Discrete analysis of spatial-sensitivity models

Kenneth R. K. Nielsen and Brian A. Wandell

Department of Psychology, Stanford University, Stanford, California 94305

Received July 27, 1987; accepted November 30, 1987

The visual representation of spatial patterns begins with a series of linear transformations: the stimulus is blurred by the optics, spatially sampled by the photoreceptor array, spatially pooled by the ganglion-cell receptive fields, and so forth. Models of human spatial-pattern vision commonly summarize the initial transformations by a single linear transformation that maps the stimulus into an array of sensor responses. Some components of the initial linear transformations (e.g., lens blurring, photoreceptor sampling) have been estimated empirically; others have not. A computable model must include some assumptions concerning the unknown components of the initial linear encoding. Even a modest sketch of the initial visual encoding requires the specification of a large number of sensors, making the calculations required for performance predictions quite large. We describe procedures for reducing the computational burden of current models of spatial vision that ensure that the simplifications are consistent with the predictions of the complete model. We also describe a method for using pattern-sensitivity measurements to estimate the initial linear transformation. The method is based on the assumption that detection performance is monotonic with the vector length of the sensor responses. We show how contrast-threshold data can be used to estimate the linear transformation needed to characterize threshold performance.

1. INTRODUCTION

It is approximately 30 years since Schade and de Lange[1–4] suggested that human threshold sensitivity can be modeled by physiological sensors whose performance can be likened to linear filters. Some 20 years ago, on the basis of detection and discrimination measurements, Campbell and Robson[5] suggested that the sensor array must consist of narrowly tuned filters with different center frequencies. The assumption that a large array of filters with different spatial sensitivities is required to model human vision leads naturally to the descriptive phrase multiple-channels model. A model based on the additional assumption that the channels are narrowly tuned with respect to the spatial-frequency dimension is called a multiple-channels spatial-frequency model, a more restrictive term.

Since the publication of the seminal paper of Campbell and Robson, a large number of empirical and theoretical studies have contributed to the specification of a multiple-channels spatial-frequency model of spatial-pattern vision.[6] Construction of a complete model requires the estimation of parameters such as the filter center frequencies, bandwidths, orientations, and spatial positions. In the past few years Wilson and co-workers,[7,8] Watson,[9] Klein and Levi,[10] and others have designed large, computationally intensive models whose goal is to incorporate all the fundamental information derived from the psychophysical studies. Some of the sensors derived from psychophysical research, particularly those of Wilson and co-workers, have been adopted as building blocks for designs of complex image-understanding systems.[11]

Psychophysical models of pattern vision can be divided into two steps. First, the stimulus contrasts are transformed into the responses of an initial set of sensors, much as low-contrast stimuli on the retinal surface are transformed into a set of nearly linear cortical-cell responses. This step is a part of all the implementations mentioned above.

The initial linear transformation is followed by a second step that takes sensor responses as input and yields the subject's response as output. The second step is the decision stage. Several authors have studied models with non-optimal decision stages as peak detectors,[12–14] whereas recently many authors have studied models incorporating more-nearly optimal decision stages.[15–18]

Some properties of the complete model are due to the initial linear transformation, some are due to the decision stage, and some are due to the interaction. In Section 2 we describe properties of the linear transformation that are inherited by all decision stages. The results in Section 2 are applicable to all models in which an initial linear transformation is assumed. In Section 3 we describe an alternative approach to the construction of linear models of spatial vision. We have sought methods of building models so that the relationship between model parameters and experimental measurements is direct. We show that if the decision stage is based on the vector length of the sensor responses, then it is possible to estimate directly the key parameters of the initial linear transformation from experimental measurements.

In this paper our development is restricted to models of sensitivity to low-contrast stimuli. We comment briefly on the related topics of masking and adaptation in the conclusion.

2. ANALYSIS OF THE INITIAL LINEAR TRANSFORMATION

Although the decision stages in various spatial pattern models are often quite different, certain properties of the initial linear transformation are inherited by all decision stages. We describe three tools for the analysis of the effects of selecting an initial linear transformation.

First, we show how to reduce the computational complexity of the initial linear transformation when the model is applied to a restricted stimulus domain. Second, we show

0740-3232/88/050743-13$02.00

how to rewrite the initial linear transformation with respect to an alternative basis set. Such a change of basis may serve conceptually to clarify the action of the linear transformation, or it may serve to reduce the computational burden. An example of a change of basis is to shift the stimulus representation from a point-by-point representation to a Fourier representation or a Gabor representation. Third, we describe how to compute the null space of the model from the initial matrix. The null space is the set of pattern stimuli that cause precisely the same sensor responses as a stimulus with zero contrast. Knowledge of the null space permits us to determine the equivalence classes of visual stimuli with respect to the model, that is, stimuli that are physically different but are encoded identically by the model. We relate the null-space analysis that we undertake here to the current analysis of spatial aliasing in uniform and nonuniform sampling arrays.

Definitions

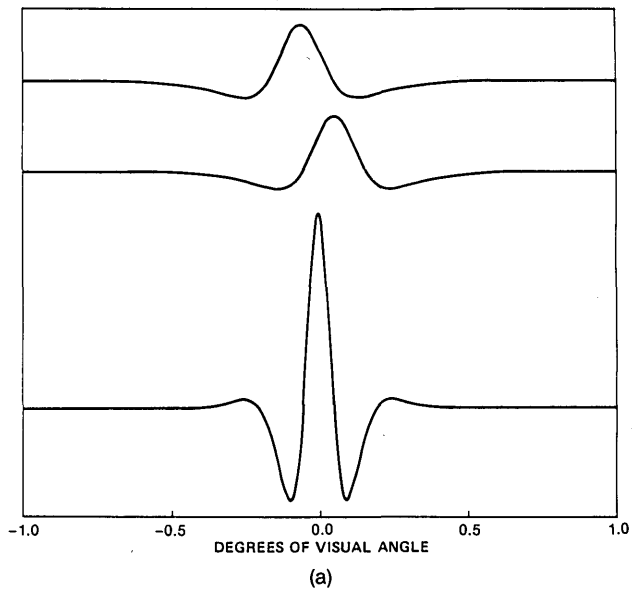
To simplify our presentation, we have chosen to use discrete matrix notation. By representing the stimuli and responses with increasing resolution, we can approach continuous calculations as closely as we desire. The symbols used in this paper are presented with their definitions in Table 1.

Two representations of image data are widely used. On occasion images are described as a matrix with horizontal and vertical dimensions N_x and N_y, respectively. Here we generally represent the stimulus input and sensor sensitivities as ($N_xN_y = N$)-dimensional vectors. We reserve the symbol $\hat{i}$ to denote any image vector of length 1, that is, $||\hat{i}|| = 1.0$. A vector of contrast α, where α is a scalar, is represented as $\alpha\hat{i}$. The vector length of the stimulus is the discrete counterpart to stimulus contrast energy defined by Watson *et al.*[19]

In describing the two-dimensional (2D) image functions as vectors, we use the following mapping from a 2D image: the first N_x entries of the vector are the first row of the image, the next N_x entries are the second row, and so forth for the $N_xN_y = N$ entries.

When the image data are represented as a vector, it is convenient to represent the spatial sensor sensitivities as vectors as well. The response of a single spatial sensor to an input image is the inner product (dot product) of the vector representing the sensor's 2D receptive field with the vector representing the image data. We require that the sensor receptive fields be represented at the same level of resolution as the image. At each sample point in the sensor representation we compute the product of the sensor sensitivity and the image contrast and then sum the results to obtain the inner product.

To represent the response of an array of sensors, we construct a matrix whose rows are the vector representations of spatial receptive fields of the individual sensors [Fig. 1(b)]. We call this matrix the sensor matrix and denote the matrix by the symbol **S**. If there are P sensors, then this matrix is $P \times N$. The matrix **S** maps the unit stimulus vector $\hat{i}$ into the sensor responses $\mathbf{r} = \mathbf{S}\hat{i}$. Given a stimulus with contrast α, the P-dimensional vector of sensor responses is the matrix product $\alpha\mathbf{r} = \mathbf{S}\alpha\hat{i}$. All the information about the initial linear stages of a spatial-vision model is described by the entries of the sensor matrix. Equivalently, the selection of P receptive-field sensitivities and a sampling resolution $N_x \times N_y$ define the $P \times N$ sensor matrix.



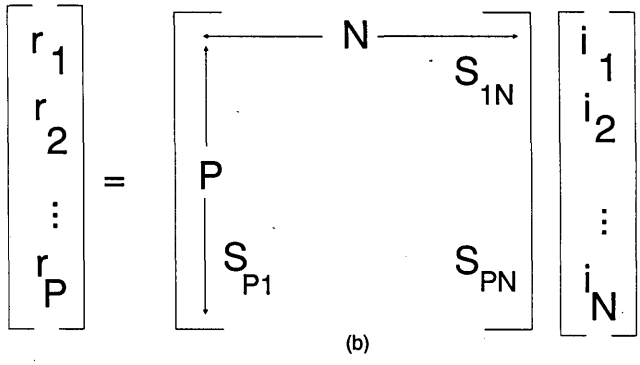


Fig. 1. (a) Linear sensor functions of 3 of the 18 Wilson–Gelb sensors. (b) Expanded representation of the matrix product $\mathbf{r} = \mathbf{S}\hat{i}$ describing the initial linear transformation. Each row of **S** is a sensor spatial sensitivity.

Table 1. Symbols and Their Definitions

Symbol	Definition
N	Number of degrees of freedom in the input image
P	Number of spatial sensors
α	Contrast of a pattern stimulus
$\hat{i}$	Unit contrast spatial pattern represented as an N-dimensional vector
S	$P \times N$ matrix whose rows are the sensor spatial sensitivities
r	P-dimensional vector of responses to unit input, $r = \mathbf{S}\hat{i}$
B	Matrix whose columns are a complete basis for all possible stimuli
R	Matrix whose columns are a reduced stimulus basis set
w	Vector representing the stimulus with respect to a reduced basis $\hat{i} = \mathbf{Rw}$
Q	Quadratic matrix, e.g., $\mathbf{Q} = \mathbf{S}^t\mathbf{S}$
Π_N	General sampling operator
l_a	Length of the sensor response vector to a stimulus a.

Example

As a familiar example, consider the sensors used by Wilson and Gelb.[20] We use these sensors because they are simple and thus are easy to use to illustrate the representation. The Wilson–Gelb sensors are one dimensional. They are constructed by forming linear combinations of three Gaussians with zero mean but different standard deviations. Figure 1 shows 3 of the 18 sensors selected by Wilson and Gelb (from a potentially much larger set) to model frequency discrimination in the central fovea. These sensors are represented in the computer computations at a sampling resolution of 120 samples per degree of visual angle. The two receptive fields at the top of the figure have the same spatial profile but are centered at different spatial locations. The third sensor has a spatial profile different from those of the first two.

The spatial sensitivity of the 18 sensors defines the 18 rows of the sensor matrix. Since the input signal is assumed to vary only in the horizontal direction, $N_y = 1$. The value of N_x can be chosen as the number of independent degrees of freedom in the stimulus. Because of the blurring imposed by the ocular optics [roughly a 60-cycle-per-degree (cpd) cutoff], retinal stimuli can be specified by using a band-limited representation with 120 degrees of freedom per degree of visual angle.[21] In the vector notation, we use the sample values across space to represent the stimulus. The Wilson–Gelb model is defined for the central 3 deg of the visual field so that $N_x = 360$, and thus $N_xN_y = N = 360$. Since there are 18 sensors, $P = 18$, and $\mathbf{S}$ is a $(P = 18) \times (N = 360)$ matrix.

Watson and co-workers[15,19,22] have described a model (hereafter referred to as the Watson model) with 2D sensors whose spatial sensitivities are defined as 2D Gabor functions: that is, as oriented sinusoids windowed by 2D Gaussian envelopes. When applied to a 2-deg field, the complete model typically consists of $P = 89{,}670$ sensors. It is impractical to perform calculations requiring this many linear sensors by using the matrix formulation that we have described here. As an alternative strategy, then, most calculations with the Watson model have been done using stimuli that are linear combinations of Gabor functions. When the sensor spatial sensitivities and stimuli are Gabor functions, the sensor response can be calculated analytically. This solution greatly reduces computational requirements and makes experimenting with the model possible. The computational efficiency for Gabor-function stimuli cannot be generalized to other stimuli. Only limited experimentation with other stimuli has been possible.

Restricted Stimulus Domain

There are two difficulties to overcome in working with these models of spatial-pattern sensitivity. The large number of sensors in the Watson model makes it difficult to use the model interactively and to experiment with its properties. Although the Wilson–Gelb calculation is far simpler, it is obvious that these 18 sensors do not capture the behavior of the entire visual system. Instead, the sensors used in the calculations of Wilson and Gelb were selected by experimenting in order to bring the model's predictions into agreement with the data. Working with a cumbersome model or working with a subset of the model determined by an attempt to come to grips with a single data set are undesirable procedures for reasoning about visual performance.

We suggest an alternative procedure. We believe that the entire set of sensors should be specified. Most theorists would choose a large number of sensors, reflecting the complexity of visual performance. To make the analysis manageable, we suggest that the calculation be simplified by incorporating stimulus restrictions into the model. This can be done with precision and with great computational savings when the stimulus domain can be described as the weighted sum of a small set of stimuli.

A set of stimuli from which all other stimuli under consideration may be formed by linear combination is called a basis set. We define the matrix $\mathbf{R}$ whose columns consist of the basis set. Each experimental stimulus can be described by a vector $\mathbf{w}$ whose entries are the weights of the basis terms that sum to the stimulus. For a stimulus of unit vector length (i.e., unit contrast energy) we will have

$$\hat{i} = \mathbf{R}\mathbf{w}. \tag{1}$$

For a general stimulus αi it follows that $\alpha\hat{i} = \alpha\mathbf{R}\mathbf{w}$. The vector $\mathbf{w}$ may be of reduced dimensionality compared with the stimulus vector $\alpha\hat{i}$. The number of arithmetic operations required to compute the sensor responses is reduced, since the equation $\alpha\mathbf{r} = \mathbf{S}\alpha\hat{i}$ becomes $\alpha\mathbf{r} = (\mathbf{SR})\alpha\mathbf{w}$. Depending on the degree of restriction in the stimulus domain, the computational savings can be significant.

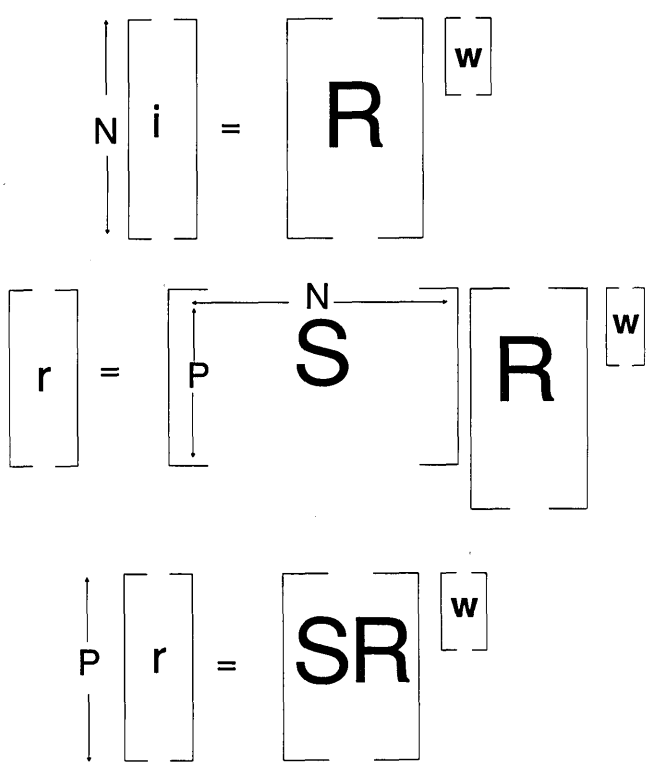


Fig. 2. Reduction of the computational size of the initial linear transformation the representation of the stimulus with respect to an efficient set of bases. P is the number of sensors; N is the number of degrees of freedom in the original stimulus. The dimension of the vector $\mathbf{w}$ depends on the stimulus basis chosen for the experiment. In the example in the text, the vector's dimension is 64.

Stimuli Limited in Frequency and Space

For example, suppose that we study performance using one-dimensional stimuli that are band limited at 16 cpd and extending over 2 deg of the retina. From the Fourier-series representation we can derive that this stimulus domain can be constructed from a basis set consisting of 64 different terms: 32 sinusoidal images and 32 cosinusoidal images. The matrix **R** consists of N rows and 64 columns, and the stimulus vector **w** is 64-dimensional. All one-dimensional, band-limited images can be constructed as the weighted sum of these 64 columns. Since a complete model will be specified for a stimulus set of $N \gg 64$ degrees of freedom, and the stimulus set described here only contains 64 independent degrees of freedom, we can incorporate the savings by a one-time calculation in which the stimulus set restriction is incorporated into the **S** matrix. To compute the response of the linear component of the system for any of the experimental images, we need only compute

$$\alpha \mathbf{r} = \mathbf{S}\alpha\hat{i} = (\mathbf{SR})\alpha\mathbf{w}. \tag{2}$$

The nature of the reductions is illustrated in Fig. 2. The matrix (**SR**) is reduced to $P \times 64$ from $P \times N$. For example, for a 2-deg field, typical values for the Watson model are $P = 89670$ and $N = 65{,}536$. Yet all predictions involving one-dimensional sinusoidal input images band limited at 16 cpd can be performed by using a matrix that is reduced in size to $(P = 89{,}670) \times 64$, yielding a net savings of 89,670 (65,536 −64) multiplications per stimulus calculation. For the small Wilson–Gelb computation, the matrix is reduced from 18×360 to 18×64, yielding net savings of (18)(296) multiplications per stimulus calculation.

Stimuli Restricted in Dimension

The computational complexity can be reduced even if the stimulus set is known only to be limited to being one-dimensional but not band limited. Recall that the 2D stimuli are represented as column vectors and that the first N_x entries describe the contrast of the first row of the image. The complete set of one-dimensional stimuli that vary only in the vertical direction can be written as weighted sums of vectors of the form $(1, 1, 1, \ldots, 0, 0, 0, \ldots, 0, 0, 0)^t$ and $(0, 0, 0, \ldots, 1, 1, 1, \ldots, 0, 0, 0)^t$, etc. The proper method of reducing a model from two dimensions to one dimension is to premultiply the complete matrix **S** by the matrix **R** whose ith column contains ones in the entries corresponding to the ith row of the image and zeros elsewhere. The reduced sensor matrix can be used for all subsequent calculations with respect to one-dimensional stimuli. In the case of the Watson calculation for stimuli in the central 2 deg, the **S** matrix is reduced in its row length from $N = 256 \times 256 = 65{,}536$ to $N = 256$. In the case of the calculations of Wilson–Regan,[8] this procedure offers a principled method of reducing that 2D model for comparison with the one-dimensional model described in Ref. 23.

Stimuli Restricted in Space

It is common to run experiments in which the stimuli are presented to only a small part of the visual field, such as the fovea. This restriction is easily incorporated by noting that all stimuli are weighted combinations of δ functions over the restricted range. If there are M points in the visual field where the stimuli might be presented, then the proper method of restricting the model is to include only those columns of the sensor matrix that correspond to points in the visual field where stimuli may be present. We can discard the parts of the sensor matrix that correspond to points in the image that are always zero.[24]

Application to Phase Discrimination

Finally, consider the simplification that is possible when we analyze the problem of discriminating the phase relationship between a cosinusoid at a frequency f and a second cosinusoid at a frequency f'. In a phase-discrimination experiment the observer is asked to discriminate between pairs of stimuli at f and f' set in different phase relationships with respect to each other.

The input stimuli for this experiment can be described by the equation $\alpha\hat{i} = \mathbf{R}\alpha\mathbf{w}$, where **w** is a three-dimensional vector. The test stimuli for a phase-discrimination experiment can be constructed by forming the sum of a fixed cosinusoid (f) and the weighted sum of a cosinusoid and a sinusoid at f'. The relative weights of the cosinusoid and the sinusoid at f' determine the phase relationship between the component at f and the component at f'. The first column of restricted stimulus set matrix **R** is the cosinusoid at f, and the second and third columns are the sinusoid and the cosinusoid at f'. For the Wilson–Gelb calculation, the linear part of the computation required for predicting discrimination is reduced to a sensor matrix that is $(P = 18) \times (N = 3)$ rather than a sensor matrix that is $(P = 18) \times (N = 360)$.

Transform to New Basis

The incorporation of stimulus restrictions is a special case of representing the model with respect to an alternative basis for the stimulus set. When we define the spatial sensitivity of the sensors with respect to individual points in space, we implicitly define the model with respect to the spatial-basis functions consisting of such points. We may wish to represent our stimuli in an alternative basis that is made complete by defining a linear transformation such that

$$\alpha\hat{i} = \mathbf{B}\alpha\mathbf{w}, \tag{3}$$

where the matrix **B** is a square nonsingular matrix whose columns are an alternative representation of the stimulus. The matrix **B** plays a role similar to that of the matrix **R** except that no stimulus restrictions are implied and any stimulus can be written uniquely as the weighted sum of columns of **B**. A restricted stimulus matrix **R** spans only a subset of the possible stimuli.

The decision to represent the computation with respect to a different complete basis set may occur for theoretical reasons or for computational reasons. Perhaps the most important change of basis is from the point representation to the Fourier representation. In that case the columns of the matrix **B** are the sinusoidal functions of space that make up the Fourier-basis terms.

Null Space

Our focus in this article is on specification of spatial pattern models that make predictions of the visibility of different target stimuli or the discriminability of pairs of stimuli. A particularly important class of stimuli comprises the stimuli that are completely invisible, that is, stimuli whose model responses are the same as an input stimulus with zero con-

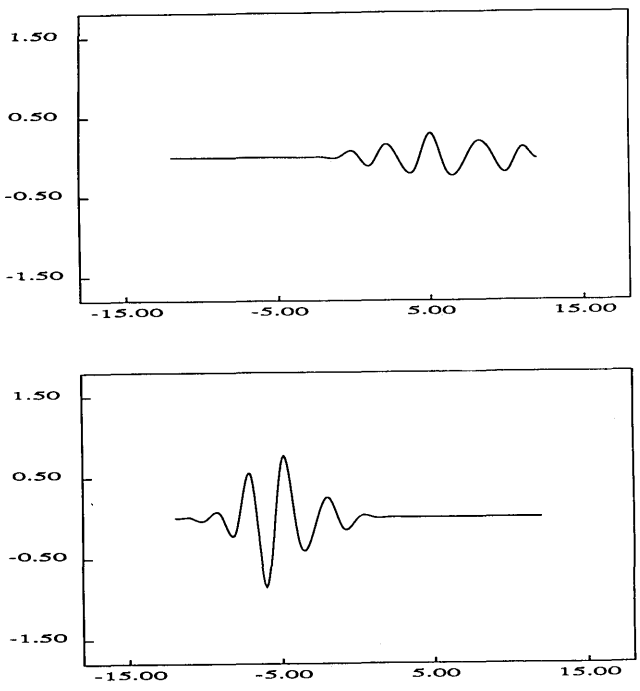


Fig. 3. Two stimuli that are indiscriminable to the Wilson–Gelb sensors. The difference between these two stimuli is within the null space of the Wilson–Gelb sensor set. The zero point marks the central fovea, so that one stimulus falls to the left of the central fovea and the other falls to the right.

trast. The set of stimuli that generate zero output of the initial matrix is called the null space of the matrix. The elements that map to zero can be considered a legitimate subspace. If z_1 maps to zero, that is, if $\mathbf{S}z_1 = 0$ and $\mathbf{S}z_2 = 0$, then $\mathbf{S}(z_1 + z_2) = 0$. Also, if $\mathbf{S}z_1 = 0$, then $\mathbf{S}(\alpha z_1) = 0$.

For some calculations, the null space can be quite large. For example, in the Wilson–Gelb calculation the initial linear stage has an enormous null space. This can be deduced from the fact that the linear transformation maps a 360-dimensional stimulus space into an 18-dimensional sensor space. Since the original representation of the 3-deg, one-dimensional stimuli requires a 360-dimensional vector and the responses of the sensors are described by an 18-dimensional vector, the null space must have a dimensionality of $360 - 18 = 342$. The set of stimuli that generate zero responses is of broad significance, since whenever the difference between any two stimuli is in the null space, the two stimuli are necessarily indiscriminable to the model, independent of subsequent processing.

The null space of the Wilson–Gelb calculation[7] can be computed from its singular-value decomposition (SVD; see Appendix A). Two stimuli whose difference is within the null space are shown in Fig. 3. The existence of a significant null space for the Wilson–Gelb calculation highlights the different strategies used by investigators in selecting their initial linear transformation. In most of their calculations, Watson and his co-workers require that the initial encoding be complete in the sense that it have a null space consisting of only the zero vector. This greatly increases the computational burden but ensures a complete representation of the signal. By contrast, in developing the calculation for frequency discrimination, Wilson and Gelb performed their calculations by selecting a subset of the sensors that would be required for a complete model. Using the formulation here, we can say that Wilson and Gelb reduced their calculation by eliminating some of the rows of the sensor matrix. We have argued that the computation should be restricted by incorporating stimulus restrictions into the sensor matrix. This can be accomplished by multiplying the matrix representing the linear model of stimulus contrasts $\mathbf{R}$ with the complete $\mathbf{S}$ matrix. The resulting matrix is a reduced sensor matrix that is applicable to all measurements in the restricted stimulus space. The effect of this procedure is to reduce the column dimension of the matrix, not the row dimension. This method of reducing the computation ensures consistency across measurement conditions.

Limitations imposed by the null space are not the same as a spatial-frequency bandpass limit. The Wilson–Gelb sensors illustrate this point. This model generates a contrast-sensitivity function (CSF) that is consistent with data obtained with human subjects. The concordance of the model predictions with respect to the CSF and the discordance of the model with respect to the predicted discriminability of weak stimuli illustrate a difficulty with relying on the CSF to assess a model's predictions. As we explain Section 3, the CSF is not generally a satisfactory characterization of the performance of a sensor set.

Aliasing and the Null Space

The proper interpretation of the null space can be illustrated by considering the set of stimuli that are equivalent with respect to the model's performance. These stimuli are analogous to metameric matches: the stimuli are physically different, but they cause the same response in the sensor set. When two stimuli are mapped by a linear operator into the same response, that is, when they are aliased, then the difference between the two stimuli is in the null space of the linear operator. Thus the stimuli that are aliased into one another by a linear operator all differ by an element of the null space of the linear operator.

As an example, consider the equivalence class of stimuli with respect to the linear operator of general sampling, which includes nonuniform sampling. If we denote the general sampling operation that maps $f(x)$ into a discrete set of values by Π_N: $f(x) \rightarrow f(i)$, $i = 1, \ldots, N$, then notice that

$$\Pi_N[\alpha f(x)] = \alpha \Pi_N[f(x)]$$

and

$$\Pi_N[f(x) + g(x)] = \Pi_N[f(x)] + \Pi_N[g(x)].$$

Since the general sampling operator is a linear operator, we represent it as a matrix. A sampling operator is a rectangular matrix consisting of only ones and zeros.[25]

The null space of the general sampling operator has the special property that, for any two functions f and g,

$$\Pi_N[f(x)g(x)] = \Pi_N[f(x)]\Pi_N[g(x)]. \quad (4)$$

It follows that if $f(x)$ or $g(x)$ is in the null space of the general sampling operator, then $f(x)g(x)$ must be as well.

Aliasing among sinusoids for the special case of the uniform sampling operator can be understood to arise in the following way. Two signals are aliased into each other by a

general linear operator precisely when their difference is in the null space of that operator. In particular, two cosinusoids will be aliased into each other by the uniform sampling operator if their difference is in the null space. The difference of two cosinusoids is

$$\cos\left[\frac{2\pi x}{N}(f+\Delta)\right]-\cos\left[\frac{2\pi x}{N}(f-\Delta)\right]=2\sin\left(\frac{2\pi x}{N}f\right)\sin\left(\frac{2\pi x}{N}\Delta\right). \quad (5)$$

It follows that if either $\sin[(2\pi x/N)f]$ or $\sin[(2\pi x/N)\Delta]$ is in the null space of the uniform sampling operator, then the cosinusoids at $f+\Delta$ and $f-\Delta$ will be indiscriminable. The sinusoid at $f = N/2$ is in the null space of the uniform sampling operator, and thus the cosinusoid at $N/2+\Delta$ will be aliased into the cosinusoid at $N/2-\Delta$. Aliasing between sinusoidal signals occurs because their difference lies in the null space of the uniform sampling operator.

This example illustrates that the null-space characterization of the initial linear transformation does not inform us about frequency selectivity but instead informs us about the equivalence classes of stimuli with respect to the linear operator. Cosinusoids that are symmetric about the Nyquist frequency are aliased into each other because their difference lies in the null space of one particular linear operator, the uniform sampling operator. If the sampling operator is not uniform, then these same cosinusoids will not be aliased pairs. The principle that aliased pairs differ by a signal in the linear operator's null space remains true for all operators. As recent studies of retinal photoreceptor sampling are extended to nonuniform regions of the retinal mosaic,[26–28] the null-space analysis can be used to extend the classic methods based on cosinusoids and Fourier transforms.

3. A QUADRATIC MEASURE

Two strategies have been used to test and develop theories of spatial vision. Some authors develop a complete computable description of the theory, including a specific choice of sensors and rules for mapping the sensor responses into performance.[7,9,10] The selection of sensors is based on physiological data, statistical signal-processing arguments, or masking data. The mapping of sensor responses to performance is usually guided by the shape of the psychometric function or the shape of masking functions.[29] A second strategy is to test hypotheses about the classes of initial linear transformations, without committing to any particular set of sensors. In this case the authors must fix their assumptions about the decision stage. A lovely example is given by Graham *et al.*,[30] who tested the hypothesis that a shift-invariant linear transformation can be used to predict the threshold sensitivity to test mixtures. Graham *et al.* used a decision stage in which performance is predicted from the length of the response vector (contrast energy). If the initial linear transformation is shift invariant, then a sinusoidal input will map into a sinusoidal output. Since sinusoids of different frequencies are orthogonal, the resulting sensor-response vectors will be orthogonal. The Pythagorean theorem permits them to predict the length of the sum of the orthogonal response vectors.

Both approaches pose the problem of spatial sensitivity in the following form: given some assumptions concerning a linear transformation from stimulus to sensor responses and a nonlinear mapping from sensor responses to performance, predict the threshold. We suggest a method of turning the problem around to pose it in this form: given threshold data and a specific nonlinear mapping from sensors to performance, specify the initial linear transformation. The approach that we describe here emphasizes the measurement of performance as opposed to the prediction of performance. To explain the approach, we find it useful to decompose the analysis into a series of steps, beginning with an initial encoding that consists of only a single sensor and then proceeding to the case of an initial encoding that is shift invariant and finally to an initial encoding that is a general linear model.

Single-Sensor Linear Encoding

Probably the simplest theory of spatial-pattern vision is to suppose that sensitivity can be modeled by a single linear sensor. In this case, $P = 1$, and the sensor response $\alpha\mathbf{r}$ is a single number calculated as the inner product of the image $\hat{i}$ and the single row of the $1 \times N$ sensor matrix $\mathbf{S}$. The matrix structure for this case is illustrated in Fig. 4. For computational clarity it is convenient to assume that the probability of a correct response in a two-alternative forced-choice detection task is monotonic with the squared sensor length (contrast energy). This imposes no additional restrictions, since the relationship between the length of the response vector and the performance is coupled by only a monotonic transformation. We write the relationship between the sensor response and the observed probability of a correct response as

$$P(\text{correct}) = F[(\alpha\mathbf{r})^2].$$

[To fix the unit of response, we assume that $F(1.0) = 0.75$. We refer to the 0.75 performance level as the threshold in a two-alternative forced-choice experiment.]

Detection data can easily be used to reject this model.[31] If the stimuli a and b are equally detectable, then from the monotonicity of F it follows that

$$\alpha_a\mathbf{r}_a = \pm\alpha_b\mathbf{r}_b,$$

where α_a and α_b are the stimulus contrasts. The linearity of the sensor guarantees that superimposing the stimuli yields a mechanism response of

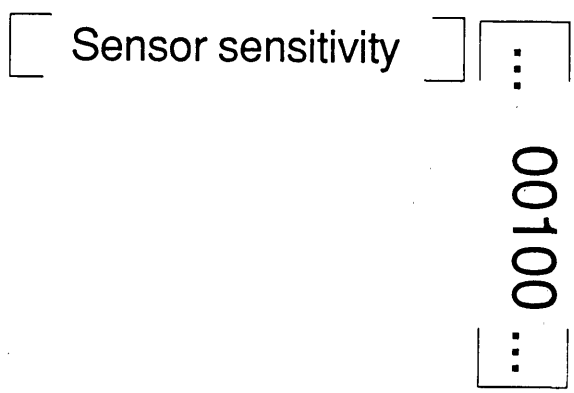


Fig. 4. Single-sensor analysis. N measurements of sensitivity to shifted delta functions permit an estimate of the values of a single-sensor model.

$$\alpha_a \mathbf{r}_a \pm \alpha_b \mathbf{r}_b = \begin{cases} 2\alpha_a \mathbf{r}_a \\ 0 \end{cases}. \quad (6)$$

It follows that the superposition of the two stimuli must be either invisible or as detectable as stimulus a at double its contrast. This prediction is easily falsified by using point stimuli at different locations in the visual field or mixtures of sinusoidal gratings at widely separated spatial frequencies.[31]

To specify this single-sensor model completely, we need estimate only the entries of the **S** matrix. If a single-sensor model holds, then we can use threshold data in order to recover all the entries **S**. The sensitivity to an input stimulus with energy only at the ith position, $\delta_i = (0, 0, \ldots, 1, \ldots, 0, 0)^t$, will be proportional to the absolute value of **S** at position $(1, i)$. (The sign is lost because of the squaring operation.) The effect of a delta input is illustrated in Fig. 4. By measuring the N thresholds to the shifted δ functions, $\delta_i, i = 1, \ldots, N$, the entries of **S** can be specified completely except for their signs. The signs of the entries can be resolved by using test-mixture experiments and Eq. (6).

Measuring the threshold to any N independent stimulus vectors provides enough information to estimate the **S** matrix for this model. As a special case, the CSF, which describes the threshold sensitivity to sinusoidal contrast patterns, provides enough information to estimate the entries of the **S** matrix (except sign).

Shift-Invariant Linear Encoding

The simplest model with a plausible set of sensors consists of sensors at evenly spaced retinal positions, with each sensor having the same receptive-field profile. In this case the rows of the matrix **S** are all identical except for a shift of one position. This matrix describes a shift-invariant linear system, as in the case analyzed by Graham *et al.*[30] (This sensor matrix is frequently referred to as a single-channel model.) Multiplication of the input $\alpha\hat{i}$ with the matrix **S** is simply the convolution of the image with the spatial-response function of the receptive field. We remind the reader that it is only for shift-invariant linear systems that a sinusoidal input yields a sinusoidal system response.

In order to relate sensor values to the probability of a correct response, we suppose that the dominant noise in the system is independent, equal-variance Gaussian noise in the sensors. For this signal the ideal observer will use the squared length of $\alpha\mathbf{r}$ as a decision statistic.[9,32] The squared length of the response vector is $\alpha\mathbf{r}^t\alpha\mathbf{r}$, or, equivalently,

$$l^2 = \alpha\hat{i}^t\mathbf{S}^t\mathbf{S}\alpha\hat{i}. \quad (7)$$

We assume that observer performance is related monotonically to the performance of the ideal observer. Specifically, we assume that the probability of a correct response in a two-alternative forced-choice experiment is monotonic with the squared length of the response vector, the decision statistic that governs ideal-observer behavior. Let l_a be the length of the response to stimulus a at unit contrast. It follows that

$$P(\text{correct}) = F[(\alpha_a l_a)^2] \quad (8)$$

so that at the threshold $[P(\text{correct}) = 0.75 = F(1)]$

$$l_a = \frac{1}{\alpha_a}. \quad (9)$$

This is a quadratic model of performance. Starting with Eq. (7) we define the symmetric matrix $\mathbf{Q} = \mathbf{S}^t\mathbf{S}$ so that the probability of a correct response for any input image can be described in matrix notation by using the quadratic form

$$P(\text{correct}) = F(\alpha\hat{i}^t\mathbf{Q}\alpha\hat{i}). \quad (10)$$

The free parameters of this model are the values of the symmetric matrix **Q** and the form of the monotonic function $F(\ )$. Thus any two matrices $\mathbf{S}_1$ and $\mathbf{S}_2$, such that

$$\mathbf{Q} = \mathbf{S}_1{}^t\mathbf{S}_1 = \mathbf{S}_2{}^t\mathbf{S}_2, \quad (11)$$

will make identical predictions (see Ref. 15).

The matrix **Q** is always square and has dimensions equal to the number of columns of the sensor matrix. When we incorporate the restrictions of the stimulus set into the sensor matrix **S**, we obtain a new sensor matrix, **SR** [see Eq. (2)]. The column dimension of **SR** is smaller than the column dimension of **S**. When we limit the number of independent degrees of freedom in the stimuli, the number of model parameters is reduced correspondingly.

We wish to estimate **Q**. A simple relationship emerges for the common experimental case in which the test stimuli are one dimensional.[33] If the initial linear transformation is shift invariant, the terms along each of the diagonals of **Q** are equal (i.e., $q_{i,i+k} = q_{j,j+k}$ for each i, j, k). Matrices with this form are called Toeplitz matrices. Notice that this matrix has only N degrees of freedom, since all diagonals are constrained to be the same and the matrix is symmetric.

When the **Q** matrix of a one-dimensional shift-invariant linear system is represented with respect to the Fourier basis, the matrix is diagonalized. Informally, this can be seen as follows. To represent the stimuli with respect to the Fourier basis, we form the matrix **SB** in which the columns of **B** are sinusoids. If the **S** matrix is shift invariant, then the output to a sinusoidal input will also be a sinusoid. The entries of $\mathbf{Q} = \mathbf{B}^t\mathbf{S}^t\mathbf{S}\mathbf{B}$ are all dot products of sinusoids with one another. Since the sinusoids are orthogonal to one another, the only nonzero terms of **Q** are the diagonal terms that correspond to the dot product of a sinusoid with itself. From the construction just described, it follows that the diagonal entries are equal to the squared contrast sensitivity (i.e., the power spectrum of the point-spread function) associated with the **S** matrix.

The simple structure of the **Q**-matrix shift-invariant system has two important effects. First, we can test whether the initial linear transformation might be modeled as shift invariant. If we measure with respect to the delta-function basis, the estimated **Q** matrix should be a Toeplitz matrix. If we measure with respect to the Fourier basis, the estimated **Q** matrix should be diagonal. Second, since a diagonal (or Toeplitz) matrix has only N independent parameters, **Q** can be estimated directly from N independent measurements. In particular, the CSF can serve as a complete estimate (except for sign).

General Linear Encoding

When the sensor matrix is not shift invariant, the formula for predicting the sensitivity remains as in Eq. (10), but two simplifications from shift invariance are lost. First, sinusoidal inputs do not map into sinusoidal outputs. Second, the matrix **Q** does not have a constant value along each of its diagonals. Rather than having only N degrees of freedom,

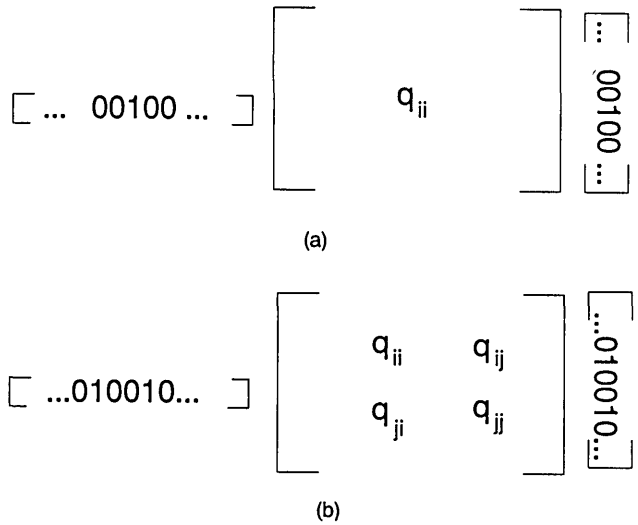


Fig. 5. Quadratic model analysis. (a) N measurements of sensitivity to shifted delta functions permit an estimate of the diagonal entries of the **Q** matrix. (b) $N(N+1)/2$ measurements of shifted delta functions and sums of shifted delta functions are required to estimate all of the parameters of the symmetric **Q** matrix. Shifted delta functions yield estimates along the diagonal. Once these are known, sums of shifted delta functions yield estimates of the off-diagonal entries.

the **Q** matrix has all the degrees of freedom of a symmetric matrix: $N(N+1)/2$.

The number of measurements needed to estimate the matrix **Q** is greater in the general case than in the shift-invariant case. First, consider the point stimuli δ_i. We measure the contrast α such that

$$F(\alpha\delta_i{}^t\mathbf{Q}\alpha\delta_i) = F(\alpha^2 q_{ii}) = P(\text{correct}) = 0.75, \qquad (12)$$

where q_{ij} is the ijth entry of **Q** and q_{ii} is the ith diagonal entry of the matrix **Q**. Using Eqs. (8) and (9), we can estimate the value of q_{ii}. By measuring the sensitivity to points of light at each of the input sample positions, we can estimate all the diagonal terms of **Q**. This is illustrated in Fig. 5(a).

The off-diagonal entries of the matrix can be estimated from the sensitivity to mixtures of points, $\delta_i + \delta_j = (0, \ldots 1, \ldots, 0, \ldots, 1, \ldots, 0)^t$. This is illustrated in Fig. 5(b). From the symmetry of the matrix we can derive that

$$(\delta_i + \delta_j)^t\mathbf{Q}(\delta_i + \delta_j) = q_{ii} + 2q_{ij} + q_{jj}. \qquad (13)$$

Since the diagonal terms are known [see Fig. 5(a)], the contrast threshold for the mixture of the two points determines the ij entry:

$$q_{ij} = \frac{1}{2}\left(\frac{1}{\alpha_{ij}{}^2} - q_{ii} - q_{jj}\right). \qquad (14)$$

From thresholds to single points and pairs of points, we can determine all the entries of the matrix **Q**.

Notice that the procedure for estimating the entries of **Q** is not dependent on the stimulus representation. It follows that experimental estimates of **Q** can be made either by using mixtures of point stimuli, as above, or by using mixtures of stimuli in alternative basis representations, such as the Fourier basis.

The CSF determines only the diagonal entries of **Q**. It follows that in the general case the CSF is insufficient to predict the threshold to other stimuli, such as mixtures of sinusoids. There are many quadratic models with identical CSF's (i.e., diagonals of **Q**) but different mixture properties (i.e., off-diagonal elements of **Q**). Each of these will make different predictions concerning test-mixture experiments and pattern sensitivity in general. Each model will have distinguishably different sensors but the same CSF. For a model in which the vector length serves as the decision statistic, the **Q** matrix serves as a generalization of the CSF when the initial linear transformation is not shift invariant.

How to Estimate Q

The experimental procedure required for estimating the matrix **Q** with respect to a small basis set is straightforward. We describe a step-by-step method here.

First, select a (small) set of stimuli for testing. For concreteness, let us suppose that we choose five Gabor patches, in cosine phase, at spatial frequencies ranging from 1 to 5 cpd. Call the matrix whose five columns are the spatial contrasts of these stimuli **R**. We shall derive a quadratic model that predicts the sensitivity to patterns generated by additive mixtures of these stimuli, **Rw**.

The underlying sensor matrix **S** of the visual system may be huge, but for present purposes the only matrix that will play a role in the computation can be derived as $\mathbf{Q} = (\mathbf{R}^t\mathbf{S}^t)(\mathbf{SR})$, which is a 5×5 matrix. To estimate the entries of this matrix we must make 15 threshold measurements: the five Gabor patches, individually, and the pairwise mixtures of Gabor patches in some fixed contrast ratio. This is a total of $n(n+1)/2$ ($n = 5$) threshold measurements. From the original 15 threshold measurements, we can use Eqs. (12) and (14) to derive the entries of the matrix **Q**. (Remember that the matrix is symmetric.) The power of the quadratic approximation to the data is this: we can use **Q** to predict the sensitivity to any stimulus that is the weighted sum of the columns of **R**. Included in this set are stimuli that are closely related to the initial 15 stimuli (e.g., mixtures of two Gabor patches but in different contrast ratios) as well as stimuli that have not been measured before (e.g., weighted sums of four Gabor patches).

We are unaware of measurements that permit us to estimate a complete **Q** matrix by using more than two stimulus components. It is possible to use data such as those collected by Graham *et al.*[30] using two components in order to get a sense of how the quadratic model can be applied to threshold measurements.

Graham *et al.* measured the contrast sensitivity to mixtures of 2- and 6-cpd Gabor patches in the fovea and periphery. The Gaussian envelope of the Gabor function fell to $1/e$ at a distance of 0.75 deg from the peak. The stimuli were presented with a Gaussian time course. The time to $1/e$ from the peak of the Gaussian was 100 msec. The 6-cpd component was added in either the peaks-added phase or the peaks-subtracted phase.

Since the stimulus set consists of two components, the **Q** matrix is 2×2 matrix. Graham *et al.* measured the threshold for 10 mixtures. The threshold contrast in the first component for the ith mixture is denoted by $\alpha_{1,i}$, the threshold contrast in the second component is denoted by $\alpha_{2,i}$, and the threshold vector is denoted by $\mathbf{w}_i = (\alpha_{1,i}, \alpha_{2,i})^t$. The quadratic model predicts that there is a matrix **Q** such that

$$\mathbf{w}_i^t \mathbf{Q} \mathbf{w}_i = 1.0, \qquad i = 1, \ldots, 10. \tag{15}$$

As we have shown above, when the initial sensor matrix is shift invariant and we measure with respect to sinusoidal stimuli, then we expect that the **Q** matrix will be diagonal. When the initial sensor matrix is general, then the matrix need only be a quadratic form.

We have estimated the best **Q** matrix required to solve the 10 simultaneous equations in Eq. (15). If the fit were perfect, then we would expect that there would be a set of values for **Q** such that Eq. (15) holds exactly for each data point.[34] Using a recent version of Chandler's STEPIT routine,[35] we estimated the best-fitting **Q** matrix for the data of Graham *et al.* under the assumption that **S** is a general linear transformation. For example, the estimated matrix for observer (NG, fovea) was

$$\mathbf{Q} = \begin{bmatrix} 0.83 & 0.088 \\ 0.088 & 0.862 \end{bmatrix}. \tag{16}$$

We now discuss two issues concerning the quality of the

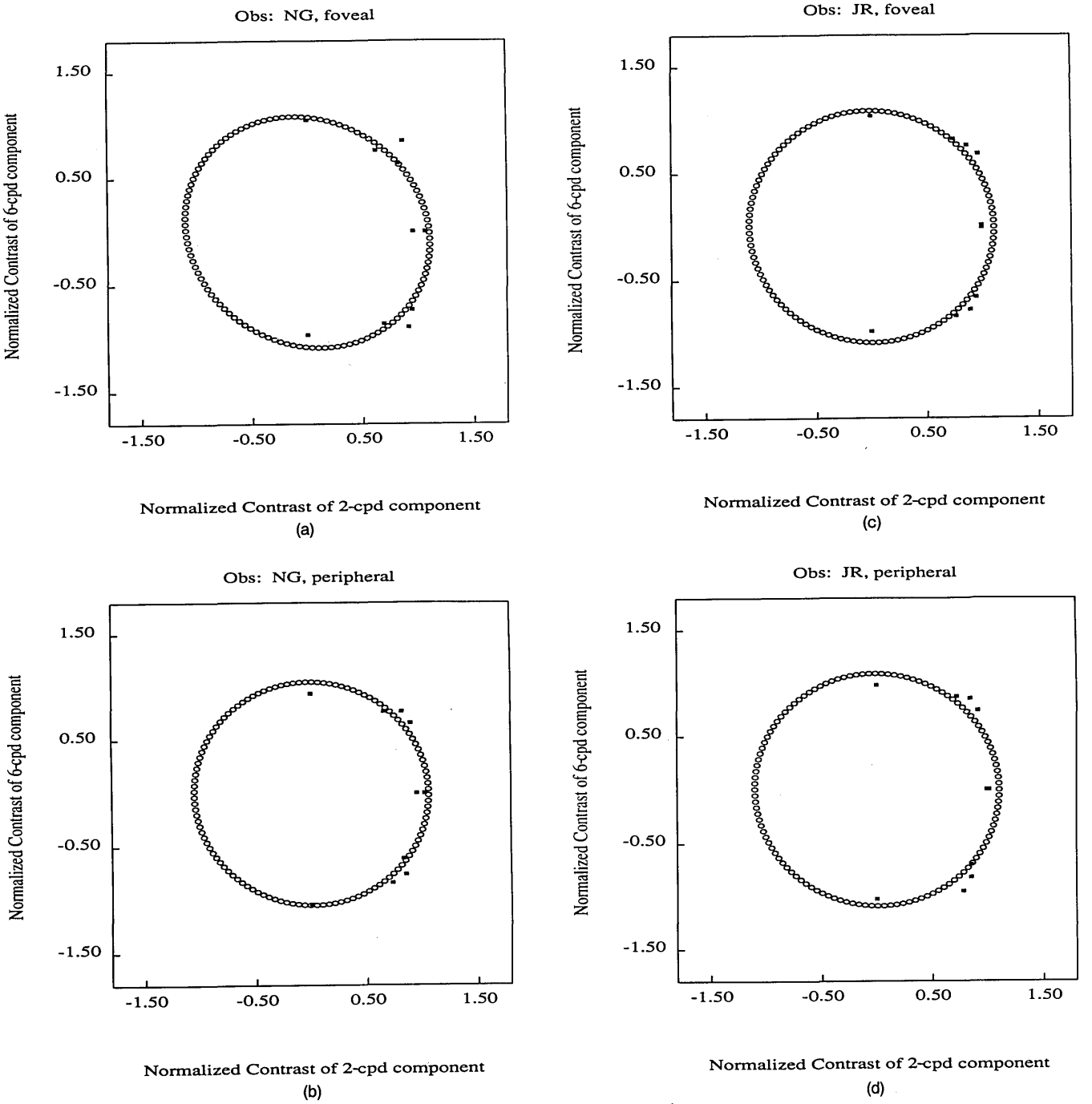


Fig. 6. The test-mixture thresholds of the two observers in two conditions from Ref. 30 are plotted along with the best-fitting ellipse. The ellipse was derived by estimating the quadratic matrix **Q**. The data fall near the ellipse, but the deviations are systematic. Data on the axes fall inside the predicted ellipse, whereas the points near the identity fall outside the ellipse. Data points plotted in the first quadrant refer to test mixtures in the peaks-added phase. Data points plotted in the fourth quadrant refer to test mixtures in the peaks-subtracted phase.

approximations of the data by the **Q** matrix. First, we ask whether the quadratic approximation to the results is an adequate summary of the data. Second, we compare the quality of the fits from a general quadratic approximation with those from a quadratic approximation based on a shift-invariant sensor set.

In Fig. 6 we plot the data along with the best-fitting ellipse. The root-mean-squared errors comparing the data points with the ellipse for observer NG were 0.088 (foveal) and 0.065 (peripheral). For observer JR the root-mean-squared error values were 0.070 (foveal) and 0.081 (peripheral). The size of the deviations is not significantly greater than the error of measurement. However, as Graham *et al.* pointed out, the deviations are systematic. For all observers, the threshold to the points on the axes falls slightly inside the predicted ellipse, whereas points along the identity line fall slightly outside the ellipse.

Our assessment of the quadratic approximation is divided into two parts. First, the model cannot be rejected on the basis of the size of the deviations. If one is interested in the practical matter of estimating contrast thresholds, and if these data are representative, then the quadratic calculation will provide a good approximation, usually within 6–8%. This is comparable with the quality of the MacAdam ellipses widely used in industrial applications of color measurement.[36,37] On the other hand, the approximation has not been tested as extensively for spatial stimuli as it has for mixtures of colored stimuli. We believe that it is important to define the range of measurement conditions over which the quadratic approximation can be applied legitimately for practical measurement.

On the other hand, as Graham *et al.* pointed out, the quadratic model can be rejected on the basis of the systematic nature of these small deviations. The small numerical size of the deviations suggests that it may be easier to construct new hypotheses from other kinds of experimental data. There is rather little left unexplained by the quadratic approximations to these sensitivity data.

If we decide to accept the quadratic approximation, then we can ask whether the data distinguish between a shift-invariant **Q** approximation and a general **Q** approximation. The large changes in pattern sensitivity as a function of retinal eccentricity imply[38–40] that for stimuli extending over substantial regions of the retina, a shift-invariant transformation is inappropriate. For stimuli with small visual extents, such as the stimuli used by Graham *et al.*, the variation in pattern sensitivity that is caused by retinal eccentricity is quite small. To evaluate the shift-invariant approximation for these stimuli, we have estimated the best-fitting diagonal **Q** matrices. (For the purpose of this discussion, we assume that a Gabor function can be used as a local approximation to a sinusoid.) The fits from the diagonal **Q** matrices are nearly as good as the fits obtained by using the unconstrained fits reported above. We think it is likely, however, that the shift-invariant approximation will fail for stimuli that extend over a substantial portion of the visual field.

Uniqueness of Sensor Estimates: A Free Unitary Transformation

We noted earlier that the predictions of the model are determined completely by the entries of **Q**. It follows that any two sensor sets such that

$$\mathbf{S}_1{}^t\mathbf{S}_1 = \mathbf{S}_2{}^t\mathbf{S}_2 \tag{17}$$

will yield equivalent predictions. Using the SVD (see Appendix A), we can write these sensor matrices as

$$\mathbf{S}_1 = U_1 D_1 V_1{}^t \tag{18}$$

and

$$\mathbf{S}_2 = U_2 D_2 V_2{}^t. \tag{19}$$

Since $U_1{}^t U_1$ is the identity matrix,

$$\mathbf{S}_1{}^t\mathbf{S}_1 = V_1 (D_1)^2 V_1{}^t, \tag{20}$$

and the same relationship applies for the second sensor set. It follows that whenever two sensor sets have the same singular values and the same right singular matrix, they will make identical predictions with respect to the theory. The final rotation U_i is irrelevant to the determination of the length of the response vector. Although the matrix **Q** can be determined exactly from empirical measurements, the matrix **S** cannot be determined exactly. Instead, **S** can be estimated only up to a final free unitary transformation. This freedom arises because we are using the length of the response vector: a unitary transformation does not change the length of the response vector.

With respect to current theories, the most important consequence of the uniqueness result is this: the free unitary transformation does not preserve the frequency selectivity of the sensors. Thus we cannot use a model that is based on the vector length of the response sensors to draw conclusions about the frequency selectivity of the sensors in **S**. Ahumada and Watson[15] made this observation clearly. They pointed out that when the threshold is assumed to depend on the vector length of the response (what they call contrast energy), then sensor "... frequency selectivity plays no real role in detection..." (see Ref. 15, p. 1138).

On a more positive note, we find that it is possible to estimate the sensors up to a unitary transformation. This is a fairly restrictive uniqueness result. We summarize the uniqueness results in two parts. Sensitivity measurements coupled with the quadratic approximation strongly constrain the possible sensors; sensitivity measurements coupled with the quadratic approximation cannot be used to estimate the frequency selectivity of the sensors.

Measurement

The quadratic analysis that we have described emphasizes the measurement of visual performance. We have developed the measurement procedure in order to maintain a close coupling of measurement with the underlying sensor representation. The quadratic form **Q** is a multidimensional measure that can be estimated across many testing conditions. If this measure is to be useful, it must guide us in two ways.

First, the measure must guide us about the experiments that should be performed under different test conditions. The quadratic model suggests that thresholds to sinusoidal patterns and mixtures of two sinusoidal patterns (or any spatial-basis set) can be used as the fundamenal empirical measures to predict an observer's complete spatial-pattern sensitivity. These measurements are all that we need to estimate the parameters of **Q**, and from this characterization we can predict thresholds to other spatial patterns. By

extending the threshold measurements to include patterns that are the mixture of more than two sinusoids (e.g., lines, edges), we can test the adequacy of the quadratic estimate of performance. If the quantitative departures are not greater than the currently demonstrated departures (see Fig. 6), the quadratic form can serve as a useful tool for many applications.

Second, the measure must provide a method of comparing pattern sensitivities for different viewing conditions. The estimated values of $\mathbf{Q}$, as a function of the viewing conditions, provide a measure that can be compared across viewing conditions. For example, consider the effect of visual masking on spatial-pattern sensitivity. For each masking condition, M_i, we can measure test-mixture thresholds that define a quadratic form, $\mathbf{Q}(M_i)$. As the masking conditions are varied, the quadratic form will vary. We can formulate hypotheses about the dependence of the quadratic form on the contrast of the masking conditions. The estimated quadratic form will be related to estimates of the underlying linear encoding of the visual system. In this way, we can ask whether the underlying sensors vary as a function of mask conditions or whether they are the same up to a unitary transformation.

4. CONCLUSIONS

To some readers it may seem odd that we have analyzed pattern sensitivity but neglected the three other fundamental types of measurement, masking, adaptation, and physiology, that are generally cited in support of the multiple-channels spatial-frequency characterization of the initial sensors. We have the conservative view that the identity of the neural mechanisms that govern these various behavioral and physiological measures has not been proved. It is important to show, not to assume, that the limits of spatial sensitivity are determined by the same mechanisms that mediate the loss of sensitivity from masking phenomena and adaptation. We must find empirical measures that characterize the spatial organization of the mechanisms that mediate masking and adaptation effects. At present there are no agreed-on measures that permit us to define meaningful spatial-sensitivity curves for masking and adapting targets. The difficulties in obtaining empirical estimates were demonstrated 10 years ago in adaptation studies[41,42] and masking studies.[43,44] Masking stimuli whose spatial-frequency energy is more than 2 octaves away from the spatial-frequency energy in the test can still have substantial effects on test visibility. Worse yet, the effects of masking stimuli used in combination cannot be predicted readily from the effects of mask stimuli used singly.

Finally, why do we not make our psychophysical arguments by using physiological evidence[45,46] that individual neurons in monkey cortex are often spatially band limited? Psychophysics must have a theory of spatial sensitivity that stands on its own. As a complete analysis of cortical performance emerges, we must be prepared to understand the connection between independently obtained behavioral measures and neural recordings. The careful development of independent psychophysical measurement procedures offers a surer route to building a close connection between the physiological neuronal measures and psychophysical performance measures.

Although the quadratic model of visual performance is inexact, it is premature to abandon it as an efficient parameterization of spatial sensitivity. On the basis of currently available measurements, the quadratic summary of performance is sufficiently precise to make it useful as a summary of additional pattern-sensitivity measurements.

It is possible as well that the quadratic measure is close enough to the true neural underpinnings that it can be used to guide our reasoning about the underlying sensor sensitivities mediating threshold performance. The quadratic measure is designed so that the relationships among the linear encoding, the stimuli selected to measure that coding, and the measure $\mathbf{Q}$ are on a firm logical footing. The analysis that we describe here allows us to state precisely what can and what cannot be inferred from performance measurements. The parameters in the estimated $\mathbf{Q}$ matrix can be tied directly to performance measures, on the one side, and can be placed within a unitary transformation of the sensor construct, on the other side. We believe that the quadratic characterization of visual performance across different viewing conditions will be useful for measuring and, we hope, for ultimately understanding visual performance.

APPENDIX A

It is possible to determine the null space of a matrix by computing the SVD of the initial matrix.[47,48] The SVD is a matrix factorization that yields three unique matrices, e.g., $\mathbf{S} = UDV^t$. Each of the matrices has a useful and significant interpretation. The matrix V is called the right singular matrix, and its column vectors are called the right singular vectors. (Notice that it is the transpose of this matrix that is used in the SVD and that we refer here to rows of the transpose.) In our application the right singular matrix is $N \times N$, where N is the number of degrees of freedom in the stimulus. The SVD factorization guarantees that the columns of V are orthogonal and have unit length. This has the immediate consequence that $VV^t = V^tV = I$. A matrix with this property is called a unitary matrix. Rigid rotations are examples of unitary matrices. Unitary transformations preserve many properties of the input vectors, including the vector length and the angle between the vectors. The effect of multiplying the stimulus vector into this matrix is to represent the stimulus with respect to the (orthonormal) basis defined by the columns of V.

The matrix $\mathbf{D}$ is $P \times N$. All its nonzero terms fall along the diagonal, and when the original matrix is not square, that is, when $P \neq N$, there will be a block region of the matrix of all zero entries. The SVD factorization is arranged so that the diagonal elements occur in descending order. The action of the matrix $\mathbf{D}$ is to scale each of the dimensions of the representation defined by V.

When the number of stimulus dimensions N exceeds the number of sensors P, the projection of the stimulus onto the rows of V^t numbered higher than P contributes zero to the resulting vector. Any stimulus that projects only onto these dimensions will be mapped entirely to zero; that is, such a stimulus is in the null space of the model. It follows that the column vectors of V beyond the Pth column form an orthogonal basis for the null space of the model.

The $\mathbf{S}$ matrix can be replaced by an equivalent matrix without any effect on the model's predictions by constructing a new matrix generated by eliminating the null space. The procedure for estimating $\mathbf{Q}$ that is described in Section

3 also permits us to estimate the minimal number of sensors required for the sensor matrix. Since the matrix **Q** arises from the product of the sensor matrix and its transpose, $\mathbf{S}^t\mathbf{S}$, it follows that the diagonal matrix derived from the SVD of **Q** is simply the square of the diagonal matrix derived from the SVD of **S**. From the estimate of **Q**, we can estimate the number of nonzero entries of the diagonal matrix of **S**. The number of nonzero entries in the diagonal matrix is equal to the number of linearly independent sensors required in the **S** matrix. Although we are not certain whether such a project would be practical, in principle it is possible to estimate the **Q** matrix at a small region of the retina and then to compute the dimensionality of the estimated matrix. If the data are well fitted by the quadratic model, then the dimensionality of the **Q** matrix is an estimate of the number of independent sensors at that location in the retina.

The final $P \times P$ matrix U is called the left singular matrix. It also consists of orthonormal column vectors and is thus a unitary matrix. The columns of this matrix provide an orthonormal basis for the sensor-response space: any sensor vector can be represented as the weighted sum of these orthonormal columns. If the input images to the model are generated by a Gaussian noise process, and if the sensors have a mean response of zero to Gaussian noise, then the column vectors are the principal components of the sensor responses generated by the Gaussian noise. The variance of the weights associated with each of the columns of U is equal to the square of the corresponding entry of the diagonal matrix D. Because these terms are ordered, the first principal component is said to explain a larger share of the variance of the sensor response than the second, the second is said to explain a larger share than the third, and so forth.

ACKNOWLEDGMENTS

We thank A. Ahumada, D. Brainard, N. Graham, J. Nachmias, and A. B. Watson for many insightful discussions and their comments on this manuscript. We thank Hugh Wilson for providing us with the details necessary to implement the Wilson–Gelb model. A preliminary report of this research was presented at the 1986 Annual Meeting of the Association for Research in Vision and Ophthalmology. Funding was provided in part by grant NC2-307 from the NASA Ames Research Center to Stanford University and by National Eye Institute grant no. EY03164.

The authors are also affiliated with the Perception and Cognition Group, N239-3 NASA Ames Research Center, Moffett Field, California 94035.

REFERENCES AND NOTES

1. O. H. Schade, "Optical and photoelectric analog of the eye," J. Opt. Soc. Am. **46**, 721–739 (1956).
2. H. deLange, "Relationship between critical flicker frequency and a set of low frequency characteristics of the eye," J. Opt. Soc. Am. **44**, 380–389 (1954).
3. H. deLange, "Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. II. Phase shift in brightness and delay in color perception," J. Opt. Soc. Am. **48**, 784–789 (1958).
4. H. deLange, "Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light," J. Opt. Soc. Am. **48**, 777–784 (1958).
5. F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," J. Physiol. London **197**, 551–566 (1968).
6. L. A. Olzak and J. P. Thomas, "Seeing spatial patterns," in *Handbook of Perception and Human Performance*, J. P. Thomas, ed. (Wiley, New York, 1986), pp. 7-1, 7-53.
7. H. R. Wilson and D. J. Gelb, "Modified line-element theory for spatial-frequency and width discrimination," J. Opt. Soc. Am. A **1**, 124–131 (1984).
8. H. R. Wilson and D. Regan, "Spatial-frequency adaptation and grating discrimination: predictions of a line-element model," J. Opt. Soc. Am. A **1**, 1091–1096 (1984).
9. A. B. Watson, "Detection and recognition of simple spatial forms," in *Physical and Biological Processing of Images*, A. C. Slade, ed. (Springer-Verlag, Berlin, 1983), pp. 100–114.
10. S. A. Klein and D. M. Levi, "Hyperacuity thresholds of 1 sec: theoretical predictions and empirical validation," J. Opt. Soc. Am. A **2**, 1170–1190 (1985).
11. D. Marr, *Vision* (Freeman, San Francisco, 1982).
12. L. T. Maloney and B. A. Wandell, "A model of a single visual channel's response to weak test lights," Vision Res. **24**, 633–640 (1984).
13. R. F. Quick, "A vector magnitude model of contrast detection," Kybernetik **16**, 65–67 (1974).
14. M. B. Sachs, J. Nachmias, and J. G. Robson, "Spatial frequency detectors in human vision," J. Opt. Soc. Am. **61**, 1176–1186 (1971).
15. A. J. Ahumada, Jr., and A. B. Watson, "Equivalent-noise model for contrast detection and discrimination," J. Opt. Soc. Am. A **2**, 1133–1139 (1985).
16. W. S. Geisler, "Ideal observer analysis of visual discrimination," in *Frontiers of Visual Science*, The Committee on Vision, National Academy of Sciences, ed. (National Academy Press, Washington, D.C., 1987), pp. 17–31.
17. A. B. Watson, "The ideal observer concept as a modeling tool," in *Frontiers of Visual Science*, The Committee on Vision, National Academy of Sciences, ed. (National Academy Press, Washington, D.C., 1987), pp. 32–37.
18. A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," Science **214**, 93–94 (1981).
19. A. B. Watson, H. B. Barlow, and J. G. Robson, "What does the eye see best?" Nature **302**, 419–422 (1983).
20. The sensors used by Wilson and his colleagues have evolved over the years. Thus the set of sensors used in the computations by Wilson and Bergen [H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," Vision Res. **19**, 19–32 (1979)] are different from that used by Wilson and Gelb[7]; both of these sets are different from that used by Wilson and Regan.[8]
21. In calculations for which the stimulus is represented before the optical blurring, a higher sampling frequency may be required to represent stimuli adequately. This may occur when sensor characterizations include both the ocular optics and the neural sensors. In our implementation of the Wilson–Gelb calculations, we used a spatial resolution of 120 samples per degree. Our simulations confirmed all their calculations.
22. K. R. K. Nielsen, A. B. Watson, and A. J. Ahumada, Jr., "Application of a computable model of human spatial vision to phase discrimination," J. Opt. Soc. Am. A **2**, 1600–1606 (1985).
23. The Wilson–Gelb sensor set[7] is not consistent, in this sense, with the sensor set described in Ref. 8.
24. In some simulations, the effects of noise are modeled by adding a random variable to each of the sensor responses. If noise is added to each sensor output, then we can never be sure that any sensor response is always zero but only that its mean response is zero. Such a sensor contributes only noise. When the total number of sensors is reduced by incorporating stimulus characteristics into the **S** matrix, then the properties of the noise added to the remaining sensors may have to be adjusted to include the effects of noise that have been deleted from the sensors with a zero-mean response.

25. W. K. Pratt, *Digital Image Processing* (Wiley, New York, 1978).
26. A. Ahumada, Jr., "Aliasing predictions from a spin-glass model for assigning cones to color sensitivity classes," Invest. Opthalmol. **28,** 137 (1987).
27. J. I. Yellott, "Spectral analysis of spatial sampling by photoreceptors: topological disorder prevents aliasing," Vision Res. **22,** 1205–1210 (1982).
28. J. I. Yellott, B. A. Wandell, and T. N. Cornsweet, "The beginnings of visual perception: the retinal image and its initial encoding," in *Handbook of Physiology: The Nervous System,* I. Darien-Smith, ed. (Easton, New York, 1984), Vol. 1, pp. 257–316.
29. These methods of selecting sensors deserve a careful review, but doing so would take us too far afield from the present analysis.
30. N. Graham, J. G. Robson, and J. Nachmias, "Grating summation of fovea and periphery," Vision Res. **18,** 815–825 (1978).
31. N. Graham and J. Nachmias, "Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models," Vision Res. **11,** 251–259 (1971).
32. C. Rashbass, "The visibility of transient changes of luminance," J. Physiol. London **210,** 165–186 (1970).
33. There will be a correspondingly simple structure in the general 2D case as well. A proper discussion of the structure of the resulting matrix, however, should take into account whether the 2D transformation is separable. The added complexity would take us far afield from our main point. The reader may consult various standard references [e.g., A. Rosenfeld and A. C. Kak, *Digital Picture Processing* (Academic, New York, 1982), Vol. 1, p. 137; W. K. Pratt, *Digital Image Processing* (Wiley, New York, 1978), p. 203] for a review of some of the issues concerning the structure in 2D linear operators and their associated quadratic forms.
34. Graham *et al.*[30] report their data in units of contrast normalized so that the threshold to each of the components alone is set near 1. Since there are four measurements along the axes and there are only two scale parameters, it is not possible for all the data along the axes to be scaled to 1.
35. J. P. Chandler, "STEPIT," *Quantum Chemistry Program Exchange* (Department of Chemistry, Indiana University, Bloomington, Ind., 1965).
36. B. A. Wandell, "Color measurement and discrimination," J. Opt. Soc. Am. A **2,** 62–71 (1985).
37. G. Wyszecki and W. S. Stiles, *Color Science* (Wiley, New York, 1982).
38. M. A. Berkley, F. Kitterle, and D. W. Watkins, "Grating visibility as a function of orientation and retinal eccentricity," Vision Res. **15,** 239–244 (1975).
39. D. H. Kelly, "Retinal inhomogeneity. I. Spatiotemporal contrast sensitivity," J. Opt. Soc. Am. A **1,** 107–113 (1984).
40. D. H. Kelly, "Retinal inhomogeneity. II. Spatiotemporal summation," J. Opt. Soc. Am. A **1,** 107–113 (1984).
41. J. Nachmias, R. Sansbury, A. Vassilev, and A. Weber,"Adaptation to square-wave gratings: in search of the elusive third harmonic," Vision Res. **13,** 1335–1342 (1973).
42. M. W. Greenlee and S. Magnussen, "Higher-harmonic adaptation and the detection of squarewave gratings," Vision Res. **27,** 249–255 (1987).
43. G. B. Henning, B. G. Hertz, and D. E. Broadbent, "Some experiments bearing on the hypothesis that the visual system analyzes patterns in independent bands of spatial frequency," Vision Res. **15,** 887–899 (1975).
44. J. Nachmias and B. E. Rogowitz, "Masking by spatially-modulated gratings," Vision Res. **23,** 1621–1630 (1983).
45. J. A. Movshon, I. D. Thompson, and D. J. Tolhurst, "Receptive field organization of complex cells in the cat's striate cortex," J. Physiol. **283,** 79–99 (1978).
46. R. L. DeValois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in the macaque visual cortex," Vision Res. **22,** 545–559 (1982).
47. G. H. Golub and C. F. van Loan, *Matrix Computations* (Johns Hopkins U. Press, Baltimore, Md., 1983).
48. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK Users Guide* (SIAM Publications, Philadelphia, Pa., 1978).