

# Temporal Encoding of Two-Dimensional Patterns by Single Units in Primate Inferior Temporal Cortex.

## III. Information Theoretic Analysis

LANCE M. OPTICAN AND BARRY J. RICHMOND

*Laboratory of Sensorimotor Research, National Eye Institute,  
and Laboratory of Neuropsychology, National Institute of Mental Health,  
Bethesda, Maryland 20892*

### SUMMARY AND CONCLUSIONS

1. Ablation and single-unit studies in primates have shown that inferior temporal (IT) cortex is important for pattern discrimination. The first paper in this series suggested that single units in IT cortex of alert monkeys respond to a set of two-dimensional patterns with complex temporal modulation of their spike trains. The second paper quantified the waveform of the modulated responses of IT neurons with principal components and demonstrated that the coefficients of two to four of the principal components were stimulus dependent. Although the coefficients of the principal components are uncorrelated, it is possible that they are not statistically independent. That is, several coefficients could be determined by the same feature of the stimulus, and thus could be conveying the same information. The final part of this study examined this issue by comparing the amount of information about the stimulus that can be conveyed by two codes: a temporal waveform code derived from the coefficients of the first three principal components and a mean rate code derived from the spike count.

2. We considered the neuron to be an information channel conveying messages about stimulus parameters. Previous applications of information theory to neurophysiology have dealt either with the theoretical capacity of neuronal channels or the temporal distribution of information within the spike train. This previous work usually used a general binary code to represent the spike train of a neuron's

response. Such a general approach yields no indication of the nature of the neuron's intrinsic coding scheme because it depends only on the timing of spikes in the response. In particular, it is independent of any statistical properties of the responses. Our approach uses the principal components of the response waveform to derive a code for representing information about the stimuli. We regard this code as an indication of the neuron's intrinsic coding scheme, because it is based on the statistical properties of the neuronal responses.

3. We measured how much information about the stimulus was present in the neuron's responses. This transmitted information was calculated for codes based on either the spike count or on the first three principal components of the response waveform. The information transmitted by each of the first three principal components was largely independent of that transmitted by the others. It was found that the average amount of information transmitted by the principal components was about twice as large as that transmitted by the spike count. The two codes also gave different relative amounts of transmitted information for individual patterns. For example, the pattern about which the most information was transmitted differed for the two codes.

4. When transmitted information calculated from a code was directly compared with another parameter of the response, its magnitude, they were only weakly correlated. The absolute values of the correlation coefficients ranged from 0.01 to 0.61. This weak correlation reflects the fact that stimuli eliciting the

largest numbers of spikes do not necessarily correspond to those stimuli about which the most information is being transmitted.

5. Previous studies have ascribed only univariate tuning properties to neurons in the visual system. We infer from our results a new hypothesis with implications for processing throughout the visual system. Neurons may be encoding multidimensional spatial properties of visual stimuli, with multivariate temporal modulation as the information code. We refer to this as the multiplex-filter hypothesis. The multivariate approach used in these studies can be extended to analyze neuronal responses to multiple physical and behavioral stimulus parameters.

## INTRODUCTION

In the search for a general description of the stimulus-response relationships of inferior temporal (IT) neurons, we found that visual stimuli could elicit different patterns of temporal modulation of the spike train (27). In this paper we show that multiple parameters were required to completely represent the stimulus-dependent part of this modulation. These results lead to the inference that neurons represent visual features in the temporal modulation of their responses.

A relationship between stimulus features and multiple response parameters can be established with Shannon's information theory (1, 15, 33). We considered the neuron as a communication channel transmitting messages (encoded in the temporal modulation of the response) about stimulus parameters (chosen from a complete set of orthogonal two-dimensional Walsh patterns). This enabled us to use the mathematical tools of information theory to quantify the information transmitted about the stimulus by the neuron.

Information theory deals with the probabilistic relationship between input and output events. These events are signaled by encoding them onto some information carrier, such as the spike train. Earlier studies of information processing in the visual system have made assumptions about what codes were being used by the nervous system, but the intrinsic neural code is not yet known (6). Perhaps the most commonly assumed form for the neural code is a mean rate, or spike count, code. Another possibility is a general binary code, which de-

pends only on the timing of spikes in the response. Several previous studies have compared information transmitted in spike count and general binary codes and have shown that the amount of information conveyed in the spike train is underestimated by the use of a spike count code (9, 14).

The general binary code depends only on the times of spike occurrences, and hence does not reflect any intrinsic limitations on spike train generation. The code thus has a very large number of degrees of freedom, which allows it to represent the amount of information transmitted without giving any indication of how the message was encoded. Our results, however, have shown that only two to four parameters are needed to represent the stimulus-dependent aspects of a neuron's responses (28). From this limited number of degrees of freedom we infer that an intrinsic neural code that reflects constraints on the spike generation or information processing mechanisms must exist. The advantage of such a code is that it will both represent the information transmitted in the spike train and provide insight into the mechanism of encoding messages.

In this paper we investigate the ability of a multiple parameter code, derived from the statistical properties of a neuron's responses, to convey information about stimulus features. We performed our information analysis on a temporal response code formed either from the spike count or from the first three principal components of the neuron's responses (28). Our results are consistent with those of earlier investigators who studied other areas of the brain: the amount of information transmitted in the temporal modulation of the response is at least twice that transmitted by the spike count (8, 10, 14). In addition, since the information transmitted by different (uncorrelated) principal components is largely independent, we can conclude that multiple stimulus features are encoded in the multidimensional temporal modulation of neuronal response. Our results provide some insight into the intrinsic method of information transmission used by IT neurons and may lead to similar insights in other visual areas.

## METHODOLOGY

This paper continues the analysis of the data described in the first two studies in this series (27, 28).

The methods for training and preparing the monkey, presenting the stimuli, and collecting the spike train data were described in the first paper of this series (27). The stimuli were members of a complete set of orthogonal patterns based on two-dimensional Walsh functions. The patterns have a resolution of one part in eight in the horizontal and vertical directions, making a total of 64 Walsh functions. The 64 contrast reversed Walsh functions were also used with many of the neurons, making a total of 128 stimuli.

The purpose of this final study in the series was to compare the ability of codes based on either the spike count or the temporal modulation of the response to represent the two-dimensional stimuli. Hence, in this study we needed to calculate information transmission in neurons from stimulus-response data. Information theory quantifies the amount of information transmitted through a communication channel (1, 15, 34). By comparing the amount of uncertainty that exists about a message (selected from a finite set of messages) before it is sent with the amount of uncertainty about the message after a signal is received (perhaps corrupted by noise), we can quantify the amount of information transmitted about the message through the channel. It is also possible to calculate a related measure, the probability of a transmission error over the channel assuming a specific detector.

### Information measures

The average uncertainty about an event is called the entropy. The reduction in entropy is called the gain in information. Since entropy is defined by the probabilities of events, any information measures based on entropies will necessarily include the effects of the probability distribution of those events. We will deal with discrete probabilities to make the computation of information easier. The discrete set of all stimuli is indicated by "S", and a member of that set by "s". The (marginal) probability of a stimulus occurring in our experiment is just  $P(s) = 1/128$ . Similarly, "R" is the set of all possible responses, and "r" is an individual response (see below). The entropy of the stimulus source,  $H(S)$ , is defined as

$$H(S) = - \sum_s P(s) \log P(s) \quad (1)$$

The measure of information is thus logarithmic, and the choice of the base of the logarithm determines the units of information. We will always use the base two logarithm, giving information the unit of "bits". There were 128 stimuli in our experiment; if they were all equally likely then the entropy of the source would be seven bits. If the message sent and the signal received were statistically independent, then one would still have seven bits of uncertainty left after receiving the signal. Alternatively, if there were a noiseless relation between the signal

and the message sent, then there would be no uncertainty, or zero bits, left after receipt of the signal. Between these two extremes, there would be a probabilistic relationship between message and signal, and the signal would tell something about the message sent, but some residual uncertainty would remain. For example, the response might indicate only that the stimulus could have been any one of, say, eight patterns. This would leave a residual uncertainty of three bits ( $\log 8 = 3$ ). Hence the information gain from receipt of the message would be four bits ( $7 - 3 = 4$  bits).

The amount of information transmitted by the neuron about the particular stimulus  $s$ , averaged over all the responses in the set R, (also called the transinformation) is

$$T(s; R) = - \sum_R P(r|s) \log \frac{P(r|s)}{P(r)} \quad (2)$$

where  $P(r|s)$  is the conditional probability of getting response  $r$  given stimulus  $s$ . The summation is over all the members of the set R. For the spike count code this is  $M$ , the number of bins (see below). For an  $n$ -dimensional temporal code based on  $n$  principal components, the number of members of the set R is  $M^n$ . The conditional probability may be calculated using Bayes' law

$$P(r|s) = \frac{P(s, r)}{P(s)} \quad (3)$$

where  $P(s, r)$  is the joint probability density function of any stimulus-response pair (see below).

The average transmitted information can then be calculated as

$$T(S; R) = \sum_s P(s) T(s; R) \quad (4)$$

### Probability of error

A measure of the performance of the neuron as an information channel is given by the probability of error,  $P_E$ , in determining which stimulus,  $s$ , gave rise to the response,  $r$ .  $P_E$  is calculated directly from  $P(s|r)$ , which is obtained from  $P(s, r)$  by Bayes' law as  $P(s, r)/P(r)$ . The value of  $P_E$  is not an invariant of the communication channel, but depends on the type of detector used to relate a specific response to a specific stimulus. The simplest detector is the maximum likelihood detector, whereby the choice of  $s$  made after receipt of  $r$  is just that stimulus with the largest probability of having given rise to the received response (maximum value of  $P(s|r)$  over all values of  $s$ ). The maximum  $P(s|r)$  over the set S is the probability of being correct, and so the probability of a mistake is  $1 - P(s|r)_{\max}$ . Summing the probability of a mistake over all the responses in the set R gives the total probability of error,  $P_E$ , for that neuron (with a maximum likelihood detector). Another detector, incorporating some a priori knowledge of the neuron's responses, would yield

smaller values of  $P_E$ . Hence the use of  $P_E$  based on a maximum likelihood detector gives a conservative estimate of errors without introducing arbitrary assumptions about intrinsic detector properties.

### Probability functions

These equations all depend on the joint probability density function (pdf). The joint pdf is the probability of the occurrence of all stimulus-response pairs during the experiment. To make these calculations on neural data we first form discrete stimulus and response codes. Since the stimuli consist of 128 orthogonal patterns, the stimulus code can be taken as the 128 discrete events representing the presentation of one of the patterns.

We define the response of the neuron as the average over the number of stimulus presentations of either the number of spikes per trial or of the multidimensional coefficient vector of the Karhunen-Loeve (KL) transform (28). These continuous variables were quantized into discrete sets to allow a discrete response code to be used. The method of quantization must be carefully considered (15, 30). The calculations needed to apply information theory to our study of neurons are based on entropy functions, which depend on the joint probability density function,  $P(s, r)$ , of all stimulus-response pairs (1, 2, 17). The purpose of quantizing the response values is to provide an efficient means of estimating the joint probability density function. If the number of trials per stimulus was large, it would be sufficient to divide the response range into bins and calculate the histogram of stimulus-response pairs and use that to estimate  $P(s, r)$ . For small numbers of trials per stimulus, however, this method is sensitive to edge effects created by the finite number of bins in the histogram. These effects give biased estimates of information measures (12).

We avoided the problems arising from the use of a discrete estimate of the pdf by actually forming a continuous estimate of the joint pdf and then quantizing that estimate. A method for estimating continuous probability density functions has been developed by Parzen (22, 24), and we followed a standard extension of that method to multivariate data (13). A Parzen estimate of a pdf is formed by convolving a function, called the kernel of the estimate, with the data. Parzen established conditions on the kernels so that the estimate would be asymptotically unbiased and uniformly convergent. Let  $x$  be the value of the response variable (a scalar for spike count, a multidimensional vector for the KL transform). The estimate of the multivariate density based on  $N$  data points is

$$\hat{P}_N(x) = \frac{1}{N} \sum_{i=1}^N g(x, x_i, h) \quad (5)$$

where the  $x_i$  are the data points,  $N$  is the number

of data points, and  $g(\ )$  is the multivariate Gaussian function

$$h^{-n}(2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left[\frac{-1}{kh^2}(x-x_i)\Sigma^{-1}(x-x_i)\right] \quad (6)$$

where  $\Sigma$  is the sample covariance matrix,  $|\Sigma|$  is its determinant,  $n$  is the number of dimensions (i.e., 1 for spike count, 3 for first three principal components),  $k$  is a constant to be determined, and  $h$  is a function of  $N$

$$h(N) = N^{-0.49/n} \quad (7)$$

where  $n$  is the number of dimensions in the response (i.e.,  $n = 1$  for spike count,  $n = 3$  for temporal code). Since the coefficients of the principal components obtained from the KL transform are uncorrelated, the covariance matrix ( $\Sigma$ ) is diagonal, which greatly simplifies these calculations. This kernel depends on the sample covariance matrix, so the statistical properties of the data, up to the second moment, are taken into account even when  $N$  is small (13). One parameter,  $k$ , is still left to be determined. The parameter  $k$  acts as a scale factor: if it is too large, the kernel will be spread out too much, leading to underestimates of the information measures, and if it is too small the kernel will be too concentrated, leading to overestimates. The nominal value of  $k$  for a Gaussian function is two. We calculated information measures based on Eq. 6 using different values of  $k$ . Figure 1A shows how the information (T) in one or two principal components depends on the value of  $k$ . Figure 1B shows how the ratio of the information in one of the principal components to the information in two principal components depends on  $k$ . Values for  $k$  less than two yield overestimates of information calculated from more than one principal component, whereas values much larger than two begin to underestimate information. In our calculations we used the value of two for  $k$  as a reasonable compromise.

Once the continuous Parzen estimate of the joint pdf is formed, it must be quantized. The quantization was done by dividing the range of the data values into equally spaced bins, and then adding an extra bin on each end to hold the "tails" of the probability function. The number of bins,  $M$ , must be large enough to avoid misrepresenting the Parzen estimate, yet it must be small to reduce the computational burden. Figure 1C shows the value of information calculated with different numbers of bins based on one or two principal components. For small values of  $M$  information is severely underestimated because many responses become indistinguishable. For large values of  $M$  the curve in Fig. 1C approaches  $\log(M)$  asymptotically (30). Hence there is no correct choice for  $M$ . The choice of  $M = 12$  is sufficient to avoid seriously underestimating information without incurring the dimin-

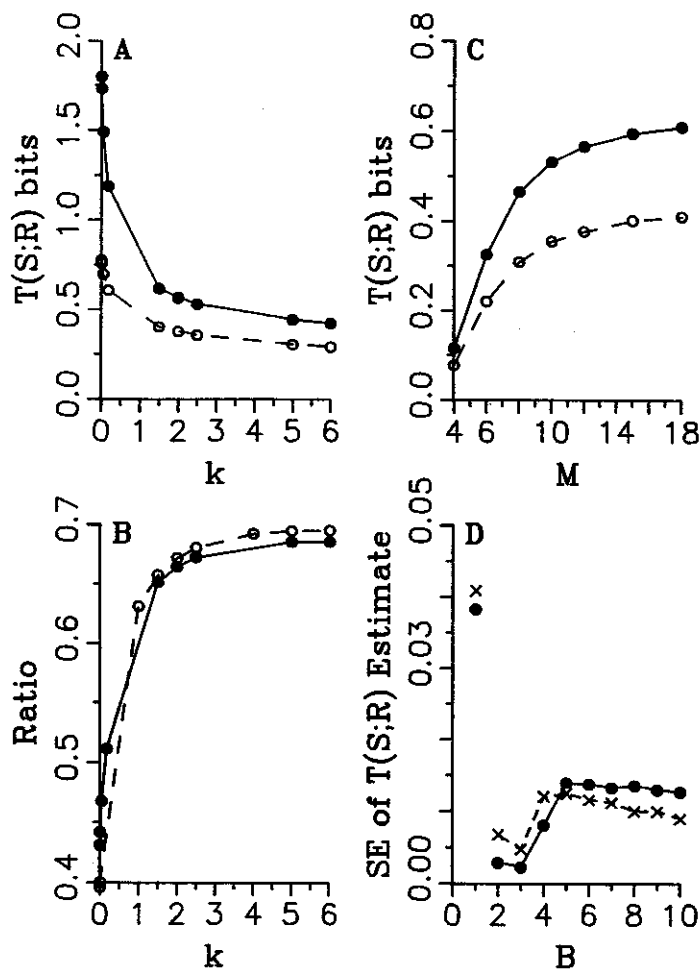


FIG. 1. Influence of scaling and quantizing on information. Calculations of information measures and variances depend on 3 arbitrary parameters,  $k$ ,  $M$ , and  $B$ . For practical reasons, the value of 2 was used for  $k$ , the value of 12 was used for  $M$ , and the value of 5 was used for  $B$ . *A*: the amount of information increases as the scale factor ( $k$ ) in the kernel of the probability estimate decreases the spread of the kernel. The information in one principal component (*open circles*) does not increase as much as the information in the first two components (*filled circles*). *B*: the ratio of the information in the first component to that in the first two components is shown for 12 bins (*filled circles*) and 18 bins (*open circles*). For values below two, the ratio is severely underestimated. *C*: the amount of information for 1 (*open circles*) or 2 (*filled circles*) principal components, as a function of the number of bins ( $M$ ) used to quantize each dimension of the probability function. For values less than  $\sim 10$ , the information falls off very rapidly. For values above 12, the curve approaches an asymptote of  $\log(M)$ , requiring more and more computation for small gains in the amount of information. *D*: calculation of the standard error of the estimate of transmitted information. *Filled circles* are data from unit PT58, and *crosses* are data from unit PT179. Closed form solutions (shown at  $B = 1$ ) give large values of the standard errors (see text and appendix). The bootstrap method gives reasonably consistent estimates as long as  $B$ , the number of bootstrap resamplings, is  $>4$ .

ishing returns of small increments in information for much longer computation times.

### Defining a code

A code is defined by naming the bins of this joint pdf. A response value is mapped onto this code by determining into which bin it falls. For example, in the spike count code every response falls into one of  $M$  bins, and in the three parameter temporal code every response falls into one of  $M^3$  bins. Once the response is assigned to a bin, it receives a name, or code symbol, such as a, b, or g for the univariate

spike count, or aaa, aab, or cgd for the trivariate temporal waveform. It must be emphasized that once converted to a code symbol, any other property of the response, in particular the magnitude, or strength, of the response is lost. This dissociation of properties and names is quite common, e.g., in the alphabet we do not associate more strength with one letter than another, even though there is an alphabetical order.

With the number of bins ( $M$ ) and the scale factor ( $k$ ) fixed, it is straightforward to fill in the multi-dimensional table, which estimates the joint prob-

ability density function,  $P(s, r)$ . Each bin in the table holds the probability that the data value is between the left and right edges of the bin. Since the Gaussian kernel chosen for the Parzen estimate is separable (i.e., the multidimensional function is the product of the unidimensional functions), we need only calculate the bin values for each dimension independently and then multiply them together to get the multidimensional bin contents. Suppose that the number of bins is  $M$ , the smallest data value is  $x_0$ , and the largest data value is  $x_1$ . The width of the bins,  $w$ , is

$$w = \frac{x_1 - x_0}{M - 2} \quad (8)$$

and the value of the right edge of the bin on the left is

$$x_0 - \frac{w}{2} \quad (9)$$

The left edge of the lowest bin is taken as  $-\infty$ , and the right edge of the highest bin is taken as  $+\infty$ . Let  $F(x)$  be the cumulative multidimensional distribution function of our Gaussian kernel

$$F(x) = \text{Prob}\{x \leq X\} \quad (10)$$

The bins are filled by taking each data value,  $x_i$ , as the center of the Gaussian kernel and adding into each bin the probability that the data point lies within the bin

$$F(x_r) - F(x_l) \quad (11)$$

where  $x_l$  and  $x_r$  are the values of the left and right edges of a bin.

After all of the data points have been accumulated the matrix is a discrete estimate of  $P(s, r)$ , the continuous joint probability density function of the stimulus-response pairs. Now marginal probabilities can be calculated

$$P(s) = \sum_R P(s, r) \quad (12)$$

and

$$P(r) = \sum_S P(s, r) \quad (13)$$

Since the purpose of this study was to determine whether the multidimensional measure of neuronal response differs from the scalar measure, any difference found with more than one principal component would suffice. Ideally, one would want to compare the one-dimensional measure of spike count with the multidimensional measure of the temporal pattern formed by the vector of all KL transform coefficients that were not randomly distributed. However, the amount of information transmitted by several principal components cannot be obtained simply by adding the transmitted information calculated for each component individually. Simple addition fails because some of the same information may be transmitted by more than one component. Hence, the transmitted informa-

tion for a multiple component code must be computed explicitly (1, 21).

Due to limitations on the size of the probability matrix easily handled on our PDP-11/44 computer, as well as on the speed of computation, we have simplified the computational burden by using a truncated measure of the temporal pattern formed from just the first three principal components. Because of the nature of the KL transform, this truncation is optimal for preserving waveform in the least-square-error sense (3, 13). Because of this truncation the information values computed in this paper should be considered conservative estimates of neuronal performance. The truncation is analogous to low-pass filtering, or smoothing, the response.

### *Bias and variance of information measures*

The calculation of the mean probability and information measures is straightforward, and derivations of these equations are available in standard texts (1, 13). Estimation of the variances of these values, however, is not simple. When the pdf is calculated from contingency tables (i.e., when symbols are assigned to bins, as in a histogram) measures of the transmitted information are biased upward. Fagen (12) provides closed form estimates of the bias and variance for information computed from contingency tables. Corrections for bias suggested by Fagen depend on both the stimulus and response sets being discrete. While Sakitt (30) has proposed a method of converting continuous data to discrete data by constructing contingency tables, it is not general enough to be used with our data, since it depends on the stimulus set being rank ordered. Our approach has thus been to minimize bias by quantizing the Parzen estimate of the continuous joint probability density function, rather than attempting to calculate the pdf of quantized data (in contingency tables) and then applying bias corrections.

Our estimate of the variance of the information measures was calculated by the bootstrap method (11, 27). We calculated each information measure,  $\hat{Z}$ , after resampling the actual data set, with replacement, to construct a bootstrap data set. This operation was repeated  $B$  times, giving  $B$  estimates of the value of the information measure

$$\hat{Z}_b, b = 1, \dots, B \quad (14)$$

A bootstrap estimate of the measure itself is then

$$\hat{Z} = \frac{1}{B} \sum_{b=1}^B \hat{Z}_b \quad (15)$$

A bootstrap estimate of the variance of this estimate of the mean is then (11)

$$\text{var} = \frac{1}{B^2} \sum_{b=1}^B [\hat{Z}_b - \hat{Z}]^2 \quad (16)$$

Since the bootstrap technique depends on re-sampling with replacement, the reconstructed data sets will usually contain duplicates (triplicates, etc.) of some of the response values, while missing others. This will artificially bias the estimate of the information measure upward, because repeated values reduce the variance of sample estimates, and information is inversely related to variance. To avoid this bias we used as our best estimate of central tendency only the calculation based on the actual data set. The estimate of the variance, however, will be largely unaffected by this bias (since it is calculated from a difference:  $\hat{Z}_b - \hat{Z}$ ), and so we used the bootstrap estimate of variance to calculate the standard error of the mean for the information measures

$$SE = \sqrt{\text{var}} \quad (17)$$

The size of  $B$  was empirically determined. Figure 1D shows that the estimate of the standard error was fairly well established after five repetitions. The bootstrap estimate is much smaller than that obtained from a closed form expression (shown at 1.0 on the abscissa) similar to those used by Fagen (12) (see APPENDIX, Eq. A2). Because of the enormous computational burden of making each repetition, we used the minimum acceptable value ( $B = 5$ ).

## RESULTS

### Average transmitted information

The amount of information transmitted by the response about a stimulus, averaged over all the stimuli, is  $T(S; R)$ . This number can be calculated from  $P(s, r)$  where the response,  $r$ , is either the spike count or the temporal code. In 20 out of 21 neurons, the amount of information transmitted in the temporal code was higher than the amount transmitted in the spike count code. The mean values of the average transmitted information carried by the first three principal components was 0.9 bits, and the mean carried by the spike count was 0.4 bits (see Table 1). The values of information calculated from different combinations of principal components are shown in Table 1. Note that the amount of information carried by each component decreases as the component number increases (i.e.,  $T_0 > T_1 > T_2$ ). The amount of information conveyed by codes made up of more than one principal component can not be determined simply by adding up the information transmitted by the individual components (e.g.,  $T_{012} = 0.883$ , whereas  $T_0 + T_1 + T_2 = 1.039$ ). Complete statistical independence of the principal components

TABLE 1. *Information measures for cells in inferior temporal cortex*

	Mean	SE	Range
$P_{Es}$	0.963	0.004	0.914–0.980
$P_{E0}$	0.964	0.004	0.922–0.980
$P_{E1}$	0.967	0.003	0.934–0.980
$P_{E2}$	0.968	0.002	0.942–0.981
$P_{E01}$	0.946	0.005	0.884–0.973
$P_{E012}$	0.927	0.006	0.852–0.962
$T_s$	0.426	0.040	0.193–0.782
$T_0$	0.403	0.039	0.191–0.853
$T_1$	0.327	0.020	0.195–0.506
$T_2$	0.309	0.015	0.197–0.451
$T_{01}$	0.642	0.045	0.307–1.089
$T_{012}$	0.883	0.054	0.457–1.297
$ r_s $	0.185	0.037	0.006–0.606
$ r_{012} $	0.306	0.041	0.094–0.772
$T_{012}/T_s$	2.25	0.13	0.827–3.237

Measures are calculated for codes based on both the spike count (subscript  $s$ ) and the principal components of the temporal waveform of the response. Subscripts indicate which principal components were used, e.g., 0 is calculated from  $\phi_0$ , 01 is calculated from both  $\phi_0$  and  $\phi_1$  together, etc.  $P_E$  is the probability of error in transmitting information assuming a maximum likelihood detector.  $T(S; R)$  is the average amount of information transmitted by the cell about the stimulus. The average of the absolute value of the correlation coefficient between information ( $T_{012}(S; R)$ ) and spike count is  $|r_s|$  and between information and the length of the waveform vector is  $|r_{012}|$  ( $n = 21$ ).

would require strict additivity of the information contributed by each. Since the information is not strictly additive, there must be some cross-talk between components. As can be calculated from Table 1, on average there is only 13% redundancy between the first two and 18% redundancy among the first three principal components. Hence, the information conveyed by the principal components can be considered largely independent. Subsequent analyses deal with the information transmitted by a temporal code formed from the first three principal components,  $T_{012}$ , which will be abbreviated as  $T_3$ . Information transmitted by the spike count code will be abbreviated  $T_s$ .

Since information is a logarithmic measure, the number of independent neurons needed to transmit as much information through spike count as a single neuron transmits through the temporal code is  $T_3/T_s$ . The mean of this ratio for the neurons analyzed is 2.25 (see Table 1). Given that the average neuron in our population can encode about one bit, seven or eight

independent neurons from this population would be needed to completely encode our 128 stimuli in the temporal pattern of their responses represented by the first three principal components. If only the information available in the spike count were used, about 16 or 17 independent neurons would be needed.

#### *Transmitted information per stimulus*

The amount of information transmitted by the response of a neuron varies with the stimulus. Figure 2 shows the transmitted information per stimulus calculated from the temporal code (solid line) and from the spike count code (dashed line) for one of the IT neurons. The amount of information transmitted per stimulus for the temporal code,  $T_3(s; R)$ , is almost always larger than that for the spike count code,  $T_s(s; R)$ . Furthermore, the distribution of information across the stimulus patterns is not the same for both measures. In particular, the peaks in transmitted information from the temporal code frequently fail to correspond with peaks in transmitted information based on the spike count code.

The distribution of information in Fig. 2 appears to show some periodicity. The stim-

ulus patterns themselves are ordered in two-dimensional sequency, i.e., the rate of alternation of black and white patches (3). It is possible to replot the transmitted information on a two-dimensional grid of increasing sequency to see whether patterns that are similar in sequency have similar amounts of information transmitted about them, i.e., whether transmitted information is "tuned" for sequency. Figure 3 shows a projection of such a three-dimensional plot. The two axes labeled V (vertical sequency) and H (horizontal sequency) correspond to the vertical and horizontal axes of the grid of Walsh functions illustrated in Fig. 2 of the first paper in this series (27). Thus the lowest corner in the figure (near the axis label) corresponds to a white square in Fig. 3A and to a black square in Fig. 3B. The height of the function (T axis) is the amount of information transmitted by three principal components for that Walsh stimulus. From the chaotic nature of the graph we infer that no broadband tuning characteristic based on sequency exists for the IT neurons studied.

#### *Information and response magnitude*

It was pointed out in the METHODS section that information is calculated from the prob-

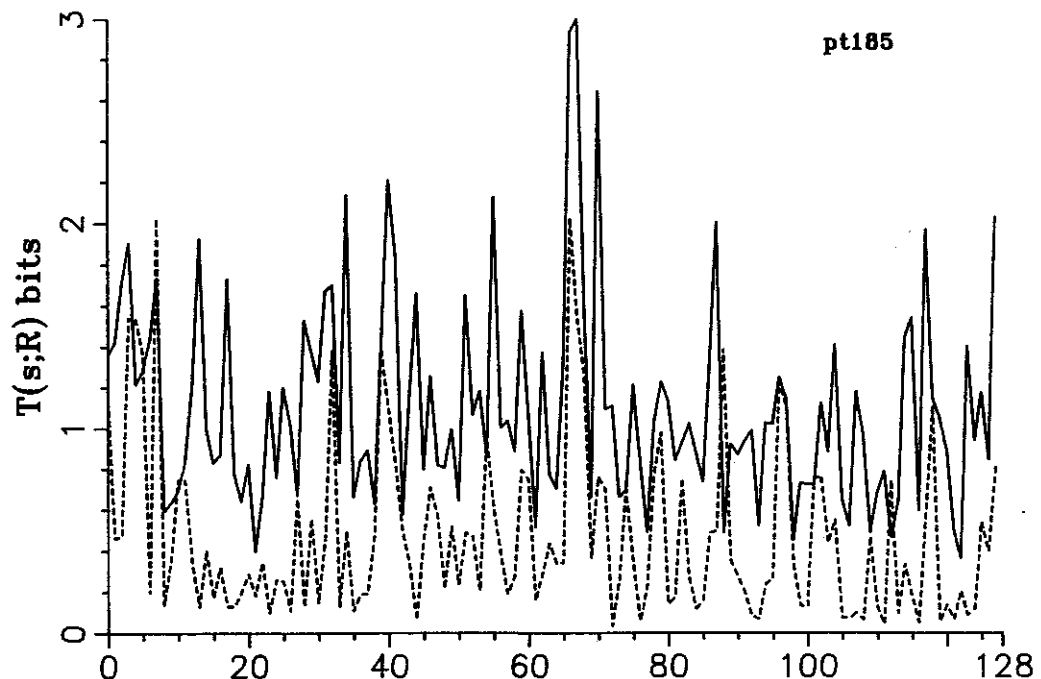
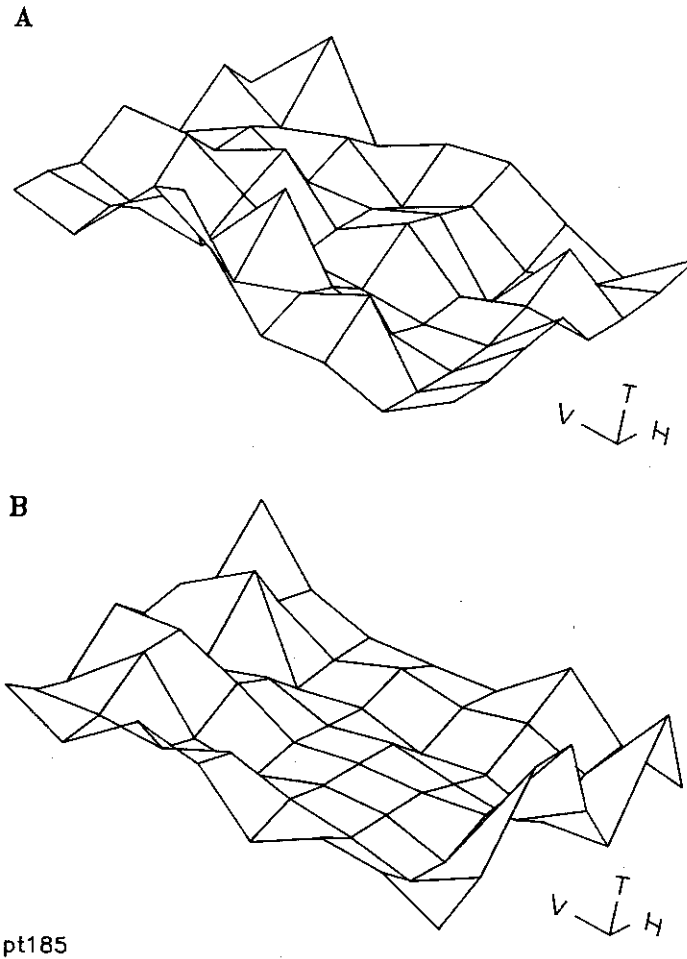


FIG. 2. Transmitted information per stimulus,  $T(s; R)$ , based on 3 principal components (*solid line*) and on the spike count (*dashed line*). Stimulus number corresponds to the serial ordering of the two-dimensional Walsh stimuli. Stimuli 0 through 63 have positive contrast, and 64 through 127 have negative contrast.





pt185

FIG. 3. Transmitted information per stimulus,  $T_3(s; R)$ , plotted on an  $8 \times 8$  grid corresponding to the increasing sequency of the Walsh stimuli. *A*:  $T_3(s; R)$  for positive contrast patterns. *B*:  $T_3(s; R)$  for negative contrast patterns. The chaotic nature of the surfaces indicates that even one step in sequency can yield a large change in the amount of transmitted information. Hence there is no evidence for sequency "tuning." *T* is information transmitted per stimulus, *H* is horizontal sequency, and *V* is vertical sequency.

abilities of symbol occurrences. Once a response is encoded in symbols (representing bins for either spike count or waveform), the knowledge of other response features, such as magnitude, is lost. Nonetheless, it is possible that the information code does correspond with other response parameters. In this section we compare the amount of information transmitted by the different codes with a measure of response magnitude. The response measure is either the number of spikes in the response or the length of the vector formed by the coefficients of the first three principal components. (Vector length, the square root of the sum of the squares of the components, is a measure of response strength in a multidimensional space.)

The magnitude of the response does not correlate well with information. The lower

graph in Fig. 4 shows the number of spikes in the response as a function of the stimulus pattern. Comparison with the information measure in the upper graph of Fig. 4 indicates that large amounts of information can be communicated by either large, medium, or small firing rates. A direct measure of the correlation between the spike count and the amount of transmitted information (based on the temporal code) can be obtained from the correlation coefficient. A scatter plot of  $T_3(s; R)$  and spike count for the 128 stimuli is shown for one neuron in Fig. 5 (top). The correlation coefficient for these data is  $-0.15$ , clearly indicating that spike count is not correlated with transmitted information.

The measure of correlation between information and spike count may be influenced by how the base line of the neuron's response

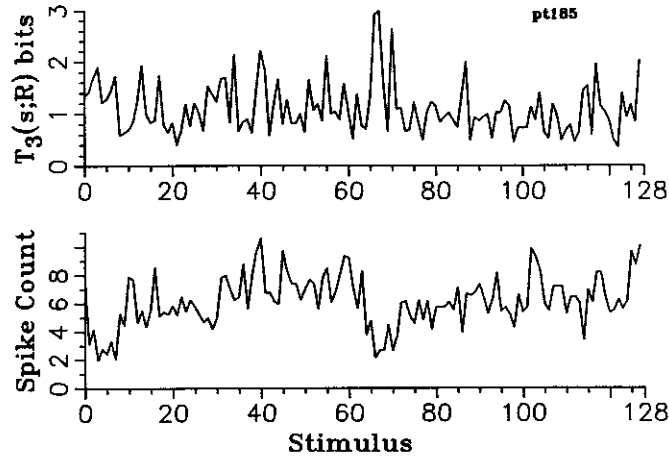


FIG. 4. Transmitted information per stimulus,  $T_3(s; R)$  and spike count for each stimulus.

varies. To avoid this complication, the information can be correlated instead with the length of the vector of the principal compo-

nents, since the principal components are defined with a base line of zero. Figure 5 (bottom) shows the scatter plot between transmitted in-

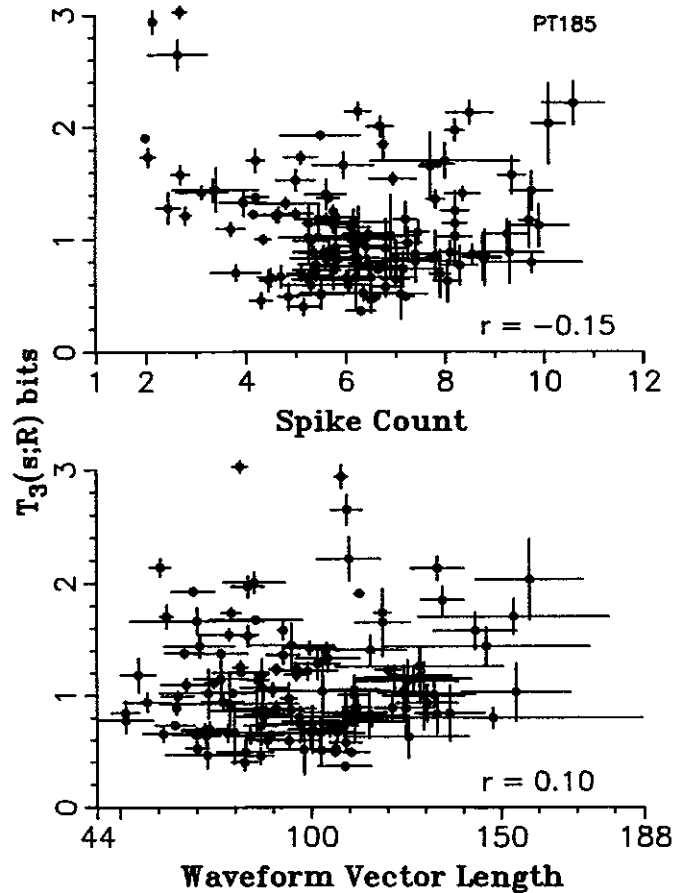


FIG. 5. Correlation between transmitted information,  $T_3(s; R)$  and response magnitude. *Top*: correlation with spike count. These points are the same data shown in Fig. 4. *Bottom*: correlation with magnitude of the temporal modulation code (the length of the vector consisting of the first 3 principal components). Note that high spike counts do not necessarily imply large amounts of transmitted information. Points are means  $\pm 1$  SE ( $n = 4$  or  $5$ ).

formation and the length of the temporal code vector. The correlation coefficient is still low (0.10), even for this multidimensional measure of response strength. Table 1 gives the average correlation coefficients for our population of neurons.

#### *Probability of error*

The probability of a transmission error, assuming a simple maximum likelihood detector, was calculated from  $P(s, r)$  for single-unit responses measured with both the spike count and the temporal code. The mean value of  $P_E$  was  $0.927 \pm 0.003$  (SE) based on the temporal code of three principal components, and  $0.963 \pm 0.004$  based on a spike count code (see Table 1). Although the probability of error by either measure is high, the value from the temporal code (based on three principal components) was smaller than that from the spike count code in every neuron.

If we assume that decisions are based on a set of independent neurons, the probability of error based on the whole set should go down as the number of neurons in the set is increased. If the probability of error of all of the neurons in the set is the same, say  $q$ , then the  $P_E$  for  $n$  statistically independent neurons would be  $q^n$ . Assume that the  $P_E$  of the neurons based on temporal code is  $p$ , and the  $P_E$  based on the spike count code is  $q$ . Then, using this simple model, one would need at least  $(\log(p)/\log(q))$  neurons to have the same  $P_E$  based on spike count code as the  $P_E$  of a single neuron based on the temporal code. For the average of our populations, this gives  $[\log(0.927)/\log(0.963)] = 2.01$ , so about twice as many neurons are needed in the population to make (maximum-likelihood based) decisions if spike count is used instead of temporal waveform.

In the METHODS section we pointed out that the use of a maximum-likelihood detector would give a conservative estimate of  $P_E$ . The ratio of identical neurons needed to obtain the same probability of error with a spike count code as with a temporal code is 2.01, whereas the ratio needed to obtain the same amount of information with a spike count code as with a temporal code is 2.25. The larger ratio from information transmission implies that a detector incorporating a priori knowledge of the neuron's properties could reduce the probability of error by 12% from that achieved with a maximum likelihood detector.

## DISCUSSION

The study of stimulus-response relations of IT neurons requires the discovery of features that stimulate the neurons and the identification of response parameters that are driven by those features. In the previous papers in this series we demonstrated that a set of two-dimensional black and white patterns, on brief presentation, determined several of the principal components of the response. The major finding of this final part of the study was that these principal components conveyed independent information about the stimulus.

This paper also compared the amount of information transmitted by single and multiple parameter temporal codes. The amount of information calculated increased with the number of response parameters (i.e., principal components) used in the temporal code. For example, the average transmitted information  $[T(S; R)]$  was approximately one bit when a three-parameter temporal code was assumed, but less than half a bit when a spike count code was assumed. The transmitted information per stimulus  $[T(s; R)]$  could be as much as three bits for some stimuli, and this information was not correlated with the magnitude of the response. These results lead to a new hypothesis about the function of visual neurons.

#### *Information theory*

The application of information theory to neurophysiology has consisted of attempts either to estimate the capacity of neural channels (19, 25, 34, 35, 36, 37) or to measure the amount of information transmitted about some stimulus by a neuron's response (8, 10, 14). Eckhorn and Pöpel (9) developed a method for calculating information based on a general binary response code. They used this method to show that information was not transmitted uniformly throughout the response in cat visual afferents but that it had a temporally modulated distribution (10). In a later study of cat lateral geniculate nucleus, Eckhorn et al. (8) showed that the information lost when using a spike count code was more severe when either the input information rates were high or when the average input spike rate was high. Fuller and Looft (14) also used a general binary code to show that more information was transmitted by touch receptors than could be accounted for by a spike count

code. All of these studies show that considerably more information is contained in the temporal modulation of a neuronal response than can be represented in a spike count code.

The results of our study demonstrate that information about spatial patterns is encoded in the temporal pattern of the neuron's responses. The use of spike count to calculate information underestimated the amount of information transmitted by the neuron. Furthermore, since spike count was not correlated with the transmitted information, it was not possible to infer anything about the significance of a response from its magnitude alone. This is consistent with the findings of Eckhorn and Pöpel (10), who found that the response magnitude could change even when there was little or no information being transmitted. These workers emphasized that transmitted information was only high when there was structure in the histogram of the responses. This implies that information is transmitted by the temporal modulation of the response, which we have represented with a principal component code.

Earlier analyses of temporal modulation of the neuronal response were made without regard to any intrinsic code that may be used by the neuron. A general binary code was used because it is based only on the knowledge of when spikes occur and thus can preserve all the information contained in a spike train. However, a general binary code cannot provide any insight into the encoding mechanism because of its large number of degrees of freedom. Our own approach has been to measure the information transmitted by a code made up of the principal components of the waveform of the neuron's responses. Since the principal components form a complete basis for any signal, it is possible to represent the same information as is contained in a general binary code by using all of the principal components (by choice of the length of the data vector, there were 64 in this study). However, we have shown that only the first few principal components of the response depend on the stimulus (28). The principal components represent the natural modes of a neuron's activity (28), so the limited number needed to encode the messages suggests that the range of time courses used by the encoding mechanisms is limited. An information code based on the first three principal components thus has an ad-

vantage over the general binary code in that it utilizes intrinsic statistical properties of the neuronal responses.

Our results show that the information transmitted by the spike train increases when more than one temporal component is used in the code (cf. Table 1). This implies that the information in different temporal components represents independent messages. Since the coefficients of the principal components are uncorrelated and largely independent, we are led to the hypothesis that a neuron is encoding more than one feature of the stimulus: i.e., a multidimensional spatial representation is encoded in a multidimensional temporal response. This can be referred to as the multiplex-filter hypothesis of neural function.

#### *Hypotheses about neural function*

In the classical view of stimulus analysis by visual system neurons (4, 18), each neuron was envisioned as representing some special feature of the stimulus. The magnitude of the response of the neuron would then signal its confidence that its preferred feature was present. The stimulus that excited the neuron the most was defined as the optimal, or trigger, feature. A contemporary outgrowth of this idea has been that neurons represent feature encoders rather than detectors. These neural encoders have been described in terms of filter properties, which are measured with tuning curves. In a tuning curve, one physical parameter (e.g., orientation) is varied systematically and a univariate measure of response (e.g., spike count) is collected. The stimulus at the peak of the tuning curve is assumed to be an optimal stimulus, characterizing the function of the neuron. For example, neurons in striate cortex have been described as tuned spatial frequency filters (7, 20, 31).

These approaches describe a stimulus in terms of several individual physical parameters, e.g., orientation, spatial frequency, or color. Such approaches establish a neuron's sensitivity to different stimulus parameters but do not address the question of the intrinsic dimensionality of the neuronal responses. What is meant by intrinsic dimensionality? If we consider a single black and white picture, then it has only two spatial dimensions, which can be represented by horizontal and vertical spatial frequency. We would say that its extrinsic dimensionality was two. Now, consider

all possible black and white pictures. There are, of course, an infinite number of them. However, they may be regarded as points in a conceptually abstract vector space of finite dimension. In other words, every black and white picture will correspond to a point in this space. Now we can ask: how many dimensions are needed to span this picture space? The answer depends only upon the spatial resolution of the pictures. If we consider pictures with infinite resolution, the picture space has an infinite number of dimensions. But, if we consider a resolution of one part in eight, then any picture can be drawn with just 64 ( $8 \times 8$ ) picture elements (called pixels). The dimensionality of the subspace is the number of pixels. So, even though there are an infinite number of such  $8 \times 8$  pictures, they only form a subspace of limited dimensionality, 64. In other words, any black and white picture (with a resolution of  $8 \times 8$ ) can be described as a weighted sum of a set of 64 pictures. There are an infinite number of sets of 64 pictures that span the subspace. An example of such a spanning, or basis, set consists of two-dimensional Walsh pictures shown in Fig. 2 of the first paper in this series (27). Note that any other basis set (e.g., Fourier patterns) would also have 64 dimensions. The number of dimensions needed to span a subspace will be referred to as its intrinsic dimensionality.

Returning to the data from the single-unit experiments, we can ask how these 64 picture dimensions are represented in the responses of a visual neuron. A basic limitation of the feature detector or tuned filter approaches is that all responses are measured along a single dimension (usually spike count). But studying univariate responses to many physical parameters, varied one at a time, does not provide any insight into the intrinsic dimensionality of the picture subspace that a neuron can represent in its responses. If only one dimension is tested, there is no possible way to estimate what the neuron may be doing in higher response dimensions. A method is needed that can give multidimensional measures of multiple physical parameters varied together, so that an estimate of intrinsic dimensionality can be made.

Intrinsic dimensionality, in statistical terms, is the number of degrees of freedom of the neuron's response that are dependent on stimulus features. In our method of analysis,

the intrinsic dimensionality is determined by the number of stimulus-dependent principal components that convey independent information. This number is about three or four (28), a significant reduction from the picture space dimensionality of 64. An intrinsic dimensionality greater than one, of course, implies that neurons are encoding multidimensional representations of pictures.

#### *The multiplex-filter hypothesis*

Our multivariate measure of a neuron's responses to our set of multidimensional stimuli can be used to determine intrinsic dimensionality. The results we have obtained imply that IT neurons have a limited intrinsic dimensionality (3 or 4). This leads to a new hypothesis of neuronal function in vision, namely the multiplex-filter hypothesis. In this hypothesis each neuron represents a picture in terms of its projection onto an intrinsic picture subspace. Each of the dimensions in this subspace can be thought of as a two-dimensional spatial filter. (The spatial filters defining the visual pattern subspace can be estimated from the ensemble of responses to the Walsh stimuli. A method for determining such estimates will be the subject of a later paper.) When a stimulus is present in the neuron's receptive field, it is weighted according to each of the neuron's spatial filters. The results of these weightings are encoded in an orthogonal set of signals (such as the principal components), and the information about the stimulus is then transmitted as a temporally modulated message.

One aspect of this hypothesis, supported by our finding that the magnitude of the neuron's response is not well correlated with the transmitted information, is that there is no optimal stimulus corresponding to a trigger feature or a single tuned filter. This lack of an optimal stimulus comes from the multidimensionality of the neuron's responses. In a system with more than one dimension, one response can not be compared with another. Scalars derived from the response, such as magnitude, can be compared, but they do not provide an ordering of multidimensional responses (e.g., although the number 1 is less than the number 2, 1 step to the left is not less than 2 steps forward).

The validity of the multiplex-filter hypothesis could be tested by implementing it in a specific quantitative model. This requires fur-

ther experimental study of the amplitude nonlinearity of the neuron's stimulus response relation. Such a model, based on preliminary studies of spatial and luminance response properties of complex cells in striate cortex, has successfully predicted temporally modulated responses to arbitrary stimuli (23, 26).

#### *Decoding neuronal messages*

In the feature detector model, the magnitude of the neuron's response indicated the degree of confidence that its trigger feature was present. The message could be passed on by assuming a threshold device of some kind, whereby the neurons that had the largest responses were considered to have detected their trigger features. In the tuned filter model, the amount of a stimulus feature present is encoded in a population of differently tuned neurons by the relative magnitudes of their responses.

The low correlations of raw measures of response (spike count or vector length) with transmitted information implies that stimulus features are being encoded in a set of response symbols and are not simply modulating the response strength. In our multiplex-filter model, the spike train is considered to be the simultaneous carrier of several different messages. Decoding a multiplexed signal would require a temporal filter scheme. Since many neurons will converge onto the same "downstream" neuron in the signal path, it seems reasonable to consider a principle of decoding that will maximize information transmission. Since information is based on a probability measure, two such principles are manifest. For an individual neuron the response should be consistent from one stimulus presentation to the next, since that will reduce the spread in the joint stimulus-response probability density function and thus increase information. For a group of convergent neurons the equivalent principle is coherence, i.e., they should all send the same message. Similar messages from different neurons will reinforce each other, since the a priori probability of such coincidence is very low. Note that such coherence need only occur for certain temporal components of the neuronal firing. If some messages were adding coherently, but others were not, the spike trains might appear noisy but would in fact be conveying information in the coherent components.

In our multiplex-filter approach to single neurons, it is possible to transmit messages about different stimulus dimensions with distinct temporal waveforms. Interpretation of stimulus parameters by downstream neurons could then be accomplished by a process of temporal filtering. Such filtering can be accomplished easily in at least two ways. For short time-course codes (say, <50 ms), the filtering could be achieved by using biophysical time constants of dendritic processes and controlling the place on the dendrites where synapses are made. For longer time-course codes, small interconnected groups of neurons could form a network with temporal filter properties. Hence, demultiplexing of messages may be a property of the biophysical characteristics of single neurons, and/or the network properties of groups of neurons.

Neuronal firing rate is influenced by many factors, such as stimulus contrast, color, orientation, and spatial frequency. For example, a spike count of 25 could correspond to a bar-shaped stimulus oriented at  $45^\circ$  with low contrast, or to a bar oriented at  $73^\circ$  with a higher contrast. Any proposed scheme for coding and decoding neurons must deal with the fact that neurons respond to many stimulus parameters. In the univariate approach, where information is conveyed by spike count alone, it is assumed that confounded stimulus parameters are separated at a later stage of processing. With a multivariate code, such confounding may also occur. One can ask, to a simple, first-order approximation, how the temporal encoding scheme compares with a univariate scheme in terms of separating confounded stimulus parameters.

For the univariate approach, suppose that the number of stimulus parameters is  $M$ , and the sensitivity of the  $i$ -th neuron to the  $j$ -th stimulus parameter is  $a_{ij}$ . If we assume that the response across  $N$  neurons is a linear combination of the stimulus sensitivities,  $a_{ij}$ , then we can think of the problem as one of solving  $N$  equations in  $M$  unknowns. It is then sufficient to have  $N = M$  neurons whose sensitivities are linearly independent, all of which converge on the same neuron in the downstream population. A downstream neuron, to decode one of the stimulus parameters, must weight its inputs, with the weights coming from a row of the inverse of the matrix with elements  $a_{ij}$ . This downstream neuron would

then have a pure one-dimensional sensitivity. Although there is no theoretical reason why this scheme could not work, neither is there any physiological evidence in support of it; no neuron is known to show sensitivity to just one stimulus parameter.

The major conceptual difficulty with this inverse decoding scheme arises when other phenomena, such as habituation or adaptation, are considered. Suppose that after adaptation, for example, new weights  $a'_{ij}$  are used to encode stimulus parameters. This means that, for the downstream cell to correctly separate the stimulus parameters, a new set of weights must be obtained by inverting the new matrix with elements  $a'_{ij}$ . A decoding scheme that needs to invert a matrix whenever a neuron's sensitivities change detracts from the simplicity of a univariate encoding scheme.

Even if some stimulus features are confounded in a temporal code, the multivariate scheme still has advantages over the univariate scheme. Suppose that some features are confounded in one code with weights  $a_{ij}$ . Then an inverse weighting must be available to separate information in the downstream cells. Unlike univariate encoding, however, these weightings may be an invariant of the neurons. Suppose, for example, that adaptation changes the weights to  $a'_{ij}$ . Furthermore, since all of these messages are coming from a single neuron, it is reasonable to assume that the  $a_{ij}$  and the  $a'_{ij}$  differ only by a scale factor. In this case a new inverse matrix need not be computed, since it is just a scaled version of the old inverse. Since the downstream neuron could also send multiple messages encoded by multidimensional temporal waveforms, it is not essential in this theory that any neurons exist with pure unidimensional sensitivities.

#### *Temporal modulation*

Even though the results of this series of studies show that the spike train is modulated in distribution as well as magnitude of response by stimulus presentation, these studies do not address the problem of whether the temporal modulation is actually used by the nervous system. It is well known that in motor systems both the number and the pattern of spikes influence motor function. For example, in the abducens nucleus, both the firing frequency and the rate of change of the frequency

influence the movement of the eye (29). If the frequency is constant the eye is stopped at a position proportional to the mean firing rate, whereas if the frequency is changing the eye moves with a speed proportional to the derivative of the firing rate. Therefore, both the mean frequency and its rate of change are important. Intracellular experiments in single neurons of isolated ganglia of *Aplysia californica* have shown that the excitatory postsynaptic potentials (EPSPs) and generated action potentials depend critically on the pattern as well as the number of the incoming spike train (32). Encoding of stimulus features by different temporal schemes has also been reported in the visual cortex. In area 17 of the cat, Cattaneo et al. (5) found that the temporal response of complex cells was modulated differently by different aspects of the stimulus. The clustering of spikes was related to spatial frequency and orientation of the stimulus, whereas the distribution of isolated spikes was related to stimulus contrast. These results are consistent with the interpretation of Eckhorn and Pöpel (10) that information is conveyed by structure, or grouping, of the spikes in the response.

The duration of neuronal codes may be reflected in psychophysical response latencies. For example, in human perception the critical period measures the interval over which the visual response is a function only of stimulus energy (the product of stimulus intensity and duration). One psychophysical experiment has shown that the critical period for stimulus evaluation in a stimulus identification task is about 200–350 ms, whereas in an otherwise similar brightness matching task the critical period is much shorter ( $\sim 100$  ms) (16). This suggests that pattern identification depends more on stimulus integration time than do simpler tasks such as brightness matching. The differences in these critical times (100–250 ms) are long enough to encompass the time course of the temporal modulation we have observed in neurons in IT cortex. These results raise the possibility that different functions (e.g., detection and discrimination) may utilize different measures of neural activity. For example, detection may be achieved just by counting spikes, whereas pattern identification may depend on the complete temporal code.

One disadvantage of a temporal encoding

scheme is that more time is required by a multivariate than by a univariate scheme to transmit information about the stimulus. A spike count code can be resolved over an integration time that is just long enough to obtain the necessary signal to noise ratio. Decoding a multiplexed message, however, depends not only on the signal to noise ratio, but also on the duration of an element of the code. Our work has not yet dealt with the crucial question of code duration. Furthermore, it must be remembered that all code words need not have the same length (e.g., in international Morse code the code word for E is dot, whereas the code word for Q is dash dash dot dash). Hence the duration of the neural code may be fixed, or it may vary so that frequent or important messages are shorter than others.

#### *Significance of a temporal code for neuronal function*

The new methodology proposed in this series of papers establishes a way to investigate quantitatively the stimulus-response relationships seen in neurons in the central nervous system. Our new methods have shown that neurons in inferior temporal cortex convey messages by temporal modulation of their firing rates. Hence, by considering information transmission through the visual system, the role of neurons in visual areas is seen to be similar to the role of neurons in other parts of the brain. The new approach also recognizes the intrinsically multidimensional nature of the neuron's response and hence can be easily and naturally extended to include the study of stimuli when more than one parameter is varied (e.g., contrast, color, orientation, or spatial frequency), or when more than one modality is stimulated (e.g., visual or auditory). Variations in nonphysical stimulus parameters, such as behavioral relevance (e.g., whether a stimulus is associated with a reward), can also be included with this type of information analysis of stimulus-response relationships.

Current neurophysiological techniques cannot determine whether neurons in sensory systems use information based on any particular code (e.g., spike count or temporal modulation). It is only possible to study what information is present in a spike train under the assumption of various codes. We have shown

that more stimulus-dependent information is present in a temporal than in a spike count code. We expect that further research into the nature of the temporal code and the representation of stimulus parameters by that code will yield insights into the role of neurons in visual processing.

#### APPENDIX

Fagen (12) used a closed form equation to estimate the variance related to the calculation of the average transmitted information. The average transmitted information (Eq. 4) can also be expressed by

$$T(S; R) = \sum_{s,R} P(s,r) \log \frac{P(s,r)}{P(s)P(r)} \quad (A1)$$

An estimate of the variance in closed form is

Var [T(S; R)]

$$= \frac{1}{N} \left[ \sum_{s,R} P(s,r) \left( \log \frac{P(s,r)}{P(s)P(r)} \right)^2 - T^2(S; R) \right] \quad (A2)$$

where  $N$  is the total number of data points.

The conditional transmitted information, or the transmitted information per code symbol (Eq. 2) can also be expressed as

$$T(s; R) = \sum_R \frac{P(s,r)}{P(s)} \log \frac{P(s,r)}{P(s)P(r)} \quad (A3)$$

Arguing in parallel with Fagen (12), we propose as an estimate of the variance of  $T(s; R)$  the same second moment type of closed form

$$\text{var} [T(s; R)] = \frac{1}{M} \left[ \sum_R \frac{P(s,r)}{P(s)} \left( \log \frac{P(s,r)}{P(s)P(r)} \right)^2 - T^2(s; R) \right] \quad (A4)$$

where  $M$  is the number of data points for stimulus  $s$ .

#### ACKNOWLEDGMENTS

We thank Dr. K. D. Pettigrew of the Theoretical Statistics and Mathematics branch, National Institute of Mental Health, for statistical help and many stimulating discussions about statistics and information. We are grateful to Drs. M. E. Goldberg, R. H. Wurtz, and M. Mishkin for reading all three manuscripts in various stages of their preparation and making many helpful comments.

Address reprint requests to: B. J. Richmond, National Institute of Mental Health, Bethesda, Maryland 20892.

Received 27 January 1986; accepted in final form 8 August 1986.



## REFERENCES

1. ABRAMSON, N. *Information Theory and Coding*. New York: McGraw-Hill, 1963.
2. ACZEL, J. AND DAROCZY, Z. *On Measures of Information and Their Characterizations*. New York: Academic, 1975.
3. AHMED, N. AND RAO, K. R. *Orthogonal Transforms for Digital Signal Processing*. Berlin: Springer-Verlag, 1975.
4. BARLOW, H. B. Summation and inhibition in the frog's retina. *J. Physiol. Lond.* 119: 69-88, 1953.
5. CATTANEO, A., MAFFEI, L., AND MORRONE, C. Patterns in the discharge of simple and complex visual cortical cells. *Proc. R. Soc. Lond. B Biol. Sci.* 212: 279-297, 1981.
6. CHUNG, S.-H., RAYMOND, S. A., AND LETTVIN, J. Y. Multiple meaning in single visual units. *Brain Behav. Evol.* 3: 72-101, 1970.
7. DEVALOIS, D. L., ALBRECHT, D. G., AND THORELL, L. G. Spatial frequency selectivity of cells in Macaque visual cortex. *Vision Res.* 22: 545-559, 1982.
8. ECKHORN, R., GRÜSSER, O.-J., KRÖLLER, J., PELLNITZ, K., AND PÖPEL, B. Efficiency of different neural codes: information transfer calculations for three different neuronal systems. *Biol. Cybern.* 22: 49-60, 1976.
9. ECKHORN, R. AND PÖPEL, B. Rigorous and extended application of information theory to the afferent visual system of the cat. I. Basic concepts. *Kybernetik* 16: 191-200, 1974.
10. ECKHORN, R. AND PÖPEL, B. Rigorous and extended application of information theory to the afferent visual system of the cat. II. Experimental results. *Kybernetik* 17: 7-17, 1975.
11. EFRON, B. *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1982.
12. FAGEN, R. M. Information measures: statistical confidence limits and inference. *J. Theor. Biol.* 73: 61-79, 1978.
13. FUKUNAGA, K. *Introduction to statistical pattern recognition*. New York: Academic, 1972.
14. FULLER, M. S. AND LOOFT, S. J. Information-theoretic analysis of cutaneous receptor responses. *IEEE Trans. Biomed. Eng.* 31: 377-383, 1984.
15. GALLAGER, R. G. *Information Theory and Reliable Communication*. New York: Wiley, 1968.
16. KAHNEMAN, D. AND NORMAN, J. The time-intensity relation in visual perception as a function of observer's task. *J. Exp. Psychol.* 68: 215-220, 1964.
17. KULLBACK, S. *Information Theory and Statistics*. New York: Wiley, 1959.
18. LETTVIN, J. Y., MATURANA, H. R., MCCULLOCH, W. S., AND PITTS, W. H. What the frog's eye tells the frog's brain. *Proc. Inst. Rad. Eng.* 47: 1940-1951, 1959.
19. MACKAY, D. M. AND MCCULLOCH, W. S. The limiting information capacity of a neuronal link. *Bull. Math. Biophys.* 14: 127-135, 1952.
20. MAFFEI, L. AND FIORENTINI, A. The visual cortex as a spatial frequency analyzer. *Vision Res.* 13: 1255-1267, 1973.
21. MCGILL, J. J. Multivariate information transmission. *IRE Trans. Info. Theory* 4: 93-111, 1954.
22. MURTHY, V. K. Estimation of probability density. *Ann. Math. Statist.* 36: 1027-1031, 1965.
23. OPTICAN, L. M. AND RICHMOND, B. J. Temporal encoding of pictures by striate neuronal spike trains: II. Predicting complex cell responses. *Soc. Neurosci. Abstr.* 12: 431, 1986.
24. PARZEN, E. On estimation of a probability density function and mode. *Ann. Math. Statist.* 33: 1065-1076, 1962.
25. RAPOPORT, A. AND HORVATH, W. J. The theoretical channel capacity of a single neuron as determined by various coding systems. *Inform. Control* 3: 335-350, 1960.
26. RICHMOND, B. J. AND OPTICAN, L. M. Temporal encoding of pictures by striate neuronal spike trains: I. The multiplex-filter hypothesis. *Soc. Neurosci. Abstr.* 12: 431, 1986.
27. RICHMOND, B. J., OPTICAN, L. M., PODELL, M., AND SPITZER, H. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J. Neurophysiol.* 57: 132-146, 1987.
28. RICHMOND, B. J. AND OPTICAN, L. M. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J. Neurophysiol.* 57: 147-161, 1987.
29. ROBINSON, D. A. AND KELLER, E. L. The behavior of eye movement motoneurons in the alert monkey. *Bibl. Ophthalmol.* 82: 7-16, 1972.
30. SAKITT, B. Visual-motor efficiency (VME) and the information transmitted in visual-motor tasks. *Bull. Psychon. Soc.* 16: 329-332, 1980.
31. SCHILLER, P. H., FINLAY, B. L., AND VOLMAN, S. F. Quantitative studies of single-cell properties in monkey visual cortex. III. Spatial frequency. *J. Neurophysiol.* 39: 1334-1351, 1976.
32. SEGUNDO, J. P., MOORE, G. P., STENSAAS, L. J., AND BULLOCK, T. H. Sensitivity of neurones in *Aplysia* to temporal pattern of arriving impulses. *J. Exp. Biol.* 40: 643-667, 1963.
33. SHANNON, C. E. A mathematical theory of communication. *AT&T Bell L. Tech. J.* 27: 379-423, 1948.
34. STARK, L. AND THEODORIDIS, G. C. Information theory in physiology. In: *Engineering Principles in Physiology*, edited by J. H. U. Brown and D. S. Gann. New York: Academic, 1973, p. 13-32.
35. STEIN, R. B. The information capacity of nerve cells using a frequency code. *Biophys. J.* 7: 797-826, 1967.
36. WALLØE, L. On the transmission of information through sensory neurons. *Biophys. J.* 10: 745-763, 1970.
37. WERNER, G. AND MOUNTCASTLE, V. B. Neural activity in mechanoreceptive cutaneous afferents: stimulus response relations. Weber functions and information transmission. *J. Neurophysiol.* 28: 358-397, 1965.