JOURNAL OF NEUROPHYSIOLOGY
Vol. 76, No. 2, August 1996. *Printed in U.S.A.*

Nature and Precision of Temporal Coding in Visual Cortex: A Metric-Space Analysis

JONATHAN D. VICTOR AND KEITH P. PURPURA
Department of Neurology and Neuroscience, Cornell University Medical College, New York, New York 10021

SUMMARY AND CONCLUSIONS

1. We recorded single-unit and multi-unit activity in response to transient presentation of texture and grating patterns at 25 sites within the parafoveal representation of V1, V2, and V3 of two awake monkeys trained to perform a fixation task. In grating experiments, stimuli varied in orientation, spatial frequency, or both. In texture experiments, stimuli varied in contrast, check size, texture type, or pairs of these attributes.

2. To examine the nature and precision of temporal coding, we compared individual responses elicited by each set of stimuli in terms of two families of metrics. One family of metrics, D^{spike}, was sensitive to the absolute spike time (following stimulus onset). The second family of metrics, $D^{interval}$, was sensitive to the pattern of interspike intervals. In each family, the metrics depend on a parameter q, which expresses the precision of temporal coding. For $q = 0$, both metrics collapse into the "spike count" metric (D^{count}), which is sensitive to the number of impulses but insensitive to their position in time.

3. Each of these metrics, with values of q ranging from 0 to 512/s, was used to calculate the distance between all pairs of spike trains within each dataset. The extent of stimulus-specific clustering manifest in these pairwise distances was quantified by an information measure. Chance clustering was estimated by applying the same procedure to synthetic data sets in which responses were assigned randomly to the input stimuli.

4. Of the 352 data sets, 170 showed evidence of tuning via the spike count ($q = 0$) metric, 294 showed evidence of tuning via the spike time metric, 272 showed evidence of tuning via the spike interval metric to the stimulus attribute (contrast, check size, orientation, spatial frequency, or texture type) under study. Across the entire dataset, the information not attributable to chance clustering averaged 0.042 bits for the spike count metric, 0.171 bits for the optimal spike time metric, and 0.107 bits for the optimal spike interval metric.

5. The reciprocal of the optimal cost q serves as a measure of the temporal precision of temporal coding. In V1 and V2, with both metrics, temporal precision was highest for contrast (ca. 10–30 ms) and lowest for texture type (ca. 100 ms). This systematic dependence of q on stimulus attribute provides a possible mechanism for the simultaneous representation of multiple stimulus attributes in one spike train.

6. Our findings are inconsistent with Poisson models of spike trains. Synthetic data sets in which firing rate was governed by a time-dependent Poisson process matched to the observed poststimulus time histogram (PSTH) overestimated clustering induced by D^{count} and, for low values of q, $D^{spike}[q]$ and $D^{interval}[q]$. Synthetic data sets constructed from a modified Poisson process, which preserved not only the PSTH but also spike count statistics accounted for the clustering induced by D^{count} but underestimated the clustering induced by $D^{spike}[q]$ and $D^{interval}[q]$.

INTRODUCTION

The manner in which sensory signals are encoded into the spike discharge of a neuron is a fundamental question in neuroscience. One emerging theme of recent physiological investigations is that the temporal pattern of a cortical neuron's discharge, and not just the number of spikes, is critically important in transmitting information. Evidence in support of this view has been obtained from single- and multi-unit studies at several levels of the visual system (Gawne et al. 1991; McClurkin et al. 1991a; Optican and Richmond 1987; Purpura et al. 1993; Richmond et al. 1987), the auditory system (Abeles 1982a, 1991; Geisler et al. 1991; Middlebrooks et al. 1994), and in the olfactory system (Skarda and Freeman 1987). It recently has been demonstrated that, despite the great variability in the discharges of cortical neurons (Softky and Koch 1993), the spike-generating mechanisms are intrinsically very precise (Mainen and Sejnowski 1995). Such precision is necessary for the propagation of information by a high-resolution temporal code. However, whether the timing per se of nerve impulses conveys information, or rather reflects the variability of neuronal inputs or spike-generating mechanisms, remains controversial (Shadlen and Newsome 1994, 1995; Softky 1995). Furthermore, even imprecise or variable neuronal responses are consistent with temporal coding, provided that differences in responses to stimuli that share a particular attribute is small in comparison with the differences in responses to stimuli that vary along that attribute. Thus to address the significance of temporal coding, it is necessary to consider not just the intrinsic variability of responses to the same stimulus, but also to compare this variability with the variability encountered as a stimulus attribute is changed. Furthermore, while temporal coding might be embodied in the precise (i.e., millisecond) times of occurrences of spikes, the degree of precision needs to be determined empirically, rather than assumed.

To determine whether a set of spike discharges depends systematically on a set of stimuli, it is necessary to have a notion of similarity of spike trains, typically expressed as a "distance" between spike trains. Most previous approaches (Chee-Orts and Optican 1993; Geisler et al. 1991, McClurkin et al. 1991a; Optican and Richmond 1987), taking their cue from methods appropriate for the analysis of multivariate data (Fukunaga 1990), have binned the spike trains and derived notions of distance from the Euclidean distance in a vector space. Within this vector space, principal components analysis (McClurkin et al. 1991a; Optican and Richmond 1987), clustering algorithms (Chee-Orts and Optican 1993), or other forms of pattern analysis (Geisler, Albrecht et al. 1991) may be applied. A practical drawback to this approach is that to achieve adequate time resolution (e.g., 2 ms) over a reasonable analysis interval (e.g., 400 ms), a very high-

dimensional vector space is required. Because this space is populated only sparsely by the existing data set, it is necessary either to synthesize artificial data to populate the space more densely (Chee-Orts and Optican 1993)—which entails assumptions about the underlying temporal code—or to limit the detail of the analysis (Geisler et al. 1991). Furthermore, there are theoretical reasons to doubt whether a vector-space approach is appropriate for this purpose (Hopfield 1995).

A second kind of strategy (Middlebrooks et al. 1994) is to develop a neural network scheme for the classification of spike discharges. While such approaches in principle can surmount the temporal resolution problem, it is difficult for the investigator to gain insight into the nature of the temporal code from an examination of the parameters of the neural network. A third kind of strategy explicitly deals with spike trains as point processes but focuses on correlations among discharges (Perkel et al. 1967a,b), the pattern of interspike intervals (Rapp et al. 1994), or the identification of similar segments of spike discharges (Abeles and Gerstein 1988), rather than on a global analysis of how the pattern of the discharge depends on the stimulus.

The approach pursued here is a novel one (Victor and Purpura 1994). We retain the notion of a distance, or measure of dissimilarity, between spike trains, so that we can determine whether there is a systematic dependence of the temporal structure of the response on stimulus parameters. However, the distances we define do not require binning of spike trains or embedding them in a vector space. Rather, spike trains are considered to be points in a metric space, and the distances between them correspond to "metrics" (Gaal 1964): a topological framework substantially more general than a vector space with a Euclidean distance. Whether our constructed metrics indeed correspond to a Euclidean distance in a vector space can in principle be determined empirically rather than assumed.

Our plan is to consider two families of metrics, each based on a simple neurobiological heuristic. Each metric determines a candidate geometry for the observed spike trains based on some aspect of their temporal structure. Within the context of each candidate geometry, we will determine to what extent responses show systematic stimulus-dependent clustering. One family of metrics, denoted $D^{spike}[q]$, embodies the idea that the absolute timing of individual impulses is significant. The rationale behind this idea is that a cortical neuron is not merely an integrator with a threshold; in some circumstances, it behaves as a coincidence detector (Abeles 1982b; Mel 1993; Softky and Koch 1993). Thus the effect of a spike train on a cortical neuron will depend on the absolute timing of the impulses and not just on the number of spikes within a given interval. A second family of metrics, denoted $D^{interval}[q]$, embodies the idea that the duration of the interspike intervals is significant. The rationale behind this idea is that, presumably as a consequence of properties of the *N*-methyl-D-aspartate receptor and Ca^{2+} channels, the effect of an action potential can depend critically on the length of the intervals since the previous potentials. This dependence can result in both short- and long-term potentiation (Bliss and Collingridge 1993), which are known to be sensitive to the pattern of interspike intervals (Bliss and Collingridge 1993; Larson et al. 1986; Rose and Dunwiddie 1986).

We apply these new tools to recordings of single units and small clusters of units in the visual cortex of the awake behaving monkey in response to presentations of gratings and texture patches. With this approach, we identify the qualitative and quantitative features of temporal codes, such as whether there are differences between temporal coding in V1 and V2, whether there are differences among the encoding of contrast, spatial frequency, orientation, and texture, and whether temporal coding might provide a way to signal multiple stimulus attributes jointly.

METHODS

Physiologic methods

Data sets consisted of neural responses elicited in cortical areas V1, V2, and V3 of two awake behaving rhesus monkeys. Details of the surgical methods for headgear and subconjunctival scleral search coil implantation (Judge et al. 1980; Richmond et al. 1983; Robinson 1963; Wurtz 1969) and extracellular recording (Crist et al. 1988; McClurkin et al. 1991b) are given elsewhere. All procedures involving the animals were performed in accordance with National Institutes of Health guidelines for the care and use of laboratory animals. Data were recorded from one hemisphere of each monkey. Of the 25 recording sites in two monkeys, 6 were determined to be in V1, 9 were determined to be in V2, 4 were in V3, and the remaining 6 sites discussed here were in either V1 or V2. The positions of the recording sites were reconstructed from histological preparation of the striate and early extrastriate cortex after the recording sessions and from the size and positions of the isolated receptive fields (Gattas et al. 1988). Ten recording sites yielded single-unit recordings, and the remainder yielded recordings from small clusters (2 or 3 neurons). In the analysis of these multi-unit data sets, the neuron of origin for each spike was not identified.

Visual responses were elicited by transient presentation of stimuli during a fixation task. Eye movements were monitored with the magnetic field search coil technique (Robinson 1963). Stimuli were produced by a microcomputer augmented with a Number Nine graphics card and displayed on a video monitor with a 60-Hz frame rate. Details of the methods used to generate the texture ensemble bit maps are available elsewhere (K. Purpura, M.-N. Chee-Orts, and L. M. Optican, unpublished data). A minicomputer recorded the timing of extracellularly recorded action potentials (to the nearest millisecond) as well as the monkey's eye position and controlled the presentation of the visual stimuli by the microcomputer and reward delivery through a program running under a real-time UNIX-based operating system (Hays et al. 1982). A successful trial consisted of maintained fixation within 0.5 deg during stimulus presentation.

Analysis was restricted to the neural responses that occurred during the 256-ms period of stimulus presentation. Stimuli consisted either of stationary sinusoidal luminance gratings or a square lattice of checks that were colored light and dark. Five kinds of parametric variation of the stimulus were used: contrast, check size, orientation, spatial frequency, and texture. Contrast $[(L_{max} - L_{min})/(L_{max} + L_{min})]$ was varied in two (0.5, 1.0), six (0.04, 0.08, 0.16, 0.24, 0.32, 0.64), or eight (0.04, 0.08, 0.12, 0.16, 0.24, 0.32, 0.64, 0.96) steps. Check size was varied in three steps spaced by factors of two, which bracketed the check size which elicited the largest response from the unit(s) under study. Orientation (grating stimuli only) was varied in eight equally spaced in steps of 22.5 deg. Spatial frequency (grating stimuli only) was varied in three or five values spaced by factors of two, which included the optimal

TABLE 1. *Summary of the experiments performed and the resulting data sets*

Cortical Area	Recording Sites	Number of Data Sets					
		Contrast	Check size	Orientation	Spatial frequency	Texture type	All
Definite V1	6	15	0	20	32	30	97
Definite V2	9	18	24	36	36	12	126
V1 or V2	6	18	0	0	0	36	54
Total V1 and V2	21	51	24	56	68	78	277
V3	4	9	6	6	6	48	75
Grand total	25	60	30	62	74	126	352

spatial frequency. Texture stimuli consisted of checkerboards colored according to one of three "isodipole" (Julesz et al. 1978) schemes (random, even, or odd). Isodipole textures, by definition, are stimulus ensembles that have equal mean luminance and equal spatial frequency spectra but differ in higher-order statistics. For the texture experiments, each stimulus class consisted of many different examples of a single isodipole texture rather than the same texture sample repeated over and over. As in our previous work with isodipole textures (Purpura et al. 1994), the rationale for this is to ensure that variations of statistics of a single texture sample from that of the ensemble cannot account for a systematic difference in responses (Victor 1994). Neural mechanisms that discriminate between isodipole texture ensembles must be sensitive to visual "features" and not just spectral content (Victor et al. 1995).

Table 1 presents a summary of the experiments performed. A typical recording consisted of 15–20 responses to all possible combinations of two of the parameters (e.g., check size and contrast) obtained in block-randomized order. These data were analyzed along each of the varied dimensions, providing a total of 352 data sets.

Mathematical methods: overview

For the reasons described in the INTRODUCTION, we constructed a method to analyze the temporal structure of spike trains based only on the bare essentials: an abstract set of points (the spike trains) and a self-consistent definition of distances between pairs of these points, known as a metric. The set of points, along with the metric (to be specified below), define a "metric space". This is in contrast to many traditional approaches to the analysis of the dynamics of neural responses, in which the responses are considered to be the elements of a vector space equipped with a Euclidean distance. The vector space structure is more restrictive than the metric space structure in that it implies a sense in which responses can be added together and multiplied by a scalar. However, defining these operations on neural responses is tantamount to making assumptions on how neural responses represent stimuli, and we would like to avoid making such assumptions whenever possible. The main justification of the Euclidean distance is that it respects the vector space operations of addition and scalar multiplication. The distances (metrics) that we consider, while well motivated by biological considerations, do not correspond to Euclidean distances in a vector space and are thus not based on assumptions about how responses should be scaled or combined.

For each choice of a metric, we will determine the extent of stimulus-dependent clustering—i.e., to what extent responses to stimuli that differ along some parameter (e.g., spatial frequency) tend to be far apart, while responses to stimuli that share a value of this parameter tend to be close.

The specific metrics that we will consider ($D^{spike}[q]$ and $D^{interval}[q]$) depend on a parameter q (s^{-1}), which is the "cost" (per unit time) to "move" a spike or to change the duration of an interval. For $q = 0$, these metrics degenerate to a metric based solely on the number of spikes. Increasing values of q correspond to a progressively greater dependence of the metric on the temporal pattern of the impulses rather than just their number. Thus a systematic dependence of the pattern of spike times or spike intervals on the stimulus will be manifest as increased clustering for some values of $q > 0$. However, sufficiently high values of q are anticipated to result in decreases in systematic clustering, because on biological grounds, the infinitesimally precise timing of impulses or intervals cannot possibly carry information (see for example Fig. 2). Thus our approach will characterize temporal coding in two ways: the amount of systematic clustering seen with $q > 0$ will indicate the extent to which absolute spike times ($D^{spike}[q]$) or spike intervals ($D^{interval}[q]$) depend on the stimulus, and the value of q for which the clustering is greatest will indicate the temporal resolution of the coding.

Mathematical formalism

The crucial mathematical step is to construct a metric for the set of spike trains so that each spike train can be regarded as a point in a metric space. (For an introduction to metric spaces, see for example Gaal 1964). A metric D is a mapping from a pair of points (here, the spike trains, denoted S_a, S_b, . . .) to a real number that represents the distance between the points. The metric must have certain properties so that it may be regarded as a distance. The metric must always be positive except for the trivial case $D(S, S) = 0$, be a symmetric function $[D(S_a, S_b) = D(S_b, S_a)]$, and satisfy the triangle inequality

$$D(S_a, S_c) \leq D(S_a, S_b) + D(S_b, S_c) \quad (1)$$

All of the metrics D that we will consider have a similar structure: in each case, $D(S_a, S_b)$ is defined as the minimum "cost" required to transform the spike train S_a into the spike train S_b via a path of allowed elementary steps. The cost assigned to a path of steps is the sum of the costs assigned to each of the allowed elementary steps; these costs and the set of allowed elementary steps determine the character of the metric. We denote the cost assigned to an elementary step from S to S' by $K(S, S')$, and require that this cost be always greater than or equal to zero, and that $K(S, S') = K(S', S)$. With these preliminaries, the distance between the spike train S_a and S_b is given by the smallest total cost of any sequence of elementary steps that connects S_a with S_b. That is

$$D(S_a, S_b) = \min \{K(S_a, S_1) + K(S_1, S_2) + \ldots + K(S_{n-2}, S_{n-1}) + K(S_{n-1}, S_b)\} \quad (2)$$

where "min" in *Eq. 2* denotes the minimum over all possible paths of spike trains S_a, S_1, S_2, . . . , S_{n-1}, S_b that begin at S_a and end at S_b and are connected by allowed elementary steps.

Metric based on spike times

We next create a family of metrics based on spike times, denoted by $D^{spike}[q]$, in which q is a parameter that expresses the relative

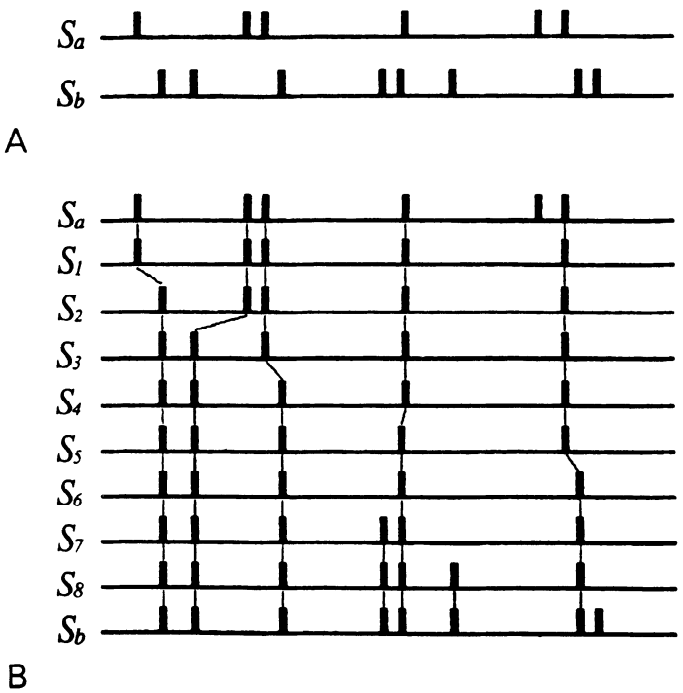


FIG. 1. *A*: two spike trains S_a and S_b for which metrics $D^{spike}[q]$ are to be determined. *B*: a path of allowed elementary steps associated with $D^{spike}[q]$ connecting 2 spike trains S_a and S_b.

sensitivity of the metric to precise timing of spikes. For the $D^{spike}[q]$ family of metrics, there are two kinds of allowed elementary steps. The first kind of step consists of adding a single spike or deleting a single spike and is assigned the cost of 1. This serves to ensure that there exists at least one path between any two spike trains. The second kind of allowed elementary step consists of a shift in the time of occurrence of a single spike by an amount Δt and is assigned the cost $q|\Delta t|$. That is, the further a spike is shifted, the more costly the elementary step. Figure 1 shows an example of two spike trains and a path of allowed elementary steps between them.

To gain an appreciation for this metric, it helps to consider two limiting cases. First, consider a cost/second q, which is zero. In this regime, any elementary step that shifts the position in time of a spike is free; costs are associated only with adding or deleting spikes. We can therefore calculate the metric of *Eq. 2* readily. Given a spike train $S_a = (a_1, a_2, \ldots, a_m)$ and a longer spike train $S_b = (b_1, b_2, \ldots, b_n)$, we construct a path in which the first m steps in the sequence of *Eq. 2* consist of shifting the m spikes of S_a until they coincide with the first m spikes of S_b, and the last $n - m$ steps consist of adding the last $n - m$ spikes of S_b. The first m steps have no cost, and the last $n - m$ steps have unit cost. Any other path would be necessarily at least as costly, because at least $n - m$ steps would have to involve addition of a spike. Thus the distance between these two spike trains is the difference in the number of spikes. That is, by setting the parameter q to zero, we have recovered the "spike count" metric, to be denoted D^{count}, in which the only relevant aspect of a spike train is its total number of spikes.

The reader may have noted that for D^{count}, the first condition is not satisfied: two spike trains S_a and S_b with the same number of spikes have $D(S_a, S_b) = 0$. There is a simple and standard formal patch for this degenerate situation. The space of spike trains is formally a "pseudometric space" (Gaal 1964) and can be restructured as a metric space of "equivalence classes of spike trains", where the equivalence class of spike trains that includes S is the set of spike trains whose distance from S is 0. The function defined by *Eq. 2* now becomes a valid metric on equivalence classes of spike trains. With this formal patch, D^{count} provides a notion of a distance.

The other extreme is that the cost/second to move a spike, q, is very large. Consider two spike trains, $\{a\}$ and $\{b\}$, each of which consist of only a single spike at times a and b, respectively. One path between these trains consists of moving the solitary spike from time a to time b and has a cost $K(\{a\}, \{b\}) = q|a - b|$. An alternate path consists of deleting the spike from S_a (to form an empty spike train) and then inserting a spike into the empty spike train at time b. This path has total cost $K(\{a\}, \varnothing) + K(\varnothing, \{b\}) = 2$. The second path (deletion and insertion of a spike) is cheaper than the first path (moving the spike) provided that $|a - b| > 2/q$. In the limit that q is very large, the lowest-cost path between any two spike trains is to remove all the spikes from one train that are not perfectly synchronous with spikes in the other train and then insert spikes to complete the new train. That is, the distance between two spike trains $S_a = \{a_1, a_2, \ldots, a_m\}$ and $S_b = \{b_1, b_2, \ldots, b_n\}$ is $m + n - 2c$, where c is the number of spike times in common, namely, the number of elements in $S_a \cap S_b$.

In essence, $D^{spike}[0]$ measures distances between spike trains in a manner that is independent of the time of occurrence of the spikes, while $D^{spike}[\infty]$ is maximally stringent. Between these two extremes, the family of metrics $D^{spike}[q]$ state that displacing a spike by an amount $1/q$ is equal in cost to deleting it altogether. This indicates that q may be taken as a measure of the precision of the temporal coding, or, equivalently, that $1/q$ determines how far a spike time can be nudged without substantially increasing the distance between the two trains in the metric space.

Metric based on interspike intervals

We also will consider a second family of metrics, $D^{interval}[q]$, which is sensitive to interspike intervals in much the same way that $D^{spike}[q]$ is sensitive to spike times. As in $D^{spike}[q]$, one kind of allowed elementary step (*Eq. 2*) consists of adding a single spike (which subdivides a single interspike interval into two adjacent intervals) or deleting a single spike (which merges two adjacent interspike intervals into one) and has a cost of 1. The second kind of allowed elementary step consists of changing the length *of an interspike interval* by an amount Δt and is assigned the cost $q|\Delta t|$. Note that a change in the length of an interspike interval necessarily changes the time of occurrence of all subsequent spikes. This is in contrast to the elementary step of a single spike time in $D^{spike}[q]$, in which exactly two intervals are modified (the intervals immediately preceding and following the shifted spike).

Efficient algorithms for the calculation of metrics

There are simple and efficient algorithms that construct the minimal path(s) required by the definition of *Eq. 2* and thereby calculate the metrics $D^{spike}[q]$ and $D^{interval}[q]$. These algorithms are related to the elegant algorithms introduced by Sellers (1974) for calculating the distance between two genetic sequences (i.e., a sequence of codons or of amino acid residues). For $D^{spike}[q]$, the algorithm is best understood via a geometric argument. Let us assume that we want to calculate the distance between the two spike trains S_a and S_b, shown in Fig. 1*A*. At first glance, we notice that S_a and S_b differ in their number of spikes (see Fig. 1*A*). Thus there will have to be an addition of two spikes to S_a to produce equal spike counts. The cost of adding two spikes is the entire cost required to transform S_a to S_b for $q = 0$. However, for $q > 0$, there will be additional costs incurred by moving the spikes so that the spike times match, and this cost may depend in a fairly complex

way on q. For example, the second spike in S_a could be moved back in time to align it with the second spike in S_b. But if the cost q of moving a spike is large, it might be better just to delete the second spike in S_a and reinsert it at an earlier time.

Let us now assume that we have identified a path of minimum cost $S_a = S_0, S_1, \ldots, S_{r-1}, S_r = S_b$ that transforms S_a to S_b. The sequence of elementary steps can be diagrammed by tracing the "life history" of each spike, as shown in Fig. 1*B*. Because this path is assumed to be of minimal cost, there are severe constraints on what this diagram can look like. The path cannot include both moving and deleting the same spike, because the cost could always be lowered by deleting the spike before it is moved. Thus all spike deletions always may be performed as the initial step(s). Similarly, no minimal path can include both inserting and moving the same spike, because the cost could be lowered by inserting the spike in its final position, and thus all spike insertions always may be performed as the final step(s). Furthermore, no minimal path can include inserting and then deleting the same spike; if it did, the cost could be lowered by eliminating both steps, along with any intervening shifts of this spike. Similarly, in a minimal path, no spike's trajectory ever includes both leftward and rightward shifts; if it did, the cost could be lowered by eliminating the redundant shifts. Finally, no minimal path includes a crossing of two spikes' trajectories, because if it did, the cost could be lowered by uncrossing them. These observations force one of three alternatives: the last spike of spike train S_a is a spike to be deleted, the last spike of spike train S_b is an inserted spike, or the last spikes of both trains are connected by a shift. Thus a path of minimum cost between two spike trains S_a and S_b must always be one of the following: deletion of the last spike of S_a, to form S_a', along with a minimal path between S_a' and S_b; deletion of the last spike of S_b, to form S_b', along with a minimal path between S_a' and S_b'; or a minimal path between S_a' and S_b', along with a shift of the final spike of S_a and S_b so that they coincide.

This argument leads to an inductive algorithm for the distance $D^{spike}[q](S_a, S_b)$ between two spike trains $S_a = \{a_1, a_2, \ldots, a_m\}$ and $S_b = \{b_1, b_2, \ldots, b_n\}$. We use $G_{i,j}^{spike}$ to denote the distance between a spike train composed of the first i spikes of S_a, and the first j spikes of S_b, [so that $G_{m,n}^{spike} = D^{spike}[q](S_a, S_b)$]. The three possibilities above imply that

$$G_{i,j}^{spike} = \min \{ G_{i-1,j}^{spike} + 1, G_{i,j-1}^{spike} + 1, G_{i-1,j-1}^{spike} + q|a_i - b_j| \} \quad (3)$$

The operation of this algorithm can be viewed as a two-dimensional spreadsheet, in which the cell in the ith row and jth column contains $G_{i,j}^{spike}$. The initial row of the spreadsheet is filled by noting that $G_{0,j}^{spike} = j$, and the initial column is filled by noting that $G_{i,0}^{spike} = i$. Subsequent cells are filled with the formula of *Eq. 3*, which depends only on the cells immediately above and to the left. The value that appears in the final column of the final row ($i = m$, $j = n$) contains $G_{m,n}^{spike}$, which is the desired distance $D^{spike}[q] \times (S_a, S_b)$.

For $D^{interval}[q]$, Sellers' original argument (Sellers 1974) leads to an algorithm of the same form as *Eq. 3*, with the sequence of interspike intervals playing the role that the spike times play in *Eq. 3*. To compute the distance $G_{i,j}^{interval}$ between a spike train composed of the first i spikes of S_a, and the first j spikes of S_b, (so that $G_{m,n}^{interval} = D^{interval}[q](S_a, S_b)$), we use

$$G_{i,j}^{interval} = \min \{ G_{i-1,j}^{interval} + 1, G_{i,j-1}^{interval} + 1, G_{i-1,j-1}^{interval} + q|e_i - f_j| \} \quad (4)$$

where e_i is the length of the ith interspike interval of S_a, and f_j is the length of the jth interspike interval of S_b.

For the family of metrics $D^{interval}[q]$, there is a technical detail that arises in assigning lengths to the first and last interspike intervals, because these are bounded by the ends of the data collection period. The first (and the last) interspike intervals of each train have unknown lengths but must be at least as long as the time between the start (or end) of the data collection period and the first (or last) spike. In the work reported here, we choose the lengths of the first and last interspike intervals of each train to be the ones that minimize the calculated distance (*Eq. 4* and hence *Eq. 2*). However, other choices (e.g., taking the first (and last) interspike intervals to be equal to the time between the first (and last) spikes and the start (or end) of the data collection period do not materially affect our findings.

Measurement of stimulus-dependent clustering

The next step in our analysis is to determine to what extent pairs of responses to the same stimulus tended to be closer to each other (in one of the senses of distance D^{count}, $D^{spike}[q]$, or $D^{interval}[q]$) than pairs of responses to distinct stimuli. That is, in the geometry determined by one of the candidate metrics D^{count}, $D^{spike}[q]$, or $D^{interval}[q]$, to what extent did the observed neural responses show systematic stimulus-dependent clustering? To make as few assumptions as possible about the geometry of the response space, our strategy was to classify responses in a manner that was a direct reflection of the candidate metric—a spike train was assigned to a particular class of responses if it was closer to these responses than to any other set of responses.

More formally, our procedure for classification of a spike train S is as follows. First, temporarily exclude S from the set of N_{tot} observed responses. Let us assume that S is elicited by a stimulus in class s_α. We consider each stimulus class (including s_α) in turn. For each stimulus class s_γ, we determine the average distance from S to each of the spike trains elicited by stimuli of class s_γ. We denote this average distance by $d(S, s_\gamma)$, which is defined as follows

$$d(S, s_\gamma) = [\langle (D[q](S, S')^z \rangle_{S' \text{ elicited by } s_\gamma}]^{1/z} \quad (5)$$

where $\langle\ \rangle$ denotes an average over all spike trains S' elicited by a stimulus in stimulus class s_γ. The spike train S then is classified into the response class r_β for which $d(S, s_\beta)$ is the minimum of all of the averaged distances $d(S, s_\gamma)$, as calculated from *Eq. 5*. Note that distances are averaged in *Eq. 5* following a power transformation (the exponent z). Negative values of the averaging exponent z bias $d(S, s_\alpha)$ toward the shortest distance between S and any response elicited by s_α. We have chosen $z = -2$—loosely corresponding to a "gravitational" attraction—but values of z in the range $[-8, -1]$ lead to substantially similar results. Positive values of the averaging exponent z bias $d(S, s_\alpha)$ toward distances from the outliers and lead to significantly lower degrees of stimulus-dependent clustering.

An interpretation of this procedure is that all of the responses except S are considered to be "training" runs, in which the stimulus is known, and S is classified according to which set of responses it most nearly matches. For binned spike data, this procedure reduces to the ideal decision rule of Geisler et al. (1991), because the probability density of a multivariate Gaussian is a monotonic function of the distance from its peak.

This procedure, applied in turn to each spike train in the data set, subdivides the set of N_{tot} observed spike trains into $N(s_\alpha, r_\beta)$ instances of a response in class r_β occurring in association with a stimulus s_α. That is, $N(s_\alpha, r_\beta)$ initially is set to 0 and then is incremented by 1 for each spike train elicited by s_α that is closer (in the sense of *Eq. 5*) to the spike trains elicited by s_β than to spike trains elicited by any other stimulus class. In case of ties [i.e, k of the distances $d(S, s_\beta)$, $d(S, s_{\beta'})$, ... share the minimum], then each of the $N(s_\alpha, r_\beta)$, $N(s_\alpha, r_{\beta'})$, ... are incremented by $1/k$.

$N(s_\alpha, r_\beta)$ can be considered to be a confusion matrix: the number of times that a stimulus from class α is classified (on the basis of

the neural response) as belonging to class β. If responses were clustered strongly, then $N(s_\alpha, r_\beta)$ would have large values on the diagonal and small values off the diagonal (i.e., stimuli rarely misclassified). If clustering were weak or nonexistent, then $N(s_\alpha, r_\beta)$ would be random and approximately uniformly distributed across the rows and columns of the matrix, subject only to the constraints on the number of stimuli presented from each class. A natural measure of clustering is thus the transmitted information (Abramson 1963) of the matrix $N(s_\alpha, r_\beta)$. The transmitted information H is given by

$$H = \frac{1}{N_{tot}} \sum_{\alpha,\beta} N(s_\alpha, r_\beta) \left[\log N(s_\alpha, r_\beta) - \log \sum_a N(s_a, r_\beta) - \log \sum_b N(s_\alpha, r_b) + \log N_{tot} \right] \quad (6)$$

For C equally probable stimulus classes, perfect clustering [$N(s_\alpha, r_\beta) = 0$ for $\alpha \neq \beta$ and $N(s_\alpha, r_\alpha) = N_{tot}/C$] corresponds to a maximal value of the transmitted information H, namely $\log C$, while random clustering ($N(s_\alpha, r_\beta) = N_{tot}/C^2$) leads to $H = 0$. Note that we use the transmitted information simply as an index of stimulus-dependent clustering. We do not intend to imply that the classification scheme we have used to derive $N(s_\alpha, r_\beta)$ and H reflect neural processes or that the estimate of H in some way corresponds to the information capacity of the neuron. Biological processes might be able to extract either more or less information than the above procedures for determination of distance and classification.

RESULTS

Detailed analysis of individual data sets

We present the analysis of several representative data sets in detail and then describe our findings across the recordings. Response rasters elicited from a small (ca. 3) cluster of neurons in V1 by random textures presented at six contrasts (0.04, 0.08, 0.16, 0.32, 0.64, and 0.96) are shown in Fig. 2*A*. The textures (64 × 64 pixels, 2 × 2 min each) covered the aggregate receptive field of the cluster (0.33 × 0.4 deg). At each contrast, responses to 28 different random textures were collected in randomized order. The elements of our analysis, the spike trains S, are segments of the raster lines shown in Fig. 2*A*. These segments start with the stimulus onset, indicated in Fig. 2*A* by the vertical lines, and extend for 256 ms, which is the duration of the stimulus, as indicated by the solid black bars shown at the bottom of the rasters.

We considered the spike time metric $D^{spike}[q]$ and the spike interval metric $D^{interval}[q]$ for a range of costs (q = 0.25–512 in steps of 2), as well as the spike count metric D^{count}. For each metric type and cost value, we calculated the degree of stimulus-dependent clustering, according to *Eqs. 6* and *5*. The thick lines in Fig. 2, *B* and *C*, show the results of these calculations.

It is well known that the information estimate calculated by *Eq. 6* is spuriously high for finite samples (Carlton 1969; Fagen 1978; Optican et al. 1991). This problem is encountered in vector-space methods because of the sparseness of the sampling of a high-dimensional vector space by the data set (Chee-Orts and Optican 1993) and is analogous to the problem of possible overfitting encountered in neural network methods (Kjaer et al. 1994) for estimation of information. In our approach, the extent of this upward bias might depend on the nature of the distance or on the parameter q. Recently, Treves and Panzeri (1995) have found an analytic approximation to this bias. However, a fundamental requirement of their approach is a stage in which each response is assigned to "bins" in a manner independent of all other responses. This is precisely the step that the metric-space approach bypasses, and our clustering algorithm features a strong and explicit interaction between responses. Thus this analytic approximation is not directly applicable to the present method. Indeed, were we to apply the Treves and Panzeri (1995) approximation, then the resampled estimates would all be independent of q (i.e., flat lines). This is true in some cases (Figs. 2, *B* and *C*, and 4, *B* and *C*), but not in others (Fig. 3, *B* and *C*).

For this reason, we took a more empirical approach. We estimated the values of H that would be expected from chance clusterings alone by repeating the above calculations for synthetic data sets, which consisted of a random reassignment of the observed responses to the stimulus categories. The results of these calculations (mean ± 2 SE as derived from 10 resamplings) are shown by the thin lines in Fig. 2, *B* and *C*. As is seen for either $D^{spike}[q]$ or $D^{interval}[q]$, there is little dependence of chance values of H on q for both classes of distances. This is important, because it indicates that the dependence of H on q and on the class of distance that we observe for the original (unresampled) dataset reflects structure in the data itself rather than intrinsic properties of the distances.

We caution the reader that the ±2 SE lines in Fig. 2, *B* and *C*, represent confidence limits for the mean of the resampled values and are only appropriate for comparison among different resampled estimates. To estimate the probability that the measured (unresampled) value of H could have come from this population, a ±2 SD criterion should be used. Because we used 10 resamplings, the standard deviation is $\sqrt{10} \approx 3.2$ times the SE.

For the spike count metric (which corresponds to q = 0 for either $D^{spike}[q]$ or $D^{interval}[q]$), there was significant stimulus-dependent clustering, as seen by a calculated information H = 0.36 bits, approximately twice that obtained from the resampled dataset. This is not at all surprising; it is simply a nonparametric measure of the fact that this neuron cluster's overall firing rate depended systematically on stimulus contrast. The behavior of information H calculated from $D^{spike}[q]$ and $D^{interval}[q]$ for $q > 0$ reveals to what extent there is a systematic dependence of the temporal firing pattern on contrast. For $D^{spike}[q]$, H rises to a maximum value of ~1.0 bits at q = 64. Recall that, according to the metrics considered here, a shift in time by an amount $2/q$ is equivalent in cost to deleting the spike altogether and reinserting it elsewhere. Thus $2/q$ reflects the temporal blur of the metric. Spikes that differ in time by more than this amount are regarded as different, whereas spikes separated by this amount or less are regarded as progressively more similar. Thus the peak of H at q = 64 means that, to within a precision of ~31 ms (0.031 ≈2/64), the time of arrival of spikes depends systematically on stimulus contrast. In other words, a decoding mechanism that is sensitive not only to the number of spikes but also to their time of occurrence (with a precision in the range of 32 ms) could, in principle,

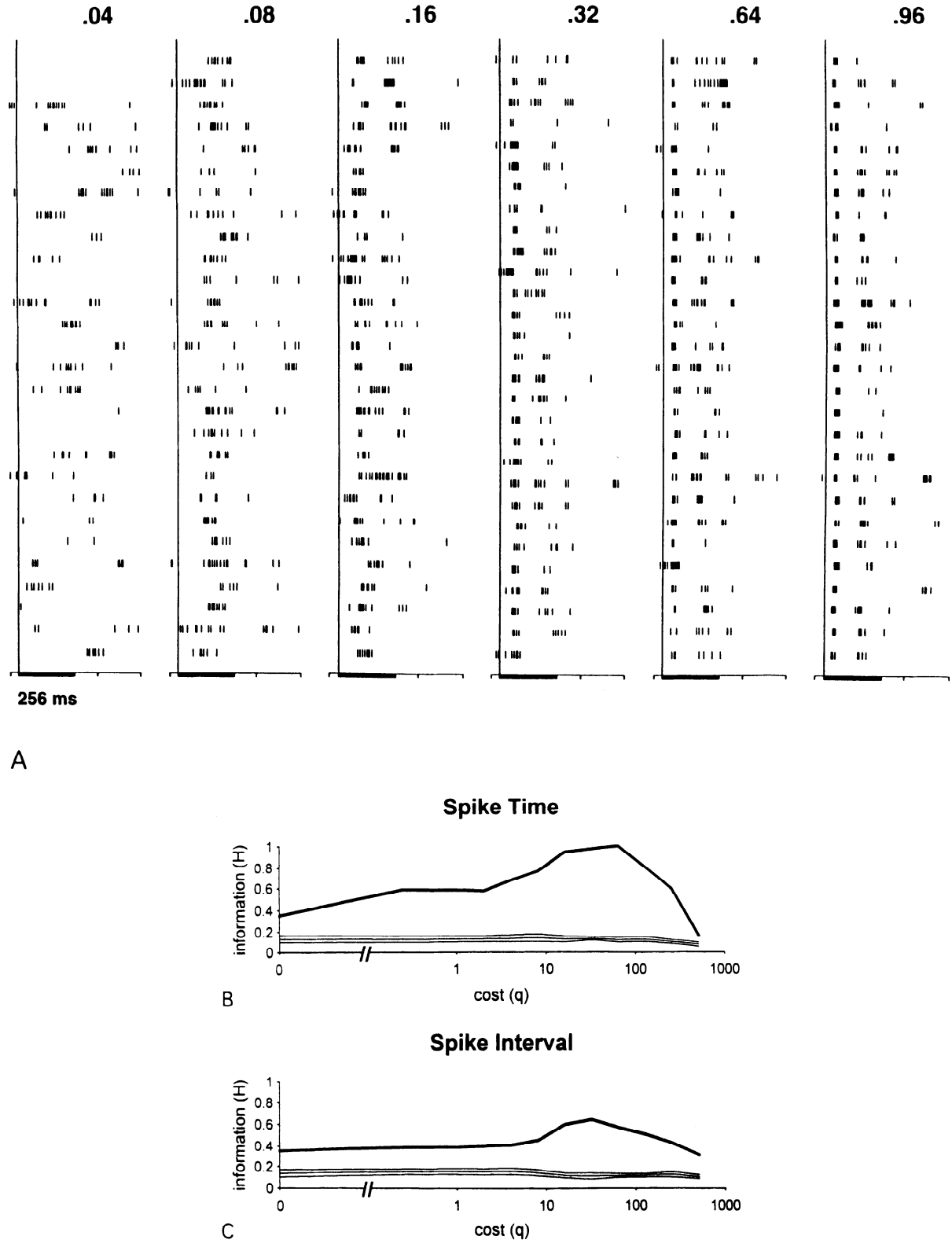
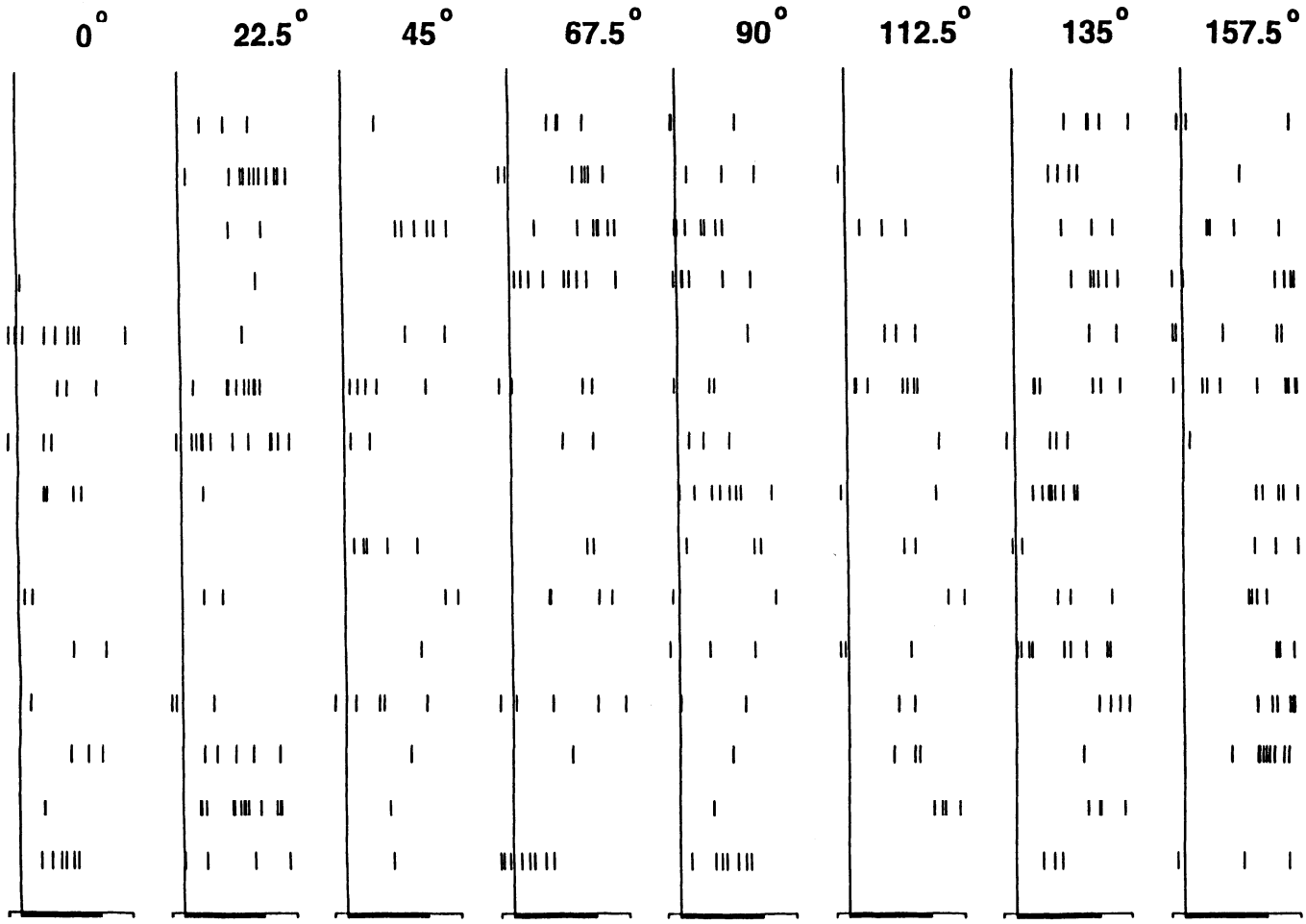
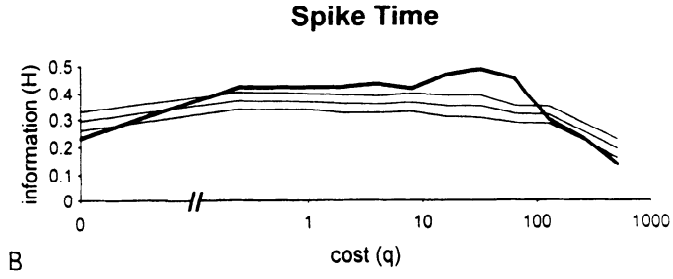


FIG. 2. *A*: responses elicited from a cluster of neurons in V1 by random textures (64 × 64 pixels). Contrasts: 0.04, 0.08, 0.16, 0.32, 0.64, and 0.96; check size: 2 × 2 min. Thick horizontal bar at the bottom of each set of rasters marks period in which stimulus was presented. In some response rasters, ticks representing spikes are confluent, indicating a burst. *B*: analysis of stimulus-dependent clustering induced by $D^{spike}[q]$. Thick line: information-theoretic measure H of clustering derived from rasters of *A*. Thin lines: mean plus and minus two standard errors of distribution of values of H calculated from synthetic data sets in which observed responses are reassigned randomly to stimuli. *C*: analysis of stimulus-dependent clustering induced by $D^{interval}[q]$. Thick and thin lines as in *B*. Note that ±2 SE lines are appropriate for comparison among different resampled estimates. To estimate probability that measured (unresampled) value of H could have come from this population, a ±2 SD criterion should be used. SD is $\sqrt{10} \approx 3.2$ times SE; recording H30031.





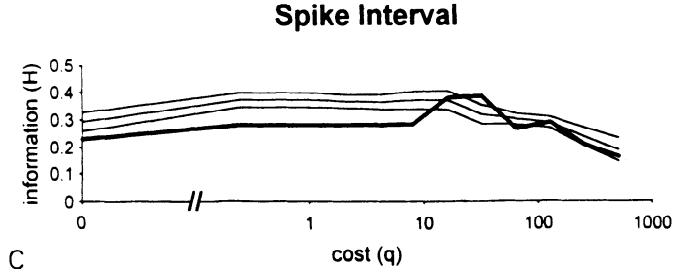


FIG. 3. *A*: response rasters elicited from an isolated neuron in V1 by gratings presented at 8 orientations (0–157.5 deg in 22.5-deg steps). Contrast, 1.0; Spatial frequency, 21 c/deg. In some response rasters, ticks representing spikes are confluent, indicating a burst. *B*: analysis of stimulus-dependent clustering induced by $D^{spike}[q]$. Thick line: information-theoretic measure H of clustering derived from the rasters of *A*. Thin lines: means ± 2 SE of distribution of values of H calculated from artificial data sets in which observed responses are reassigned randomly to stimuli. *C*: analysis of stimulus-dependent clustering induced by $D^{interval}[q]$. Thick and thin lines as in *B*; recording H28031.

provide a more informative representation of stimulus contrast than one that merely counted spikes. However, the time of occurrence of spikes depends on the stimulus only up to a limited precision. For higher values of q, H descends, which indicates that the timing of individual spikes does not systematically depend on the stimulus at higher spike time resolutions. One expects that for sufficiently high q, H will eventually become lower than the value measured for the spike count metric. This is because it is biologically implausible that the precise (e.g., on a microsecond scale) time of spikes conveys stimulus-specific information, and thus a metric that regarded two spikes as unrelated, even if their times of occurrence differed by minuscule amounts, would not be likely to represent a stimulus. In this dataset, H, as measured from $D^{spike}[q]$, descends below the value measured for the spike count metric at $q = 512$. This means that a decoding mechanism that attached significance to spike times with a 4-ms precision ($0.004 \approx 2/512$) would be inferior to a decoding mechanism that simply counted spikes. We will use q_{max} to denote the value of q for which H reaches its maximum value and, given the rationale described above, consider $2/q_{max}$ to be an index of the precision of the temporal code.

The analysis for $D^{interval}[q]$ reveals a similar, but less marked, dependence on the cost q. The maximum value achieved by H is 0.66 bits (at $q_{max} = 32$), and H descends below the value measured for the spike count metric at $q = 512$. This means that a decoding mechanism that is sensitive to the pattern of spike intervals with a precision in the range of 63 ms ($0.063 \approx 2/32$) could, in principle, provide a more precise representation of stimulus contrast than one that merely counted spikes. However, the gains in information seen for the spike interval metric are less marked than the gains in information seen for the spike time metric.

Response rasters elicited from an isolated neuron in V1 by gratings presented at eight orientations (0–157.5 deg in 22.5-deg steps) are shown in Fig. 3*A*. The gratings had a contrast of 1.0 and a spatial frequency of 21 c/deg, and covered the unit's receptive field (0.59 × 0.46 deg). Responses (15) to each of the eight orientations were collected in randomized order. The spatial frequency optimum for this neuron, as measured by the number of spikes elicited, was 5 c/deg. The spatial frequency studied in Fig. 3 is far from the "best" spatial frequency in terms of spike counts. There was no significant stimulus-dependent clustering in the spike count metric ($q = 0$), as is seen from the fact that the calculated values of information were not higher than what would be expected from chance alone. However, for $q > 0$, there was a progressive increase in the information-theoretic measure of clustering, with H reaching a maximum value of 0.49 bits for $D^{spike}[q]$ and 0.39 bits for $D^{interval}[q]$, both at $q_{max} = 64$. For $D^{spike}[q]$, this was significantly higher than the distribution of values obtained from the resampled data sets, but for $D^{interval}[q]$, the peak information value was within two standard deviations of this distribution. This dataset shows that the spike time metric can reveal significant stimulus-dependent clustering, even when there is little evidence of tuning in the traditional spike-count sense.

Response rasters elicited from an isolated neuron in V3 by even and odd isodipole textures are shown in Fig. 4*A*. Textures had a contrast of 0.32 and a check size of 4 × 4 min and covered nearly all of the unit's receptive field (4.9 × 4.5 deg). Responses to 32 different examples of each texture were collected in randomized order. The spike count metric ($q = 0$) showed some significant stimulus-dependent clustering ($H = 0.14$ bits, small but far in excess of the value of 0.03 bits that would be expected from chance alone). For $D^{spike}[q]$ (Fig. 4*B*), this measure was substantially larger (0.24 bits) at $q_{max} = 16$. For $D^{interval}[q]$ (Fig. 4*C*), H remained above that expected for chance alone, but not significantly so.

These three detailed analyses all reveal a rise in the information-theoretic measure of clustering H for moderate values of $q > 0$ for $D^{spike}[q]$, with q_{max} in the range 16–64, followed by a descent below the clustering seen for the spike count metric at still higher values of q (typically >128). For $D^{interval}[q]$, similar behavior is seen, but the increase in H is typically less marked and less often significant (as judged by a comparison with resampled data sets).

Summary across all data

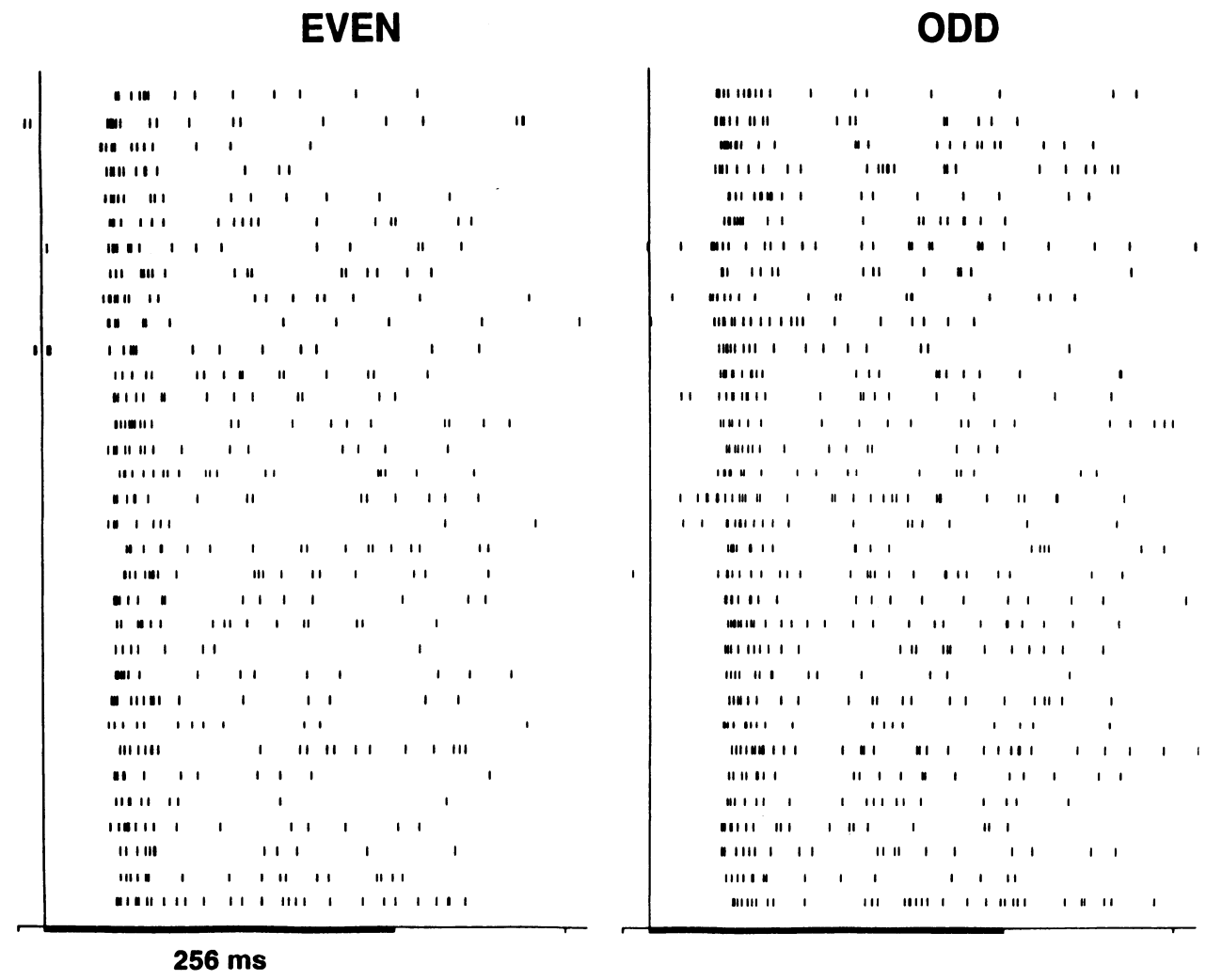
In all, we performed the above calculations on 352 data sets (Table 1). In 170 of these data sets, the information-theoretic measure of clustering was greater than chance for the spike count ($q = 0$) metric, and the average value of H was 0.176 bits across all data sets. When analyzed by the spike time metric, 294 of the 352 data sets led to a greater-than-chance value of H, and the average value of H was 0.287. For the spike interval metric, 272 of the 352 data sets showed a greater-than-chance value of H, and the average value of H was 0.227 bits. Similar results were obtained when the analysis was restricted to the first 100 ms after stimulus onset: there was significant clustering in 151 data sets via the spike count metric, 270 data sets via the spike time metric, and 232 data sets via the spike interval metric.

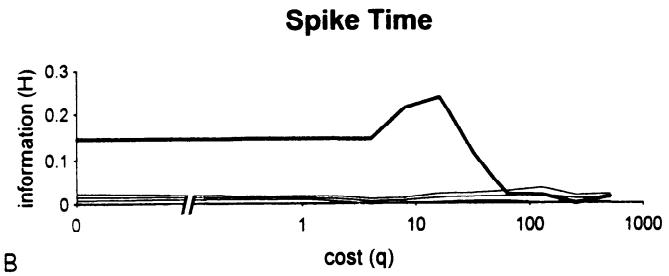
Of the 170 data sets in which neural responses showed significant tuning based on spike counts (over the full 256-ms stimulus period), 113 showed evidence of additional temporal coding in the spike time metric and 109 showed evidence of additional temporal coding in the spike interval metric.

Across the entire collection of data sets, the information not attributable to chance clustering averaged 0.042 bits for the spike count metric, 0.171 bits for the optimal spike time metric, and 0.107 bits for the optimal spike interval metric. There was a very large amount of scatter in the estimated information values, even when normalized for experimental design (see Fig. 5 below), likely reflecting the presence of many data sets for which there was little neural response.

We did not note any systematic difference between analyses of single- and multi-unit data, but this may well be related to our recording method, which does not label spikes according to neuron of origin. The manner in which spike trains from multiple neurons combine to transmit information is a matter of great interest, but not one that we address here.

Table 2 shows a breakdown of the temporal coding analysis by stimulus modality. The table is restricted to recordings from V1 and V2 because of the small number of experiments performed in V3. Because we did not find significant or consistent differences in comparing analyses from single-





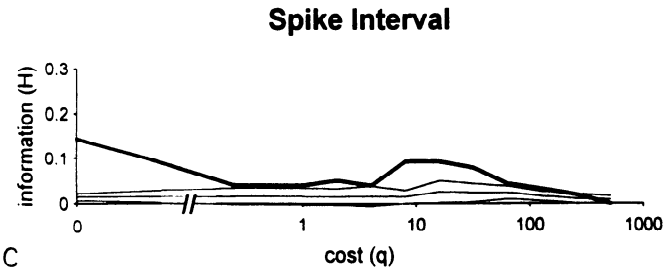


FIG. 4. *A*: response rasters elicited from a V3 neuron by even and odd isodipole textures. Contrast, 0.32; check size, 4 ×4 min. In some response rasters, ticks representing spikes are confluent, indicating a burst. *B*: analysis of stimulus-dependent clustering induced by $D^{spike}[q]$. Thick line: information-theoretic measure H of clustering derived from rasters of *A*. Thin lines: means ± 2 SE of distribution of values of H calculated from artificial data sets in which observed responses are reassigned randomly to stimuli. *C*: analysis of stimulus-dependent clustering induced by $D^{interval}[q]$. Thick and thin lines as in *B;* recording: H14071.

TABLE 2. *Fraction of data sets in which the information-theoretic measure of clustering H was greater than by chance*

	Fraction of Data Sets with Greater-Than-Chance Values of H					
	Contrast	Check size	Orientation	Spatial frequency	Texture type	All
Definite V1						
Spike count	0.73		0.65	0.72	0.20	0.55
Spike time	0.93		0.80	0.88	0.53	0.76
Spike interval	0.87		0.80	0.84	0.47	0.72
Definite V2						
Spike count	0.44	0.38	0.44	0.50	0.25	0.43
Spike time	1.00	1.00	0.92	0.89	0.75	0.92
Spike interval	0.89	0.92	0.94	0.83	0.83	0.89
Total V1 and V2						
Spike count	0.71	0.38	0.52	0.60	0.22	0.48
Spike time	0.98	1.00	0.88	0.88	0.60	0.83
Spike interval	0.90	0.92	0.89	0.84	0.55	0.79

and multi-unit recordings, they are pooled. In all cases, analysis of neural responses via the spike time metric and the spike distance metric showed evidence of tuning (i.e., greater-than-chance values of clustering, as measured by H) more often than did analyses via the spike count metric. The largest differences were seen for the texture experiments, in which the number of recordings that would have been classified as "tuned" more than doubled. There was little consistent difference in the number of recordings that were tuned as classified by the spike time metric versus the spike interval metric.

Figure 5 shows a more detailed comparison of the extent of stimulus-dependent clustering by the spike time and the spike interval metric. All V1 and V2 experiments are presented. To compare values of H across modalities, we have normalized their values by dividing them by $\log_2(N)$, where N is the number of conditions in each experiment. (Thus the maximum value of this normalized H in any dataset is 1, independent of the number of conditions). There is a clear tendency for recordings to show greater evidence of tuning in the spike time metric, for contrast ($P < 0.00001$), check size ($P < 0.006$), orientation ($P < 0.004$), and spatial frequency ($P < 0.0001$) (two-tailed paired t-test for each modality). However, there was no statistically significant difference between the clustering induced by these metrics in the texture-type experiments ($P \approx 0.9$, two-tailed paired t-test).

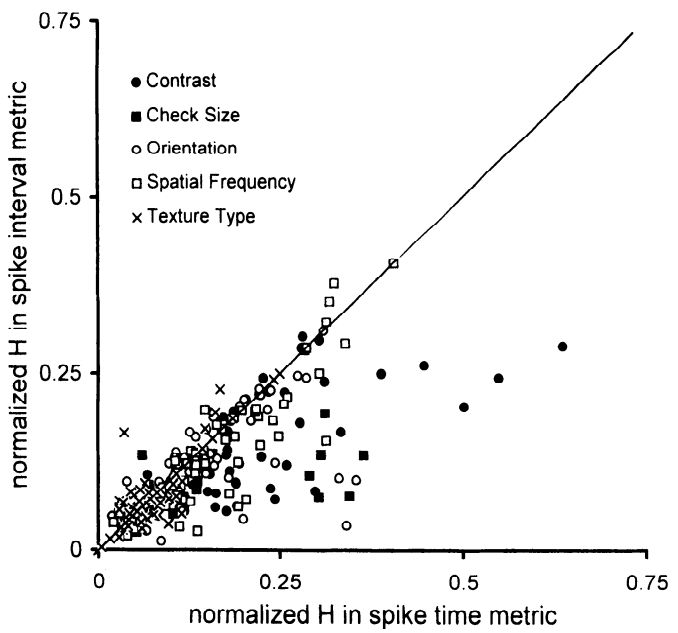


FIG. 5. A comparison across stimulus modalities of clustering induced by the spike time metric $D^{spike}[q]$ and the spike interval metric $D^{interval}[q]$ for all recordings in V1 and V2. The information-theoretic measure of clustering H has been normalized by logarithm of number of stimuli in each data set, so that maximum possible value of H is 1.

Figure 6 compares the temporal precision of the optimal spike time metric (Fig. 6, *A* and *B*) and the optimal spike interval metric (Fig. 6, *C* and *D*) across stimulus modality for recordings in V1 and V2. For both metrics, contrast is encoded with the highest temporal precision and texture type and spatial frequency are encoded with the lowest temporal precision. Both V1 and V2 show these trends, but the trend is stronger in V1 for the spike interval metric (compare Fig. 6, *A* with *C*) and is stronger in V2 for the spike time metric (compare Fig. 6, *B* with *D*). Note that precision, as measured by q_{max}, is not the degree of stimulus-dependent clustering or the amount of information transmitted, but rather, it is a characterization of the nature of the coding. For example, the fivefold higher value of q_{max} for D^{spike} from V1 to V2 does not imply that a fivefold greater amount of information is present, but rather that the structure of the code has changed to one that relies more critically on spike times. Similarly, even though q_{max} is higher in V1 for $D^{interval}$ than for D^{spike}, the amount of information transmitted by spike-time encoding is greater than that transmitted for spike interval coding.

Poisson hypothesis

Perhaps the above results could be accounted for by the hypothesis that neurons fired in a Poisson fashion with the rate of the Poisson process varying in time in a stimulus-dependent manner. If this were the case, this hypothetical envelope function (the underlying stimulus-dependent function that governs the mean rate) would be the source of temporal coding, and although the time of occurrence of individual spikes might convey information, individual spike times would be significant only to the extent that they helped to establish this envelope.

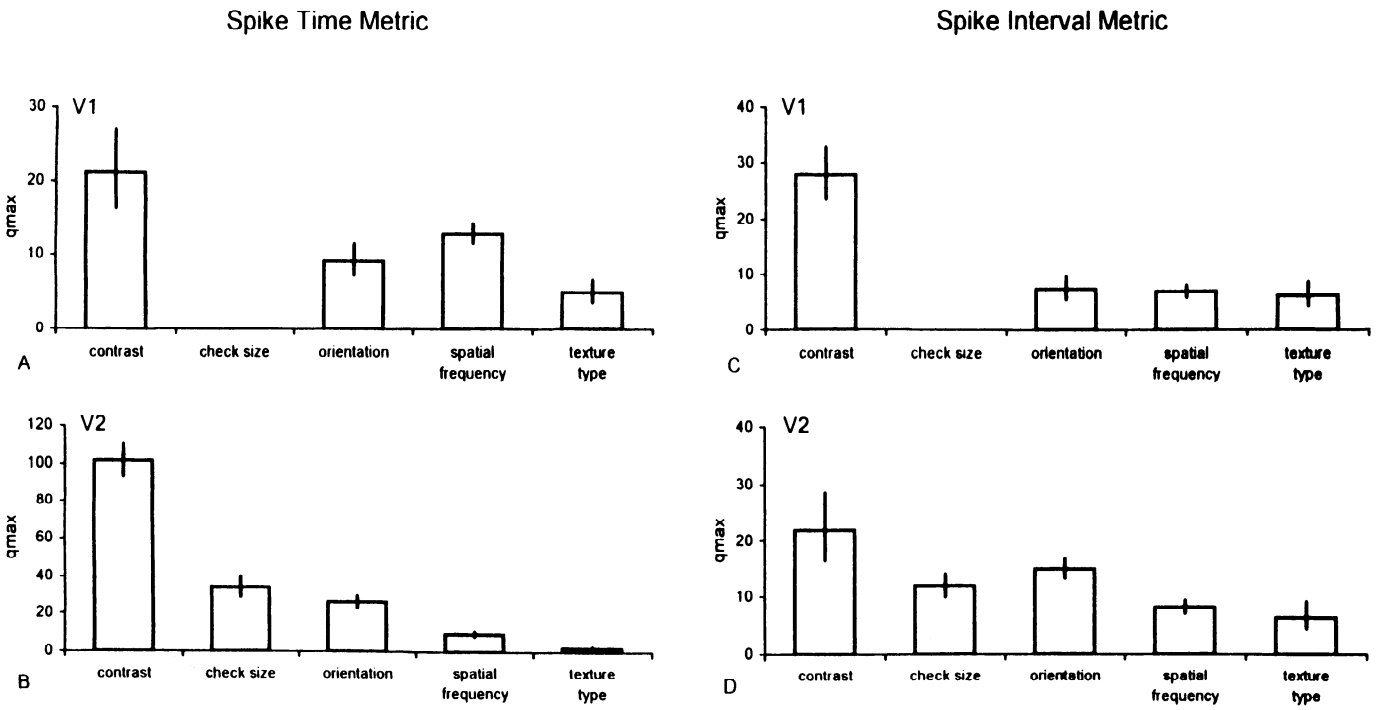
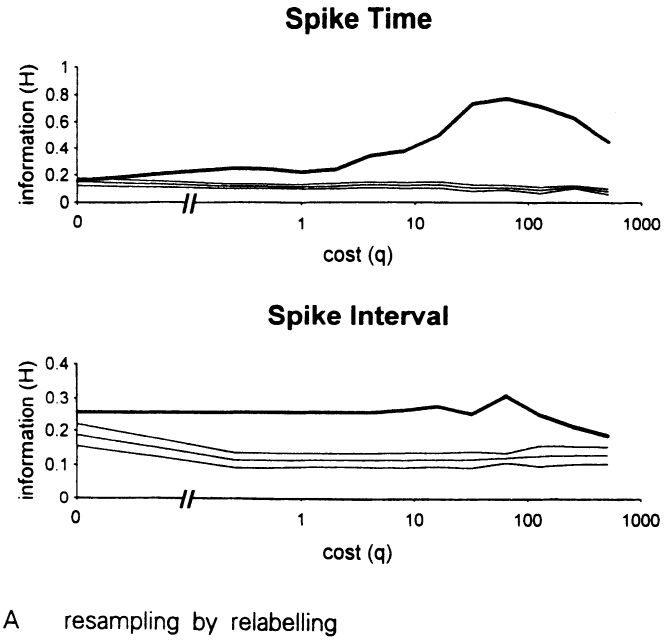


FIG. 6. A comparison across stimulus modalities of temporal precision q_{max} for maximal clustering. Geometric means and SE (following logarithmic transformation) are shown. Recording sites shown in V1 (A) and V2 (B) as clustered by spike time metric $D^{spike}[q]$ and in V1 (C) and V2 (D) as clustered by spike interval metric $D^{interval}[q]$.

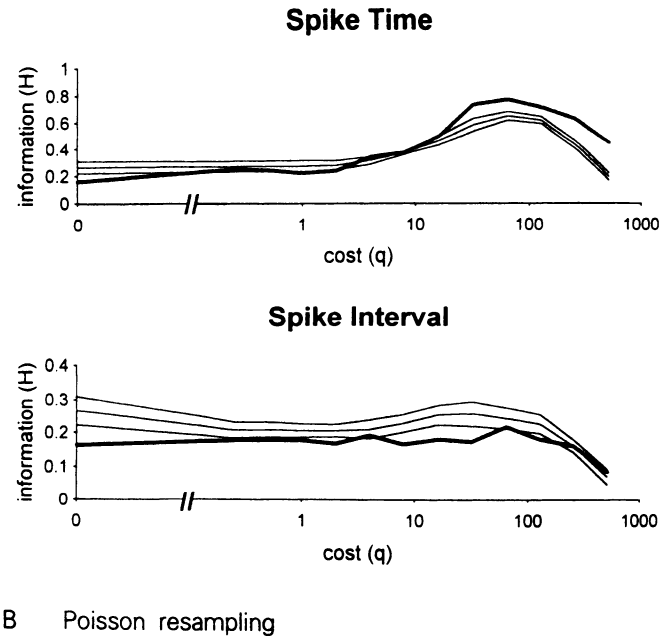
To address this possibility, we repeated the above metric-space analysis on synthetic data derived from experimental data in the following fashion. We begin (for example) with the contrast dataset illustrated in Fig. 7A, in which we recorded 30 responses of a V2 neuron to random textures presented at each of six contrasts. For each of the six contrasts, the times of occurrence (relative to stimulus onset) of all recorded spikes were placed in a single set. Then, each of these spike times was assigned randomly (without repetition) to any of 30 responses. This maneuver preserves the average poststimulus time histogram (PSTH), and, indeed, uses the PSTH as the best estimate of the underlying Poisson process. Thus if spikes were indeed generated in a time-dependent Poisson fashion, the synthetic data sets obtained by shuffling spike times among the responses should demonstrate the same kind of temporal coding as the original data. Results of the Poisson resampling analysis are shown in Fig. 7B. The synthetic data sets indeed show an increase in information for $D^{spike}[q]$ and a similar value of q_{max}. However, there are two discrepancies between the actual data and the synthetic data. The first is that for $q > 32$, the synthetic data sets show less evidence of temporal coding via the spike time metric as manifest by a lower value of H. For example, for $D^{spike}[64]$, H was 0.770 bits for the original data, but 0.650 $\pm$ 0.096 bits (mean $\pm$ 2 SD). This indicates (at least for this data set) that the degree of clustering seen in the original data is more than can be explained by rate-dependent Poisson firing. While it could be argued that our use of the PSTH as an estimate of the underlying envelope necessarily leads to an estimate of the histogram whose peaks are too sharp, any smoothing of this histogram would be tantamount to using a lower value of q.

The second discrepancy is that for low values of q, the resampled data sets show a higher level of stimulus-dependent clustering as manifest by a larger value of H calculated from the synthetic data sets. In this data set, for $D^{spike}[0] = D^{count}$, H was 0.163 bits for the original data, but 0.266 $\pm$ 0.132 bits (mean $\pm$ 2 SD) for the synthetic data sets. This increase in information with resampling is not statistically significant by itself, but a similar trend, significant at $P < 0.001$, was seen across all the single- and multi-unit data (see Table 3 and discussion below).

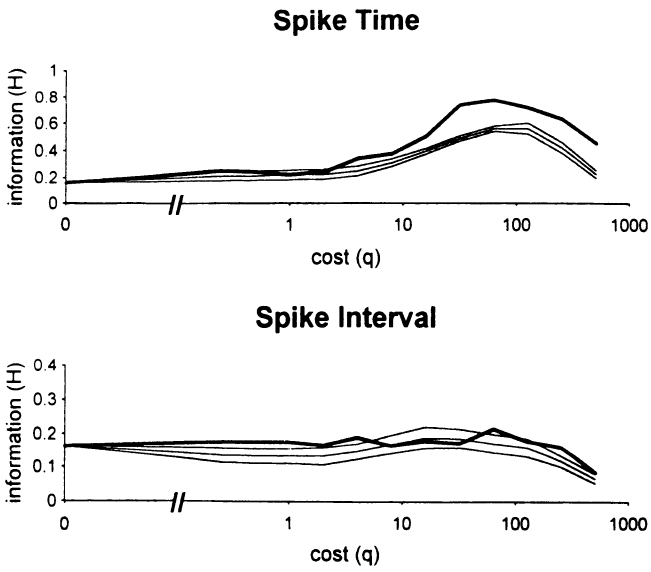
This observation is at first counterintuitive because it indicates that the application of a stimulus-independent randomization of the data results in an increase in the amount of apparent information. However, one effect of this randomization is that it forces the spike count statistics to be Poisson. In the real data, the distribution of the number of spikes elicited by each presentation of a stimulus has a higher variance than a Poisson process of equal mean—in the data set illustrated in Fig. 7, the variance of the spike counts recorded in response to the six contrasts ranged from 1.6 to 2.7 times the mean. Because the resampling procedure forces the spike count statistics to be Poisson but leaves unchanged the mean number of spikes elicited by each stimulus type, the original data set will have a larger variance than the synthetic data sets. If this more-than-Poisson variance is due to influences unrelated to the stimulus, then its removal will increase the amount of information. Indeed, we see that the synthetic data sets have a larger value of H not just for the spike count code ($D^{spike}[q]$ for $q = 0$), but also for those spike time codes dominated by the spike count component (q close to 0). Additionally, analysis of $D^{interval}[q]$ (lower portion of Fig. 7B) revealed a reduction in H for all values of q, implying that the more-than-Poisson variance of the original data set was the dominating influence for this family of metrics.



A resampling by relabelling



B Poisson resampling



C exchange resampling

FIG. 7. *A*: analysis of stimulus-dependent clustering induced by $D^{spike}[q]$ and $D^{interval}[q]$ for responses of a single neuron in V2 to random textures (64 × 64 pixels). Contrasts: 0.04, 0.08, 0.16, 0.32, 0.64, and 0.96; check size: 4 × 4 min. Thick line: H calculated from observed responses. Thin lines: means ± 2 SE of distribution of values of H calculated from synthetic data sets in which responses were reassigned arbitrarily to stimuli. *B*: comparison of H calculated from observed responses (thick lines, as in *A*) with distribution of values of H calculated from synthetic data sets in which spikes are assumed to be generated by a time-dependent Poisson process whose poststimulus time histogram (PSTH) for each condition is identical to that of data ("Poisson resampling"—see text for further details). *C*: comparison of H calculated from observed responses (thick lines, as in *A*) with distribution of values of H calculated from synthetic data sets whose PSTH for each condition is identical to that of data and whose spike count statistics are identical to that of data ("exchange resampling"—see text for further details). As in Figs. 2–4, the ±2 SE lines are appropriate for comparison among different resampled estimates. To estimate probability that measured (unresampled) value of H could have come from this population, a ±2 SD criterion ($\sqrt{10} \approx 3.2$ times SE) should be used.

This analysis suggests another hypothesis: perhaps firing times are governed in part by a time-dependent Poisson process, but this spike rate envelope is modulated (i.e., multiplied) by a slowly varying influence unrelated to the stimulus. This additional factor would account for the more-than-Poisson variance we observe and perhaps account for the clustering detected by $D^{spike}[q]$ or $D^{interval}[q]$. To test this hypothesis, we introduced a second statistical transformation of the data. We randomly selected two spikes from different responses to the same stimulus and exchanged them so that the spike that actually occurred in one response was now placed into the other response and vice versa. Many such exchanges (more than 4 times the number of spikes) were performed so that the particular responses in which each spike occurred were scrambled thoroughly. However, because each exchange introduces no new spike times and also results in no net change in the number of spikes in each response, it retains not only the observed PSTH but also the observed spike count statistics. Under the modified Poisson hypothesis, synthetic data generated by exchange resampling should behave in a manner identical to the original data.

Figure 7*C* shows the analysis of exchange resampling applied to this dataset. For the spike count metric ($q = 0$), H must be unaffected by the exchange resampling, because the number of spikes in each response are unchanged. For $D^{spike}[q]$, there is significantly more clustering in the actual data set than in the synthetic data sets for $q \geq 16$. For $D^{interval}[q]$, the level of stimulus-dependent clustering identified in the actual data set is similar to that in the synthetic data sets.

Table 3 shows a summary of this analysis across all of the data sets. The reduction in spike count variance due

TABLE 3. *Effects of Poisson resampling and exchange resampling on the information-theoretic measure of clustering H*

	n	Original Data Mean H	Poisson Resampling			Exchange Resampling		
			Mean H	Difference	P	Mean H	Difference	P
Single-unit data	152							
Spike count		0.158	0.210	0.052	<0.001	0.158	†	†
Spike time		0.299	0.286	−0.014	0.083	0.247	−0.052	<0.001
Spike interval		0.215	0.202	−0.013	0.111	0.168	−0.047	<0.001
Multi-unit data	201							
Spike count		0.189	0.267	0.078	<0.001	0.189	†	†
Spike time		0.278	0.300	0.021	0.027	0.249	−0.029	<0.001
Spike interval		0.236	0.255	0.020	0.020	0.203	−0.032	<0.001
Pooled	352							
Spike count		0.176	0.243	0.067	<0.001	0.176	†	†
Spike time		0.287	0.294	0.006	0.336	0.249	−0.039	<0.001
Spike interval		0.227	0.232	0.005	0.367	0.188	−0.039	<0.001

Levels of significance were calculated by paired t-tests within data sets. † For these conditions, the shuffling operation preserves H exactly.

to Poisson resampling resulted in a greater information-theoretic measure of clustering in the synthetic data with the average H rising from 0.176 bits (pooled data) to 0.243 bits (Poisson resampled data) for D^{count}. There was no change, on the average, in the maximum information calculated from $D^{spike}[q]$ or $D^{interval}[q]$. This appears to be the result of an overall increase in H associated with the decrease in spike count variance and a decrease in H near q_{max}. The exchange resampling process, which eliminates the change in spike count statistics, leads to synthetic data sets in which clustering is significantly reduced, for both $D^{spike}[q]$ and $D^{interval}[q]$.

The rise in H with Poisson resampling, and the fall in H with exchange resampling, was seen for all stimulus modalities and in all cortical regions studied. However, there was a suggestion of a significant difference between the single- and the multi-unit recordings. For the single-unit recordings, Poisson resampling led to a small and borderline-significant reduction in H for $D^{spike}[q]$ and $D^{interval}[q]$ despite the increase in H for D^{count}. For the multi-unit data, there was an increase in H for Poisson resampling for all three metrics, although this increase was much greater for D^{count}. Thus in single unit responses, the temporal structure produces an increase in H that is larger than the penalty incurred by the large variance (as is illustrated for $D^{spike}[q]$ in Fig. 7*B*), but this temporal structure is less evident in the multi-unit responses because of the intermingling of responses from multiple units without separate identification. That is, in addition to the more-than-Poisson scatter in spike counts, there is internal structure in the discharges of single units, which is masked when multiple single-unit discharges are superimposed.

In summary, a Poisson model cannot account for our observations: if spike count statistics are assumed to be Poisson across trials, then H is overestimated for D^{count}; if empirical spike count statistics across trials are preserved so that D^{count} is accounted for, then H is underestimated for $D^{spike}[q]$ and $D^{interval}[q]$.

DISCUSSION

The aim of this paper is to learn about the nature and precision of temporal coding in the primate visual cortex. The approach is based upon introducing several notions of the "distance" between two spike trains (D^{count}, $D^{spike}[q]$, and $D^{interval}[q]$) and a determination of the extent to which, in the sense of these distances, the observed neural responses cluster in a stimulus-dependent fashion. D^{count} corresponds to the traditional notion of a spike count, or spike rate, code: the "distance" between two spike trains is the difference in the number of spikes. The other notions of distance, $D^{spike}[q]$ and $D^{interval}[q]$, are sensitive to the temporal structure of the response—the time of occurrence of the impulses ($D^{spike}[q]$) and the pattern of interspike intervals ($D^{interval}[q]$). The cost parameter q in the metrics $D^{spike}[q]$ and $D^{interval}[q]$ expresses the precision of temporal coding, in that spike times, or spike intervals, separated by $>2/q$ are regarded as "different", whereas spike times or spike intervals separated by $\leq 2/q$ or are regarded as similar.

One major distinction between the present approach and most previous approaches is that there is no reliance on embedding responses in a vector space and no assumption of a Euclidean (or near-Euclidean) geometry for the set of spike trains. This strategy has both practical and theoretical advantages. On a practical level, the dimension of the embedded vector space is equal to the duration of each data segment divided by the width of the time bin. Thus a vector-space embedding that retained even 10-ms resolution would be of very high dimension and necessarily would be only very sparsely populated by the data. There are computationally intensive algorithmic solutions available that confront the problem of sampling in high-dimensional vector spaces (Chee-Orts and Optican 1993), but the present approach enables us to examine the data with high precision directly and with a lower computational burden. On a theoretical level, the metric-space approach avoids imposing a priori notions of the geometry of the response space, which, as has been emphasized recently (Hopfield 1995), is unlikely to be Euclidean.

Qualitative and quantitative features of temporal codes in macaque visual cortex

Our analysis of neural responses in macaque V1 and V2 demonstrates that each of the stimulus attributes studied (contrast, check size, orientation, spatial frequency, and texture) systematically change not only the absolute number of

spikes, but also their temporal pattern. Indeed, ~30% of recordings (Table 2) would be regarded as showing a lack of dependence on the stimulus attribute under study if one only considered spike count but demonstrated substantial tuning when temporal pattern was taken into consideration. In most recordings, the time of occurrence of spikes, rather than the interspike intervals per se (see below), showed the most systematic dependence on stimulus attributes (Fig. 5).

Although strong evidence for temporal coding was seen for all stimulus attributes, there were significant differences in the nature of this dependence, as characterized by q_{max}, the value of the cost parameter q that maximizes stimulus-dependent clustering. As seen in Fig. 6, contrast was associated with the highest value of q_{max} and texture was associated with the lowest values of q_{max}. This does not necessarily mean that temporal pattern plays a greater role in signalling contrast than texture. Rather, it means that for contrast, the systematic dependence of spike (or interval) times on the stimulus is manifest with a greater precision (10–30 ms) than for texture (ca. 100 ms); precision for the other modalities studied was between these extremes. Heller and colleagues (Heller et al. 1995) obtained a similar value (ca. 25 ms) for overall precision of temporal coding in response to Walsh patterns but did not identify a dependence of precision on stimulus attribute. Additionally, Gawne and coworkers (1991), in LGN recordings, found evidence for temporal multiplexing of stimulus attributes. They observed, via principal components analysis, that luminance was selectively encoded in the transient component of the response. This is analogous to our findings, but distinct in two ways: we examined contrast, not luminance, and we examined temporal precision, not response waveform.

This finding suggests the possibility that multiple modalities can be represented simultaneously in a spike train with some degree of independence–the firing pattern, viewed with high temporal resolution, might represent contrast, while the same pattern, viewed with a substantially lower resolution, might represent texture or another correlate of visual form. This conclusion is supported by a study of steady state field potential recordings in the primary visual cortex of the anesthetized, paralyzed macaque with a completely different stimulation and analysis strategy (Victor et al. 1994). Fourier components of the population response to abrupt pattern interchange carried contrast information at high harmonics. Spatial frequency and orientation information were carried at somewhat lower harmonics, and texture information was carried at the lowest harmonics. Yet a third line of evidence for temporal coding in primary visual cortex recently has been obtained from intracellular recordings in cat visual cortex (Volgushev et al. 1995): although the preferred orientation was unchanged in the first 100 ms after stimulus presentation, orientation tuning generally increased with time. This finding suggests that the circuitry within primary visual cortex leads to a neural response whose initial transient is influenced more strongly by contrast and whose subsequent components are influenced more strongly by orientation—i.e., multiplexing of contrast and form information. A similar conclusion was reached by an analysis of temporal coding in spike trains based on a vector-space embedding (Purpura et al. 1993). Together, these observations extend the original identification of temporal coding in extrastriate visual cortex (Optican and Richmond 1987; Richmond et al. 1987) to the earliest stages of cortical processing.

Not consistent with Poisson spike trains

It might be suggested that our findings could be explained on the basis of neurons that fire in a Poisson fashion, but the rate of this Poisson process varies with time in a stimulus-dependent manner. This view permits temporal multiplexing—for example, different modalities could be represented in different frequency bands or principal components of the envelope. In this view, spike times carry information, but only insofar as they report the shape of the underlying envelope. Furthermore, the Poisson model would account for the finding that clustering in $D^{spike}[q]$ typically exceeded clustering in $D^{interval}[q]$ (Fig. 5 and Table 2), because in a variable-rate Poisson model, spike times have significance in that they serve to mark peaks in the underlying rate envelope, but spike intervals have significance only inasmuch as they are consequences of neighboring peaks in the envelope.

However, a more detailed analysis (Fig. 7 and Table 3) indicates that time-dependent Poisson firing cannot account for our observations: Poisson models cannot simultaneously account for the variability of the spike counts and the high degree of stimulus-dependent clustering seen at q_{max}. Thus one cannot regard the time of occurrence of individual spikes simply as "a random instantiation of the average synaptic activity" (Shadlen and Newsome 1995), and one is forced to the alternative that patterns in the spike train are both reproducible and stimulus dependent.

Decoding and propagation

In contrast to studies of the processing of temporal patterns in the sensory input (Bialek et al. 1991), the temporal pattern in the response to a simple stimulus appearance does not directly represent the input, and thus presumably, needs to be decoded. One can take the view that a decoding operation, as such, is not necessary for perception because the similarities and differences in responses, by themselves, represent the perceptual space (Edelman and Cutzu 1995). Nevertheless a kind of decoding is necessary to propagate information from one set of neurons to another. Our experiments do not address how temporal patterns in spike trains might be propagated or decoded, but work by a number of investigators indicates that the cortical circuitry indeed possesses the requisite machinery. In our analysis, the "absolute" time of occurrence of a nerve impulse is determined with reference to stimulus onset; any implementation of spike time coding requires a neural signal that corresponds to stimulus onset. In these experiments, stimulus onset consisted of the abrupt appearance of a stationary pattern; in natural circumstances, stimulus onset is more likely to correspond to an abrupt change in the inputs to a neuron's receptive field due to an eye movement. In either case, stimulus onset corresponds to a burst of excitatory input to visual cortex, which can be viewed as a resetting or synchronization event for its intrinsic circuitry (M.-N. Chee-Orts, K. P. Purpura, and L. M. Optican, unpublished data; Purpura and Schiff 1996. After this synchronization event, propagation of information based on spike times is likely to be based on coincidences

across neurons. Coincidence detection with high temporal precision appears necessary to account for the variability observed in cortical discharges (Softky and Koch 1993), and multiple candidate mechanisms exist for coincidence detection with a wide range of temporal precision (Bourne and Nicoll 1993). Furthermore, it recently has been shown that the spike generation mechanisms of cortical neurons are capable of millisecond precision (Mainen and Sejnowski 1995), so that temporal coding, once generated at this scale, can be propagated.

Spike times, spike intervals, and other metrics

Spike times determine the interspike intervals (and vice versa, with the notion that the first "interspike interval" is the interval between the onset of data collection and the first spike). However, it is not true that the distance between two spike trains in the sense of $D^{interval}$ determines the distance between the trains in the sense of D^{spike}, or vice versa, and there is no built-in reason that clustering in the sense of one metric will determine clustering in the sense of the other. [Indeed, the topologies induced by these metrics are inequivalent, with D^{spike} "stronger" than $D^{interval}$ (Gaal 1964)]. Thus in principle, one could find strong evidence of temporal coding in D^{spike} but not in $D^{interval}$. For example, if the main source of variability was that spikes were deleted at random, then the times of the surviving spikes would still be capable of encoding stimulus-dependent information, but the intervals might be nearly random. Conversely, $D^{interval}[q]$ can distinguish a neuron whose firing pattern consists of intervals with a chaotic nonlinear recursion or a grammar (Rapp et al. 1994) from one whose firing pattern is determined by a renewal process with equal interval statistics; $D^{spike}[q]$ cannot make this distinction.

Hopfield (1995) recently has proposed a specific hypothesis for how sensory information might be represented in, and propagated by, spike trains. This proposal is not limited to the representation of temporal information (Bialek et al. 1991) and features a non-Euclidean distance between spike trains, but one that does not directly correspond to the metrics considered here. In this view, temporal coding is embodied in the pattern of intervals between spikes that occur during successive oscillations and thus would be manifest more strongly in $D^{interval}$ than in D^{spike}. However, the main role of the oscillations is to provide a synchronization; their periodicity is less crucial. If the oscillations were replaced by aperiodic synchronizing pulses related to stimulus onset, one would expect stronger encoding in the sense of D^{spike} (K. Purpura, M.-N. Chee-Orts, and L. M. Optican, unpublished observations).

We stress that we make no claims to have found the optimal metric. We have only explored two families of metrics, chosen because of their simplicity and the biological motivations discussed in the INTRODUCTION. It is easy to generate many other metrics, for example, by adding other kinds of allowed elementary steps (e.g., involving spike clusters (Abeles 1982a,b; Dayhoff and Gerstein 1983a,b), altering the cost function, or permitting the metrics to be time-dependent. Furthermore, the present analysis does not examine the relationship of temporal patterns within a single neuron's discharge and temporal patterns across neurons (Aertsen et al. 1989). Additional investigations will provide additional insights concerning temporal coding, but it is unlikely that they will dispute the relevance of measures of temporal precision or our finding that temporal precision is modality dependent and thus provides a substrate for temporal multiplexing even in the face of the apparently high variability that is common in cortical neurons.

The experimental work described here was conducted in the Laboratory of Sensorimotor Research in the National Eye Institute of the National Institutes of Health in Bethesda, MD. The authors are indebted to Drs. Lance Optican and Robert Wurtz for supporting this component of the project and to Prof. Bruce Knight, Dr. Ken Miller, and two referees for helpful comments. The authors also acknowledge Dr. John McClurkin for his expert advice and assistance on all aspects of the experimental work. We thank M. Conte for assistance with the preparation of the manuscript. A portion of this work was presented at the 1994 meeting of the Society for Neuroscience, Miami (Victor and Purpura 1994).

This work was supported by NIH Grants EY-9314 (J. D. Victor) and NS-01677 (K. P. Purpura), The McDonnell-Pew Foundation (K. P. Purpura), The Revson Foundation (K. P. Purpura), and The Hirschl Trust (J. D. Victor).

Address for reprint requests: J. D. Victor, Dept. of Neurology and Neuroscience, Cornell University Medical College, 1300 York Ave., New York, NY 10021.

Received 10 October 1995; accepted in final form 21 February 1996.

REFERENCES

ABELES, M. *Local Cortical Circuits. An Electrophysiological Study.* Berlin: Springer, 1982a.

ABELES, M. Role of the cortical neuron: integrator or coincidence detector? *Isr. J. Med. Sci.* 18: 83–92, 1982b.

ABELES, M. *Corticonics. Neural Circuits of the Cerebral Cortex.* Cambridge, UK: Cambridge University Press, 1991.

ABELES, M. AND GERSTEIN, G. L. Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol.* 60: 909–924, 1988.

ABRAMSON, N. *Information Theory and Coding.* New York: McGraw-Hill, 1963.

AERTSEN, A. M., GERSTEIN, G. L., HABIB, M. K., AND PALM, G. Dynamics of neuronal firing correlation: modulation of effective connectivity. *J. Neurophysiol.* 61: 900-917, 1989.

BIALEK, W., RIEKE, F., DE RUYTER VAN STEVENINCK, R. R., AND WARLAND, D. Reading a neural code. *Science Wash. DC* 252: 1854–1857 1991.

BLISS, T. V. P. AND COLLINGRIDGE, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature Lond.* 361: 31–39, 1993.

BOURNE, H. R. AND NICOLL, R. Molecular machines integrate coincident synaptic signals. Cell 72/Neuron 10, *Suppl.*: 65–85, 1993.

CARLTON, A .G. On the bias of information estimates. *Psychol. Bull.* 71: 108–109, 1969.

CHEE-ORTS, M. N. AND OPTICAN, L. M. Cluster method for analysis of transmitted information in multivariate neuronal data. *Biol. Cybern.* 69: 29–35, 1993.

CRIST, C. F., YAMASAKI, D. S. G., KOMATSU, H., AND WURTZ, R. H. A grid system and a microsyringe for single cell recording. *J. Neurosci. Methods* 26: 117–122, 1988.

DAYHOFF, J. E. AND GERSTEIN, G. L. Favored patterns in spike trains. I. Detection. *J. Neurophysiol.* 49: 1334–1346, 1983a.

DAYHOFF, J. E. AND GERSTEIN, G. L. Favored patterns in spike trains. II. Application. *J. Neurophysiol.* 49: 1347–1363, 1983b.

EDELMAN, S. AND CUTZU, F. Representation of complex parametric similarity among 3-D shapes. Invest. Ophthalmol. *Vis. Sci.* 36: 1068, 1995.

FAGEN, R. M. Information measures: statistical confidence limits and inference. *J. Theor. Biol.* 73: 61–79, 1978.

FUKUNAGA, K. *Introduction to Statistical Pattern Recognition* (2nd ed.). New York: Academic Press, 1990, p. 591.

GAAL, S. A. *Point Set Topology.* New York: Academic Press, 1964, p. 317.

GATTAS, R., SOUSA, A. P. B., AND GROSS, C. G. Visuotopic organization and extent of V3 and V4 of the macaque. *J. Neurosci.* 8: 1831–1845, 1988.

GAWNE, T. J., MCCLURKIN, J. W., RICHMOND, B. J., AND OPTICAN, L. M.

Lateral geniculate neurons in behaving primates. III. Response predictions of a channel model with multiple spatial-to-temporal filters. *J. Neurophysiol.* 66: 809–823, 1991.

GEISLER, W. S., ALBRECHT, D. G., SALVI, R. J., AND SAUNDERS, S.S. Discrimination performance of single neurons: rate and temporal-pattern information. *J. Neurophysiol.* 66: 334–362, 1991.

HAYS, A. V., RICHMOND, B. J., AND CHU, F. C. A Unix-based multiple process real-time data acquisition and control. *Wescon Conf. Proc.* 2/1: 1–10, 1982.

HELLER, J., HERTZ, J. A., KJAER, T. W., AND RICHMOND, B. J. Information flow and temporal coding in primate pattern vision. *J. Comput. Neurosci.* 2: 175–193, 1995.

HOPFIELD, J. J. Pattern recognition computation using action potential timing for stimulus representation. *Nature Lond.* 376: 33–36, 1995.

JUDGE, S. J., RICHMOND, B. J., AND CHU, F. C. (1980) Implantation of magnetic search coils for measurement of eye position: an improved method. Vision Res. *20*, 535-538.

JULESZ, B., GILBERT, E. N., AND VICTOR, J. D. Visual discrimination of textures with identical third-order statistics. *Biol. Cybern.* 31: 137149, 1978.

KJAER, T. W., HERTZ, J. A., AND RICHMOND, B. J. Decoding cortical neuronal signals: network models, information estimation, and spatial tuning. *J. Comput. Neurosci.* 1: 109–139, 1994.

LARSON, J., WONG, D., AND LYNCH, G. Patterned stimulation at the theta frequency is optimal for the induction of hippocampal long-term potentiation. *Brain Res.* 368: 347–350, 1986.

MAINEN, Z. AND SEJNOWSKI, T. J. Reliability of spike timing in neocortical neurons. *Science Wash. DC* 268: 1503–1506, 1995.

MCCLURKIN, J. W., GAWNE, T. J., OPTICAN, L. M., AND RICHMOND, B. J. Lateral geniculate neurons in behaving primates. II. Encoding information in the temporal shape of the response. *J. Neurophysiol.* 66: 794–808, 1991a.

MCCLURKIN, J. W., GAWNE, T. J., RICHMOND, B. J., OPTICAN, L. M., AND ROBINSON, D. L. Lateral geniculate neurons in behaving primates I. Responses to two-dimensional stimuli. *J. Neurophysiol.* 66: 777–793, 1991b.

MEL, B. W. Synaptic integration in an excitable dendritic tree. *J. Neurophysiol.* 70: 1086–1101, 1993.

MIDDLEBROOKS, J. C., CLOCK, A. E., XU, L., AND GREEN, D. M. A panoramic code for sound location by cortical neurons. *Science Wash. DC* 264: 842–844, 1994.

OPTICAN, L. M., GAWNE, T. J., RICHMOND, B. J., AND JOSEPH, P. J. Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. *Biol. Cybern.* 65: 305–310, 1991.

OPTICAN, L. M. AND RICHMOND, B. J. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J. Neurophysiol.* 57: 162–178, 1987.

PERKEL, D. H., GERSTEIN, G. L., AND MOORE, G. P. Neuronal spike trains and stochastic point processes. I. The single spike train. *Biophys. J.* 7: 391–418, 1967a.

PERKEL, D. H., GERSTEIN, G. L., AND MOORE, G. P. Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.* 7: 419–440, 1967b.

PURPURA, K. P., CHEE-ORTS, M.-N., AND OPTICAN, L. M. Temporal encoding of texture properties in visual cortex of awake monkey. *Soc. Neurosci. Abstr.* 19: 771, 1993.

PURPURA, K. P. AND SCHIFF, N. D. The thalamic intralaminar nuclei: a role in visual awareness. *Neuroscientist.* In press.

PURPURA, K. P., VICTOR, J. D., AND KATZ, E. Striate cortex extracts higher-order spatial correlations from visual textures. *Proc. Natl. Acad. Sci. USA* 91: 8482–8486, 1994.

RAPP, P. E., ZIMMERMAN, I. D., VINING, E. P., COHEN, N., ALBANO, A. M., AND JIMÉNEZ-MONTAÑO, M. A. The algorithmic complexity of neural spike trains increases during focal seizures. *J. Neurosci.* 14: 4731–4739, 1994.

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.

RICHMOND, B. J., WURTZ, R. H., AND SATO, T. Visual responses of inferior temporal neurons in the awake rhesus monkey. *J. Neurophysiol.* 50: 1415–1432, 1983.

ROBINSON, D. A. A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.* 10: 137–145, 1963.

ROSE, G. M. AND DUNWIDDIE, T. V. Induction of hippocampal long-term potentiation using physiologically patterned stimulation. *Neurosci. Lett.* 69: 244–248, 1986.

SELLERS, P. H. On the theory and computation of evolutionary distances. *SIAM J. Appl. Math.* 26: 787–793, 1974.

SHADLEN, M. N. AND NEWSOME, W. T. Noise, neural codes, and cortical organization. *Curr. Opin. Neurobiol.* 4: 569–579, 1994.

SHADLEN, M. N., AND NEWSOME, W. T. Is there a signal in the noise? *Curr. Opin. Neurobiol.* 5: 248–250, 1995.

SKARDA, C. A. AND FREEMAN, W. J. How brains make chaos to make sense of the world. *Behav. Brain Sci.* 10: 161–195, 1987.

SOFTKY, W. R. Simple codes versus efficient codes. *Curr. Opin. Neurobiol.* 5: 239–247, 1995.

SOFTKY, W.R. AND KOCH, C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* 13: 334–350, 1993.

TREVES, A. AND PANZERI, S. The upward bias in measures of information derived from limited data samples. *Neural Comput.* 7: 399–407, 1995.

VICTOR, J. D. Images, statistics, and textures: a comment on Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture. *J. Opt. Soc. Am.* A11: 1680–1684, 1995.

VICTOR, J. D., CONTE, M. M., PURPURA, K. P., AND KATZ, E. Isodipole textures: a window on cortical mechanisms of form processing. In: *Early Vision and Beyond* edited by T. Papathomas, C. Chubb, A. Gorea, and E. Kowler. Cambridge: The MIT Press. 1995, p. 99–107.

VICTOR, J. D. AND PURPURA, K. P. A new approach to the analysis of spike discharges: cost-based metrics. *Soc. Neurosci. Abstr.* 20: 315, 1994.

VICTOR, J. D., PURPURA, K. P., KATZ, E., AND MAO, B. Population encoding of spatial frequency, orientation, and color in macaque V1. *J. Neurophysiol.* 72: 2151–2166, 1994.

VOLGUSHEV, M., VIDYASAGAR, T. R., AND PEI, X. Dynamics of the orientation tuning of postsynaptic potentials in the cat visual cortex. *Vis. Neurosci.* 12: 621–628, 1995.

WURTZ, R. H. Visual receptive fields of striate cortex neurons in awake monkeys. *J. Neurophysiol.* 32: 727–742, 1969.