of nonlinear analysis in the frequency domain. *Biophysical Journal, 29*, 671–689.

Weiss, T. (1966). A model of the peripheral auditory system. *Kybernetic, 3*, 153–175.

Wiener, N. (1958). *Nonlinear problems in random theory.* New York: John Wiley.

Acknowledgements—We thank Jonathan Victor and Haim Sompolinsky for their detailed comments and criticisms on an earlier version of this manuscript. Special thanks to Mike Hawken and Ferenc Mechler for valuable discussions and help in data collection. Supported by a Sloan Foundation Postdoctoral Fellowship to DLR and by NIH-EY01472 to RS.

APPENDIX

Proof of Theorem 1: Consider the system in Fig. 1(C). $z(n)$ and $\mathbf{x}(n)$ are linearly related by $z(n) = \sum_{k=0}^{\infty} \mathbf{h}(k) \cdot \mathbf{x}(n-k)$. Thus, we have that

$$R_{\mathbf{xz}}(j) = E\{\mathbf{x}(n-j)z(n)\} \tag{A1}$$

$$= E\left\{\mathbf{x}(n-j)\sum_{k=0}^{\infty}\mathbf{h}(k)\cdot\mathbf{x}(n-k)\right\} \tag{A2}$$

$$= E\{\mathbf{x}(n-j)(\mathbf{h}(j)\cdot\mathbf{x}(n-j))\}$$

$$+ \sum_{k=0,k\neq j}^{\infty} E\{\mathbf{x}(n-j)(\mathbf{h}(k)\cdot\mathbf{x}(n-k))\} \tag{A3}$$

$$= \frac{1}{Q}P_s\mathbf{h}(j) + \sum_{k=0,k\neq j}^{\infty} E\{\mathbf{x}(n-j)E\{\mathbf{h}(k)\cdot\mathbf{x}(n-k)\,|\,\mathbf{x}(n-j)\}\} \tag{A4}$$

$$= \frac{1}{Q}P_s\mathbf{h}(j) + \sum_{k=0,k\neq j}^{\infty} E\{\mathbf{x}(n-j)\,E\underbrace{\{\mathbf{h}(k)\cdot\mathbf{x}(n-k)\}}_{=0}\} \tag{A5}$$

$$= \frac{1}{Q}P_s\mathbf{h}(j). \tag{A6}$$

In equation (A4) we have used the fact that for any two random variables a and b we have $E\{ab\} = E\{aE\{b\,|\,a\}\}$. In equation (A5) the second expectation term is zero because the contributions of $+\mathbf{e}_i$ and $-\mathbf{e}_i$ cancel each other out.

We now have to compute $R_{\mathbf{xy}}(j) = E\{\mathbf{x}(n-j)y(n)\}$. Due to the symmetry of the input and the fact that $z(n)$ is linearly related to $\mathbf{x}(n-j)$ we have that $E\{\mathbf{x}(n-j)z(n)\,|\,z(n)>0\} = E\{\mathbf{x}(n-j)z(n)\,|\,z(n)<0\}$. In addition $P\{z(n)<0\} = P\{z(n)>0\} = {}^1/_2$. This means that $E\{\mathbf{x}(n-j)z(n)\} = E\{\mathbf{x}(n-j)z(n)\,|\,z(n)>0\}$

$$R_{\mathbf{xy}}(j) = E\{\mathbf{x}(n-j)y(n)\} \tag{A7}$$

$$= \frac{1}{2}E\{\mathbf{x}(n-j)z(n)|z(n)>0\} \tag{A8}$$

$$= \frac{1}{2}E\{\mathbf{x}(n-j)z(n)\} \tag{A9}$$

$$= \frac{1}{2Q}P_s\mathbf{h}(j) \tag{A10}$$

which completes the proof. □

We now obtain the first order approximation in the case of other types of nonlinearities.

Proof of Theorem 2: As before, let us denote $z(n) = \sum_{k=0}^{\infty}\mathbf{h}(k)\cdot\mathbf{x}(n-k) \equiv Z_j(n) + \mathbf{x}(j)\cdot\mathbf{h}(n-j)$, where $Z_j(n)$ contains all the other elements in the sum except the one for $k=j$. We have:

$$R_{\mathbf{xy}}(j) = E\{\mathbf{x}(n-j)y(n)\} \tag{A11}$$

$$= E\{\mathbf{x}(n-j)\,\phi(Z_j(n) + \mathbf{h}(j)\cdot\mathbf{x}(n-j))\} \tag{A12}$$

$$= E\{\mathbf{x}(n-j)\sum_{k=0}^{\infty}\frac{\phi^{(k)}(Z_j(n))}{k!}(\mathbf{h}(j)\cdot\mathbf{x}(n-j))^k\} \tag{A13}$$

$$= \sum_{k=0}^{\infty}\frac{E\{\phi^{(k)}(Z_j(n))\}}{k!}E\{(\mathbf{h}(j)\cdot\mathbf{x}(n-j))^k\mathbf{x}(n-j)\} \tag{A14}$$

Here, $\phi^{(k)}$ denotes the k-th derivative of ϕ. In equation (A13) we have written the Taylor expansion of $\phi(Z_j(n) + \mathbf{h}(j)\cdot\mathbf{x}(n-j))$ around $Z_j(n)$. We assume this expansion exists. Note that $E\{(\mathbf{h}(j)\cdot\mathbf{x}(n-j))^k\,\mathbf{x}(n-j)\} = 0$ for all even k, and that for $k=1$ we have that $E\{(\mathbf{h}(j)\cdot\mathbf{x}(n-j))\,\mathbf{x}(n-j)\} = \frac{1}{Q}P_s\mathbf{h}(j)$. Thus, the error term in the first order approximation is at least of third order and we have that,

$$R_{\mathbf{xy}}(j) = \frac{E\{\phi'(Z_j(n))\}}{Q}P_s\mathbf{h}(j) + \mathcal{O}\left(E\{(\mathbf{h}(j)\cdot\mathbf{x}(n-j))^3\mathbf{h}(j)\}\right) \tag{A15}$$

$$\approx \frac{E\{\phi'(Z_j(n))\}}{Q}P_s\mathbf{h}(j). \tag{A16}$$

The approximation improves as the number of nonzero $\mathbf{h}(j)$ increases. In this case, the contribution of $\mathbf{h}(j)\cdot\mathbf{x}(n-j)$ to $z(n)$ is small and $E\{\phi'(Z_j(n))\} \approx E\{\phi'(z(n))\}$. We obtain

$$R_{\mathbf{xy}}(j) \approx \frac{C_\phi}{Q}P_s\mathbf{h}(j), \tag{A17}$$

where the constant $C_\phi \equiv E\{\phi'(z(n))\}$. □