

Information Tuning of Populations of Neurons in Primary Visual Cortex

Kukjin Kang,^{1,2} Robert M. Shapley,¹ and Haim Sompolinsky³

¹Center for Neural Science and ²Courant Institute of Mathematical Sciences, New York University, New York, New York 10003, and ³Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel

Neurons in macaque primary visual cortex (V1) show a diversity of orientation tuning properties, exhibiting a broad distribution of tuning width, baseline activity, peak response, and circular variance (CV). Here, we studied how the different tuning features affect the performance of these cells in discriminating between stimuli with different orientations. Previous studies of the orientation discrimination power of neurons in V1 focused on resolving two nearby orientations close to the psychophysical threshold of orientation discrimination. Here, we developed a theoretical framework, the information tuning curve, that measures the discrimination power of cells as a function of the orientation difference, $\delta\theta$, of the two stimuli. This tuning curve also represents the mutual information between the neuronal responses and the stimulus orientation. We studied theoretically the dependence of the information tuning curve on the orientation tuning width, baseline, and peak responses. Of main interest is the finding that narrow orientation tuning is not necessarily optimal for all angular discrimination tasks. Instead, the optimal tuning width depends linearly on $\delta\theta$. We applied our theory to study the discrimination performance of a population of 490 neurons in macaque V1. We found that a significant fraction of the neuronal population exhibits favorable tuning properties for large $\delta\theta$. We also studied how the discrimination capability of neurons is distributed and compared several other measures of the orientation tuning such as CV with Chernoff distances for normalized tuning curves.

Key words: primary visual cortex; orientation selectivity; population coding; macaque monkey; Chernoff distance; discrimination

Introduction

Neurons in primary visual cortex (V1) are selective for the movement direction or the orientation of line-like simple visual patterns. The shape of the response tuning curve and orientation selectivity of neurons in macaque V1 are diverse (Ringach et al., 2002). Our motivation was to understand the possible functional use of the observed diversity in V1 orientation tuning.

The orientation selectivity of neurons in V1 has been studied mainly in two different ways. First, the most informative point of a tuning curve, which is usually the steep flank part of the tuning curve, is selected, and discrimination capability of the neuron for two angles is computed using ROC analysis or neurometric functions (Bradley et al., 1987; Hawken and Parker, 1990; Vogels and Orban, 1990; Parker and Newsome, 1998). But these studies only analyzed discrimination for two nearby angles and did not clarify the functional use of broadly tuned neurons. In addition, discrimination capability computed in this way depends only on the local shape of the tuning curve. The advantage of having diversity in the global shape of tuning curves may be clear only in terms of population coding.

Discrimination capability of a population of neurons is more

difficult to study mainly because a practical measure for it has been lacking. In several studies (Seung and Sompolinsky, 1993; Abbott and Dayan, 1999; Sompolinsky et al., 2001), Fisher information was used to study population coding. But Fisher information can be used only when angles are very near to each other. Other well known measures such as mutual information are computationally too expensive to calculate for a population of neurons.

Here, we studied the relationship between the shape of a tuning curve and the discrimination capability of a population of neurons using the Chernoff distance (Cover and Thomas, 1991; Kang and Sompolinsky, 2001). The Chernoff distance is a measure of the difference between two probability distributions and has direct relationships with other information measures such as Fisher information, mutual information, and the error of maximum likelihood discrimination.

For a population of neurons with preferred orientations that are distributed isotropically, the Chernoff distance between two distributions of spike counts corresponding to two different orientations depends on only $\delta\theta$, the difference in the orientations. The information tuning curve is a plot of Chernoff distance as a function of $\delta\theta$. The shape of the information tuning curve characterizes how different orientations are represented by the activities of a population of neurons. In Results, we studied how the information tuning curve depends on various features of the response tuning curve.

We applied the theoretical analysis to macaque V1 data. The results suggest that diversity may exist in V1 because different

Received Sept. 18, 2003; revised March 1, 2004; accepted March 1, 2004.

This work was supported by grants from the National Eye Institute to R. M. S., M. J. Hawken, and D. L. Ringach. We thank the Sloan and Swartz Foundations and the US-Israel Bi-National Science Foundation for support.

Correspondence should be addressed to Dr. Kukjin Kang, Center for Neural Science, New York University, 4 Washington Place, Room 809, New York, NY 10003. E-mail: kkj@cns.nyu.edu.

DOI:10.1523/JNEUROSCI.4272-03.2004

Copyright © 2004 Society for Neuroscience 0270-6474/04/243726-10\$15.00/0

neurons are optimal for different discrimination tasks. It also shows that neurons in macaque V1 are not optimized for the discrimination of nearby angles. Finally, we discussed the relationship between Chernoff distance and several other measures of orientation tuning such as circular variance (CV) and tuning width.

Materials and Methods

Preparation and recording. Acute experiments were performed on 40 adult Old World monkeys (*Macaca fascicularis*) in the laboratories of R. M. Shapley, M. J. Hawken, and D. L. Ringach and colleagues (cf. Ringach et al. 2002). The methods of preparation and single-cell recording are the same as those described by Ringach et al. (2002). Each cell was stimulated monocularly via the dominant eye and characterized by measuring its steady-state response to drifting sinusoidal gratings (the non-dominant eye was occluded). With this method, basic attributes of the cell, including spatial and temporal frequency tuning, orientation tuning, contrast response function, and color sensitivity, as well as area, length, and width tuning curves, were measured. Orientation tuning curves were measured at high contrast (0.8). Spike times were recorded for 18 directions (every 20°). Spatial frequency, temporal frequency, and size of the sinusoidal gratings were optimized for each cell separately to maximize the peak response.

A model for the directional tuning of the spike count. We introduced a Gaussian model for the directional tuning of mean spike count and fit the model to the measured mean spike counts for 18 directions to reduce noise in the experimental data and to extract a small number of parameters to describe the shape of the tuning curve. The model tuning curve $\lambda(\theta)$ is described in Equation 1:

$$\lambda(\theta) = A + B_1 \exp(-\theta'^2/2\sigma^2) + B_2 \exp(-\theta''^2/2\sigma^2), \quad (1)$$

where $\theta' = R(\theta, \theta^0)$ and $\theta'' = R(\theta, \theta^0 + \pi)$. θ^0 is the preferred direction of the neuron. $R(x, y) = \min\{|x - y|, 2\pi - |x - y|\}$ is the angle between x and y . See Figure 2 for examples of tuning curves. For each neuron in the V1 data, we minimized the squared error, $Er(A, B_1, B_2, \sigma, \theta^0)$:

$$Er(A, B_1, B_2, \sigma, \theta^0) = \sum_{i=1}^{18} (\lambda(\theta_i, A, B_1, B_2, \sigma, \theta^0) - m(\theta_i))^2, \quad (2)$$

where $m(\theta_i)$ is the mean spike counts of the neuron for the direction, θ_i . We also defined the error ratio, R_{ER} to measure the goodness of the fit to a Gaussian model:

$$R_{ER} = Er(A^*, B_1^*, B_2^*, \sigma^*, \theta^{0*})/Er_0, \quad (3)$$

where $Er_0 = \sum_{i=1}^{18} [m_0 - m(\theta_i)]^2$ and m_0 is the mean of $m(\theta_i)$. $A^*, B_1^*, B_2^*, \sigma^*$ and θ^{0*} are the values of parameters minimizing the error $Er(A, B_1, B_2, \sigma, \theta^0)$.

In this study, we ruled out neurons with a maximum firing rate lower than five spikes per second. Seventy-six neurons among 897 neurons were discarded in this way. We fitted the observed mean spike counts to our Gaussian model (see Eq. 1) and did not study further those neurons that did not show a good fit to the proposed model ($R_{ER} > 0.3$). Three hundred thirty-one neurons among 821 neurons are discarded in this way. The total number of neurons in the resulting database was 490. Most of the discarded neurons should be considered as “noninformative” in any sense. For most of the discarded neurons, the tuning curves were very irregular, and baseline firing rates were relatively large. Spiking activities of those neurons were less reliable so that the statistics of the spike count had larger variance. For a few neurons (<1%), our model was bad because the distance between the peaks of the tuning curve was different from π . But such neurons were rare and ignored in this study.

Classification of neurons. Neurons are classified into orientation-selective (OS) neurons and direction-selective (DS) neurons based on the

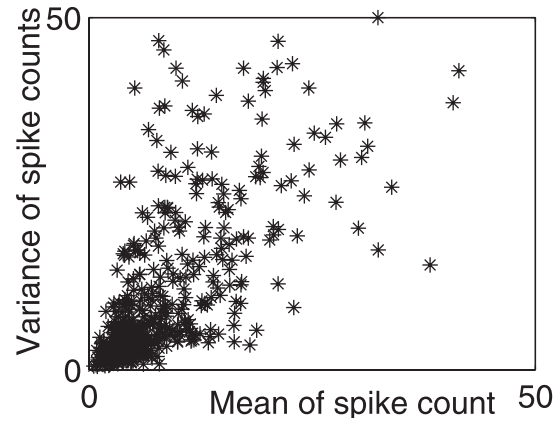


Figure 1. Mean and variance of spike counts of 490 neurons for preferred directions of each neuron. For each neuron, the number of spikes for one period of sinusoidal grating stimulus was counted. The average value of the ratio of variance and the mean is 1.9, but the distribution of the ratio between mean and variance has a peak at 1, which is the value for Poisson distributions.

ratio of the heights of two peaks of tuning curves R_B . R_B is $\min(B_1, B_2)/\max(B_1, B_2)$ where B_1 and B_2 are the height of two peaks (see Eq. 1 and Fig. 4). R_B is a ratio of the responses for the preferred direction and the opposite direction. For tuning curves of ideal OS neurons, R_B is 1, and for ideal DS tuning curves, R_B is 0. We classified neurons as OS if $R_B > 0.5$ or as DS otherwise. We found that 240 neurons are OS and 250 neurons are DS among 490 neurons. A similar method was used in a previous study (Hawken et al., 1988).

Spike count statistics. As for the statistics of the spike count, we assumed that it follows a Poisson distribution, the mean of which is the same as the variance. It is observed in experiments that the variance is often approximately proportional to mean spike count (Tolhurst et al., 1983). Real distributions show some deviations from Poisson distributions. Figure 1 shows a scatter plot of the mean and the variance of spike count at the preferred orientation for 490 neurons. Here, we just assumed Poisson distributions and focused on studying the role of the shape of tuning curves in the neuronal representation of sensory information.

Significance of correlation. We calculated correlation coefficients between several features of tuning curves. To show the significance, we randomly shuffled the indices of one of two quantities with which the correlation coefficient is calculated and calculated the correlation coefficient again. We used the frequency that the absolute value of this correlation coefficient after random shuffling is larger than the absolute value of the correlation coefficient before random shuffling as a measure of the significance. We did this 1000 times. If none of the trials generated a correlation coefficient larger than the original, we took the significance as <0.1%.

Results

Distance measures in the representation space of a population of neurons

To study the relationship between the shape of a tuning curve and the capability to discriminate angles, a measure of discrimination capability should be defined and calculated. Here, we used Chernoff distance as a measure of orientation discrimination capability for a population of neurons.

Chernoff distance measures the difference between two distributions. For two distributions, $P(\vec{r}|\theta_1)$ and $P(\vec{r}|\theta_2)$, Chernoff distance $D_C(\theta_1, \theta_2)$ is defined in the following way:

$$D_\alpha(\theta_1, \theta_2) = -\log \text{Tr}_{\vec{r}} P^\alpha(\vec{r}|\theta_1) P^{1-\alpha}(\vec{r}|\theta_2) \quad (4)$$

$$D_C(\theta_1, \theta_2) = \max_{\alpha} D_\alpha(\theta_1, \theta_2) \quad (5)$$

$D_C(\theta_1, \theta_2)$ is the maximum value of $D_\alpha(\theta_1, \theta_2)$ in terms of α within an interval $0 \leq \alpha \leq 1$. θ_i is the orientation of a sinusoidal grating, and \vec{r} is a vector of spike counts for a population of neurons. $P(\vec{r}|\theta_i)$ is the distribution of activity across the population \vec{r} when the stimulus with the orientation θ_i is presented. $Tr_{\vec{r}}$ is a summation over all possible \vec{r} .

$D_C(\theta_1, \theta_2) \geq 0$ for any pairs of distributions. $D_C(\theta_1, \theta_2)$ is 0 if and only if two distributions are the same. $D_C(\theta_1, \theta_2) = D_C(\theta_2, \theta_1)$ so that $D_C(\theta_1, \theta_2)$ is uniquely defined for a given pair of orientations (Cover and Thomas, 1991; Kang and Sompolinsky, 2001).

Before deriving the form of Chernoff distance for a population of neurons, we introduced its relationship with Euclidean distance and the error of maximum-likelihood discriminator to explain the meaning of the Chernoff distance. For the relationships with Fisher information and mutual information, see Appendix. We discuss the advantage of Chernoff distance later (see Discussion).

The relationship between Chernoff distance and Euclidean distance

A simple way to measure the difference between two distributions is to calculate Hellinger distance (Cam and Yang, 2000), which is the Euclidean distance between $\sqrt{P(\vec{r}|\theta_i)}$:

$$D_H(\theta_1, \theta_2) = \sqrt{Tr_{\vec{r}} \left(\sqrt{P(\vec{r}|\theta_1)} - \sqrt{P(\vec{r}|\theta_2)} \right)^2} \quad (6)$$

Chernoff distance D_C is the maximum value of D_α in terms of α , and D_α often has its maximum at $\alpha = 0.5$. In this case, D_C has the following relationship with Hellinger distance:

$$D_C(\theta_1, \theta_2) = -\log(2 - D_H^2) + \log 2. \quad (7)$$

Hellinger distance is a more intuitive measure than Chernoff distance and often gives a very good approximation of Chernoff distance through Equation 6 if D_α has its maximum near $\alpha = 0.5$. In fact, for the population of neurons with orientation symmetry as considered later here, D_α has a maximum at $\alpha = 0.5$ (see Appendix). So Chernoff distance and Hellinger distance have the above relationship here.

Relationship with the error of maximum-likelihood discriminator

Another way of measuring the difference between two distributions is to perform discrimination using a discriminator and calculate the error. If two distributions are well separated, the discrimination error is small. The error of maximum-likelihood discriminator provides an error of the optimal discriminator.

When $D_C(\theta_1, \theta_2) \gg 1$, the error of the maximum-likelihood discriminator P_C has an exponential dependence on the Chernoff distance $D_C(\theta_1, \theta_2)$ (Kang and Sompolinsky, 2001):

$$P_C \propto \exp[-D_C(\theta_1, \theta_2)] \quad (8)$$

The error of the maximum-likelihood discriminator P_C is defined in the following way:

$$P_C = Tr_{\vec{r}} P(\vec{r}|\theta_1) \Theta(\log[P(\vec{r}|\theta_2)/P(\vec{r}|\theta_1)]), \quad (9)$$

where $\Theta(x)$ is 1 for $x > 0$ and 0 for $x \leq 0$. Equation 8 shows that if the Chernoff distance $D_C(\theta_1, \theta_2)$ is larger than 1, discrimination between two stimuli can be done with small error.

Whether the condition of $D_C(\theta_1, \theta_2) \gg 1$ is satisfied or not depends on the size of the population, the size of the time interval, and the shape of tuning curves in general. For two far-away ori-

entations, this condition will be satisfied in most cases. For two orientations very close to each other, this condition may not be satisfied for a population of neurons with small size. For example, for a population of ~ 100 typical neurons in V1, time interval ~ 100 msec and angles larger than a few degrees, the Chernoff distance for this population is typically of the order of 1 if not much larger than that. So the condition required for the relationships between Chernoff distance and other information measures are satisfied in physiologically plausible situations. Chernoff distance has exponential relationships with mutual information and the error of maximum-likelihood discriminator. So, in practice, it is enough for Chernoff distance to be 3–4 to show good convergence to its asymptotic behavior.

Chernoff distance for a population of neurons

Here, we calculated the Chernoff distance for a population of neurons to get a quantitative relationship between the shape of the tuning curve and the discrimination capability of a population of neurons. We assumed that each neuron observed in the experiment represents a population of neurons with tuning curves that have the same shape as the observed one but in which preferred directions are different. We calculated $D_C(\theta_1, \theta_2)$ for this population of neurons.

When the tuning curve of a neuron is $\lambda(\theta)$, we generate tuning curves for a population of neurons using the operation of rotation and reflection:

$$\lambda^{k,a}(\theta) = \lambda[(\theta - \theta^k)(-1)^a], \quad (10)$$

where $\theta^k = 360^\circ k/N$, $k = 0 \dots N-1$ and $a = 1$ or 2. k is an index for rotation of the tuning curve, and a is an index for its reflection. The number of neurons in this population is $2N$.

For this population of neurons, the Chernoff distance in Equation 4 has the following form (see Appendix for the derivation):

$$D_C(\theta_1, \theta_2) = \sum_{a=1,2,k=0,\dots,N-1} \frac{1}{2} \left(\sqrt{\lambda^{k,a}(\theta_1)} - \sqrt{\lambda^{k,a}(\theta_2)} \right)^2. \quad (11)$$

The summation in Equation 11 can be approximated by an integration for large N :

$$D_C(\theta_1, \theta_2) \approx \frac{N}{720^\circ} \int d\theta \left(\sqrt{\lambda(\theta + \delta\theta)} - \sqrt{\lambda(\theta)} \right)^2 + \left(\sqrt{\lambda(\theta - \delta\theta)} - \sqrt{\lambda(\theta)} \right)^2. \quad (12)$$

Equation 12 shows us how the shape of the tuning curve is related to the discrimination capability of a population of neurons. The Chernoff distance is an extensive quantity so that it is proportional to the size of the neuronal population. Here, this N will be assumed to be divided out so that the Chernoff distance will be Chernoff distance per neuron in the population. $D_C(\theta_1, \theta_2)$ will be also written as $D_C(\delta\theta)$ because $D_C(\theta_1, \theta_2)$ depends on θ_1 and θ_2 only through $\delta\theta$.

The information tuning curve

We introduced the information tuning curve, a plot of $D_C(\delta\theta)$ as a function of $\delta\theta$, and discussed what it shows. Then we studied how $D_C(\delta\theta)$ depends on the features of a tuning curve.

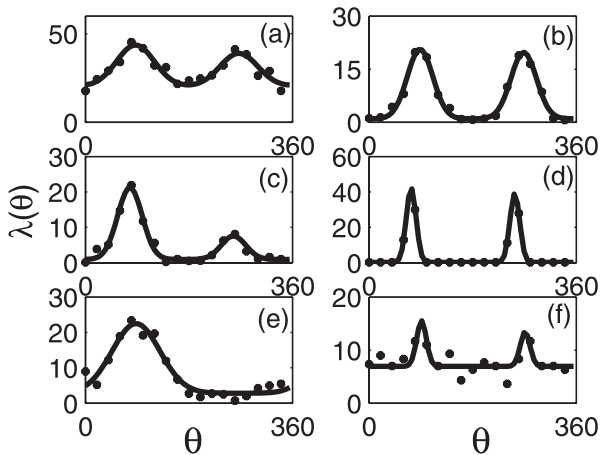


Figure 2. The response tuning curves of six neurons in V1. Solid lines are models of tuning curves fitted to experimental results. Filled dots present the observed mean spike counts for 1 sec.

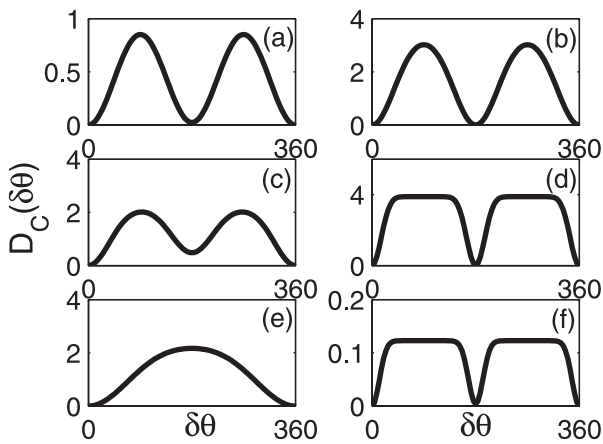


Figure 3. Plots of $D_C(\delta\theta)$ (i.e., examples of information tuning curves). Response tuning curves of corresponding neurons are shown in Figure 2.

Examples of information tuning curves

Figure 2 shows response tuning curves with various shapes. There are broad tuning curves and narrow tuning curves. There are neurons with large baselines and neurons with baselines at zero. Neuron (a) has a bigger peak response than neuron (b). In the previous section, we found the relationship between the shape of a tuning curve and discrimination capability (Eq. 12). Figure 3 shows $D_C(\delta\theta)$ as a function of $\delta\theta$ and how the diversity in the shapes of response tuning curves affects the discrimination capabilities of neurons.

The information tuning curve shows how the distance between two orientations in the neuronal representation space changes as the angle between them, $\delta\theta$, increases. Consider information tuning curves (a) and (e) in Figure 3. One thing very easy to notice in the shape of those information tuning curves is that OS tuning curves like (a) in Figure 2 have information tuning curves with two peaks whereas a DS tuning curve like (e) in Figure 2 has an information tuning curve with one peak. Figure 2a is an OS tuning curve so that it is not able to discriminate two opposite directions. It is represented by a minimum of $D_C(\delta\theta)$ at $\delta\theta = 180^\circ$ in Figure 3. Discrimination capability of an ideal DS neuron such as (e) in Figure 2 should be maximized for two opposite directions. It is represented by a maximum of the information

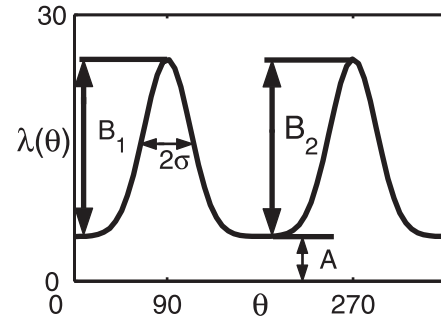


Figure 4. A model of tuning curve $\lambda(\theta)$. $\lambda(\theta) = A + B_1 \exp(-\theta^2/2\sigma^2) + B_2 \exp(-\theta'^2/2\sigma^2)$. θ' (θ'') is the angle between θ and 90° (270°). For this example, $A = 5$, $B_1 = B_2 = 20$, $\sigma = 22.5^\circ$.

tuning curve at $\delta\theta = 180^\circ$ in Figure 3e. There are also information tuning curves between these two cases like Figure 3c.

The information tuning curve also enables us to compare the discrimination capability of neurons quantitatively. Consider (a) and (b) in Figure 3. The information tuning curves have similar shapes, but the overall scale is more than three times bigger for (b), which means that we need three times as many neurons like (a) as neurons like (b) to achieve the same discrimination power. This is because of the big baseline of the response tuning curve of (a) (Fig. 2). Because the spike count is Poisson, a large baseline means spike counts are more stochastic. In fact, the modulation of the tuning curve for (a) is bigger than for (b), suggesting that without a large baseline, neuron (a) should be the more informative neuron. Neurons (d) and (f) also have information tuning curves with similar shape and different overall scales. Neuron (d) has a discrimination capability about 40 times bigger than neuron (f).

Information tuning also shows which tuning curves are good for the discrimination of nearby angles or faraway angles. Consider the information tuning curves of (d) and (e) in Figure 3. For (d), $D_C(\delta\theta)$ increases with a large slope as $\delta\theta$ increases from 0. For (e), the information tuning curve has a much smaller slope. For narrow response tuning curves like Fig. 2d, information tuning curves increase with large slopes as $\delta\theta$ increases from 0 and saturate soon. For broad response tuning curves like Figure 2e, information tuning curves increase with small slopes and do not saturate. This makes neurons with narrow response tuning curves have a discrimination capability larger for small $\delta\theta$ and smaller for large $\delta\theta$ than neurons with broad tuning curves.

Parameters to determine the Chernoff distance

Here, we study which features of tuning curves determine $D_C(\delta\theta)$. Consider the model of tuning curve $\lambda(\theta)$ shown in Figure 4. $\lambda(\theta)$ has two peaks at opposite directions. A is the level of baseline. B_1 and B_2 are the size of Gaussian peaks on the top of the baseline. σ is the width of the Gaussian functions. For simplicity, we considered only the case of OS tuning curves ($B_1 = B_2 = B$) here.

Equation 12 shows that $D_C(\delta\theta, A, B, \sigma)$ is $M_B D_C(\delta\theta, A/M_B, B/M_B, \sigma)$, where $M_B = A + B$ is the peak response of the tuning curve. Note that $D_C(\delta\theta, A/M_B, B/M_B, \sigma)$ is the Chernoff distance for a normalized tuning curve, the peak response of which is 1 because $A/M_B + B/M_B = 1$. We found that it is convenient to factor out the peak response M_B and study how Chernoff distance depends on the remaining parameters because once we understand how $D_C(\delta\theta, A/M_B, B/M_B, \sigma)$ behaves, it is easy to see how the original Chernoff distance depends on the peak response M_B : it is proportional to M_B . For this reason, we factored

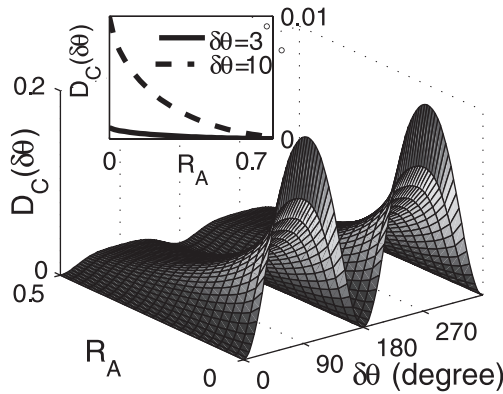


Figure 5. A surface plot of $D_C(\delta\theta)$ as a function of R_A and $\delta\theta$. $\sigma = 17.2^\circ$. The inset has plots of $D_C(3^\circ)$ and $D_C(10^\circ)$ as a function of R_A . The solid line is for $D_C(3^\circ)$. The dashed line is for $D_C(10^\circ)$.

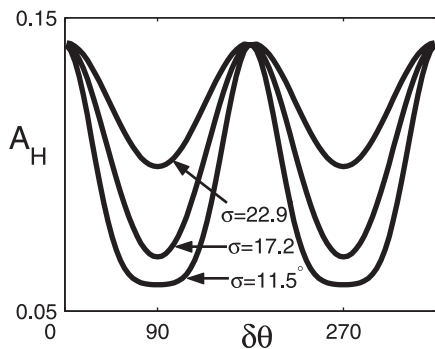


Figure 6. Plot of the half-width for relative baseline A_H as a function of angular difference $\delta\theta$. The dependence on tuning width is revealed by comparing these curves for tuning width $\sigma = 11.5, 17.2$, and 22.9° .

out peak response M_B and studied how $D_C(\delta\theta)$ depends on three parameters: $\delta\theta$, relative baseline $R_A = A/M_B$, and tuning width σ . Relative baseline $R_A = A/M_B$ is also the ratio of the responses to preferred orientation and orthogonal orientation (Gegenfurtner et al., 1996) and has been considered as a measure of orientation selectivity.

Dependence on the relative baseline R_A

$D_C(\delta\theta)$ decreases monotonically as R_A increases because R_A is a nontuned component of the tuning curve. Figure 5 shows a two-dimensional plot $D_C(\delta\theta, R_A, \sigma)$ for $\sigma = 17.2^\circ$, a typical value of σ for neurons in V1. For R_A as large as 0.5, $D_C(\delta\theta)$ is already very small for all $\delta\theta$. Also note that $D_C(\delta\theta)$ for large R_A looks flatter than $D_C(\delta\theta)$ for small R_A . This point is more clearly shown by the inset in Figure 5. $D_C(\delta\theta)$ decreases monotonically as R_A increases and decreases faster for larger $\delta\theta$.

To study these points more quantitatively, we use the “half-width” value of R_A, A_H , where $D_C(\delta\theta, R_A = A_H, \sigma)$ is $D_C(\delta\theta, R_A = 0, \sigma)/2$. A_H measures how fast $D_C(\delta\theta)$ decreases as R_A increases. If A_H were small, it would mean $D_C(\delta\theta)$ decays very fast as R_A increases. If R_A of a tuning curve were much larger than A_H , the discrimination power of the tuning curve would be small, unless peak response M_B were very big.

Figure 6 shows a plot of A_H for several different values of σ and for all possible values of $\delta\theta$. Note that $D_C(\delta\theta)$ is very sensitive to R_A . For R_A as large as 0.15, $D_C(\delta\theta)$ is already significantly smaller than $D_C(\delta\theta)$ for $R_A = 0$ because a typical value of A_H is 0.1. A_H is smaller than 0.142 and larger than 0.059 for any σ and

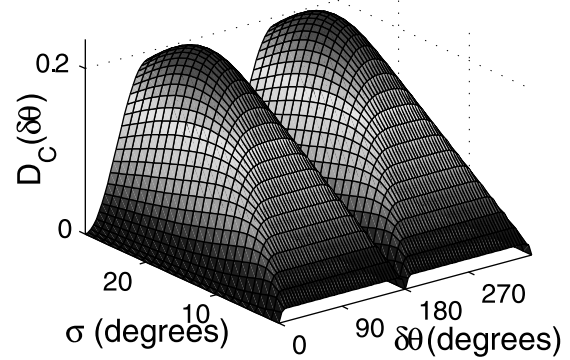


Figure 7. A plot $D_C(\delta\theta)$ as a function of tuning width σ and $\delta\theta$. Relative baseline $R_A = 0$.

$\delta\theta$. These values of A_H provide a scale for R_A to be “too big” or “small enough.” For example, the response tuning curve of Figure 2a has too large a relative baseline because R_A is about 0.5.

The effect of R_A is not the same for different $\delta\theta$. A_H is 0.142 for any tuning width σ . Figure 6 shows that when $\delta\theta$ is close to 90° or 270° , A_H tends to be smaller, which means that degradation of the discrimination capability is bigger for such $\delta\theta$. This is the reason why $D_C(\delta\theta)$ for large R_A is flatter than for smaller R_A as a function of $\delta\theta$ in Figure 5.

Dependence on tuning width σ

Consider the case that the relative baseline $R_A = 0$. In this case, it is possible to calculate $D_C(\delta\theta)$ analytically. Performing the integration in Equation 14 gives the following result:

$$D_C(\delta\theta) = 4N\sigma(1 - e^{-R(\delta\theta,0)^2/8\sigma^2} - e^{-R(\delta\theta,\pi)^2/8\sigma^2}), \tag{13}$$

where $R(x, y)$ is the angle between x and y .

As a function of σ , $D_C(\delta\theta)$ has a maximum at a nonzero value of σ . Figure 7 shows a surface plot of $D_C(\delta\theta)$ as a function of σ and $\delta\theta$. For small σ , exponential terms in Equation 13 are very small for nonzero $\delta\theta$, and $D_C(\delta\theta)$ rapidly saturates to $4N\sigma$ as $\delta\theta$ increases [for another example see (d) in Fig. 3]. This makes $D_C(\delta\theta)$ flat as a function of $\delta\theta$. Figure 7 also shows that a very narrow tuning curve does not produce large $D_C(\delta\theta)$ because $D_C(\delta\theta)$ converges to a value proportional to σ as $\sigma \rightarrow 0$. For larger σ , $D_C(\delta\theta)$ has round shape. In this case, the $4N\sigma$ factor is larger but the exponential terms in Equation 13 decrease $D_C(\delta\theta)$.

For each $\delta\theta$, there is an optimal tuning width maximizing $D_C(\delta\theta)$. Maximization of Equation 13 gives this optimal width, σ^* , which is proportional to $\delta\theta$:

$$\sigma^* = \delta\theta / \sqrt{8x^*} \approx 0.316\delta\theta, \tag{14}$$

where x^* satisfies $1 - e^{-x^*} - 2x^*e^{-x^*} = 0$. We assumed that $\delta\theta \ll \pi$.

An optimal value is more important when $D_C(\delta\theta)$ decreases rapidly as the difference between tuning width σ and optimal tuning width σ^* increases. We defined σ_H to measure how fast $D_C(\delta\theta)$ decreases as σ departs from σ^* . It is defined in a way similar to A_H , such that $D_C(\delta\theta, R_A, \sigma_H) = D_C(\delta\theta, R_A, \sigma^*)/2$. There are two σ_H for a given $\delta\theta$ and R_A . Because the optimal tuning width σ^* is non-zero, $D_C(\delta\theta)$ would be decreased if σ deviated from σ^* either by increasing it or decreasing it.

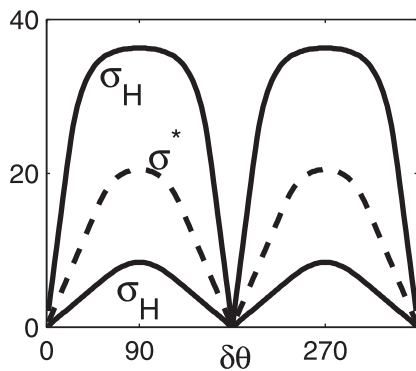


Figure 8. Half-width for tuning width σ , σ_H , and σ^* . Solid lines are for σ_H . The dashed line is for optimal tuning width σ^* . Relative baseline $R_A = 0$.

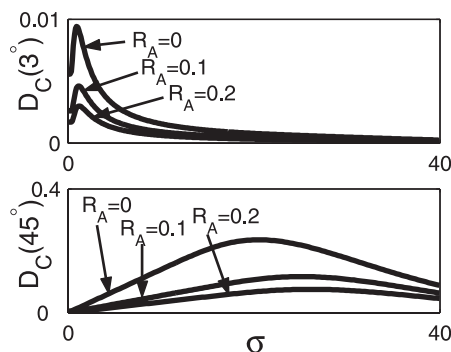


Figure 9. Plots of $D_C(3^\circ)$ and $D_C(45^\circ)$ as functions of tuning width σ . Each line is for a different value of relative baseline R_A . From top to bottom, $R_A = 0, 0.1$, and 0.2 , respectively.

$D_C(\delta\theta)$ depends on σ more sensitively for smaller $\delta\theta$. Analytical study of Equation 13 shows that $|\sigma^* - \sigma_H|$ is $O(\delta\theta)$ for small $\delta\theta$, that is, smaller for smaller $\delta\theta$. Figure 8 shows σ_H and σ^* together for $A = 0$. It is clear that σ_H is closer to σ^* for small $\delta\theta$. It means that neurons with broad tuning curves have poor capability to discriminate two nearby angles because tuning width is very different from the optimal value for nearby angles and $D_C(\delta\theta)$ depends on σ very sensitively. In contrast, for a large $\delta\theta$, $D_C(\delta\theta)$ depends on σ more weakly so that informative neurons do not need to have σ very close to σ^* .

We now consider the more general case in which the relative baseline $R_A > 0$. This cannot be calculated analytically but can be calculated numerically. Figure 9 shows $D_C(3^\circ)$ and $D_C(45^\circ)$ for various values of R_A . These should be compared with the $R_A = 0$ case in Figure 7.

$D_C(\delta\theta)$ has non-zero and finite optimal widths, σ^* for non-zero R_A , too. A smaller value of σ decreases the number of neurons active for the stimuli making $D_C(\delta\theta)$ smaller. But it also increases the slope of the tuning curve making $D_C(\delta\theta)$ bigger. This competition of two effects results in the existence of an optimal σ to discriminate two orientations in general.

Optimal tuning width σ^* is bigger for larger R_A . There is a small shift of σ^* peaks as R_A goes from 0 to 0.3. Figure 10 shows σ^* for various values of R_A . This graph also shows the $\delta\theta$ dependence of the optimal tuning width σ^* .

Neurons in V1

Here, we studied how the features of tuning curves are distributed within the population of neurons in V1 of macaque monkeys.

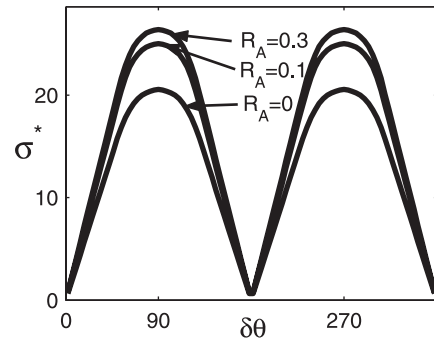


Figure 10. Optimal tuning width σ^* for several different values of relative baseline R_A .

After that, we discussed the distribution of discrimination capability of neurons in V1.

Features of V1 tuning curves

We studied how the features of tuning curves are distributed in V1 separately for OS and DS neurons. Figure 11 shows histograms of peak responses, relative baselines, and response tuning widths for 240 OS neurons and 250 DS neurons.

Peak responses, $M_B = \max\{B_1, B_2\} + A$, to 80% contrast stimuli are < 100 spikes/sec for most of the neurons in V1. The means of M_B are 38.6 and 49.7 spikes/sec for OS neurons and DS neurons, respectively. Only 37 neurons among 490 neurons have peak response higher than 100 spikes/sec. Thirty-three neurons among them are DS neurons.

Figure 11 shows the histograms of relative baseline $R_A = A/M_B$, too. Remember that typical values of A_H are between 0.059 and 0.142 (Fig. 4). A_H gives us a scale to see whether there is a significant degradation of discrimination capability attributable to the baseline. Figure 11 shows that R_A is smaller than these values for most neurons. It means for most neurons R_A of the tuning curve is not too large to degrade discrimination capability. The mean of R_A is bigger for the OS population than the DS population.

Finally, the histograms of tuning width σ show that the distributions of σ are broad or nearly flat within intervals of allowed values. σ cannot be much larger than 40° for OS neurons because two Gaussians overlap if the peaks of Gaussians are too broad. Only DS neurons with one peak can have σ as large as 60° . This gives an upper bound condition on the value of σ . There is also a lower bound for σ . Because our experiment was done only for 18 directions (every 20°), this resolution limitation requires that σ should be larger than 7° . If neurons with tuning width smaller than this existed, our estimation of tuning width would be an overestimation.

It should be emphasized that the tuning widths of neurons in V1 are not optimized for the discrimination of nearby angles. Equation 14 and Figure 10 show that the optimal tuning width σ^* is about $0.3\delta\theta$. This means for $\delta\theta$ as large as 10° , σ^* is only 3 or 4° . Such a small tuning width is hard to find, if not impossible, in macaque V1. As orientation discriminators, neurons in V1 are optimized for $\delta\theta$ larger than 20° .

Table 1 shows the means and median values of distributions shown in Figure 11. The features of tuning curves are not independent of each other. There seems to be several different types of tuning curves in the neuronal population such as narrow OS tuning curves with zero baselines or broad OS tuning curves with large relative baselines and large peak responses. This gives correlations between different features of tuning curves within the

neuronal population. We calculated correlation coefficients between different features for the OS and the DS populations. Table 2 shows these correlation coefficients.

There are three significant correlations. For the OS population, the relative baseline R_A showed a significant correlation with the peak response M_B (correlation coefficient, 0.24). So there is a tendency that OS neurons with a large peak response have a large baseline. The mean of peak responses for OS neurons with relative baseline $R_A > 0.2$ (80 cells) is 47.8 spikes, whereas that for OS neurons with relative baseline $R_A < 0.2$ (160 cells) is 34.0 spikes.

Another significant correlation is found between R_A and σ in the OS population (correlation coefficient, 0.39). Many narrow tuning curves of OS neurons do not have a baseline. The mean of tuning width σ for OS neurons with relative baseline $R_A > 0.2$ (80 cells) is 26.4° , whereas that for OS neurons with relative baseline $R_A < 0.2$ (160 cells) is 20.0° .

The last significant correlation is between M_B and σ in the DS population. DS neurons with large peak responses tend to have broad tuning curves. For example, the average tuning width of 33 DS neurons with $M_B > 100$ spikes/sec is 34.2° . The mean of σ for the other 217 neurons is 23.4° .

Specialization of neurons to different tasks

Here, we studied specialization in V1. We showed in Equation 14 and Figure 9 that optimal tuning width is different for different $\delta\theta$ so that neurons with different tuning width may be specialized to discrimination between angles with different ranges of $\delta\theta$.

One way to study the specialization of neurons to different tasks is to compare the discrimination capability of neurons for two different angles. We made scatter plots of $D_C(\delta\theta)$ for two different values of $\delta\theta$ for normalized tuning curves of 490 neurons. Figure 12 shows three scatter plots of $D_C(\delta\theta_1)$ and $D_C(\delta\theta_2)$ for $\delta\theta_1 = 3^\circ$ and $\delta\theta_2 = 10, 45$ and 90° , respectively. We can see that as $\delta\theta_2$ increases, the spread in the scatter plots increases.

We calculated the correlation coefficient between $D_C(3^\circ)$ and $D_C(\delta\theta)$ as a function of $\delta\theta$. Figure 12d shows this correlation coefficient decreases almost linearly and becomes negative at $\delta\theta = 156^\circ$. This shows that neurons with large tuning widths do not have large discrimination capability for small $\delta\theta$.

Comparison with other measures of orientation selectivity

Here, we compared Chernoff distance with several other measures of orientation selectivity such as CV (Swindale, 1998; Ringach et al., 2002), tuning width, and the ratio of the responses to preferred orientation and orthogonal orientation (Gegenfurtner et al., 1996). These measures were used as a measure of orientation selectivity without rigorous theoretical background. Here, we calculated each measure for 490 neurons and made scatterplots for various values of $\delta\theta$. These measures weakly correlated with Chernoff distance in general because Chernoff distance is proportional to the overall scale of a tuning curve and the three measures we are comparing do not depend on it. It means that they measure should be orientation selectivity in terms of the shape of tuning curve ignoring overall scale. Therefore, we compared these measures with Chernoff distance after factoring out the peak response.

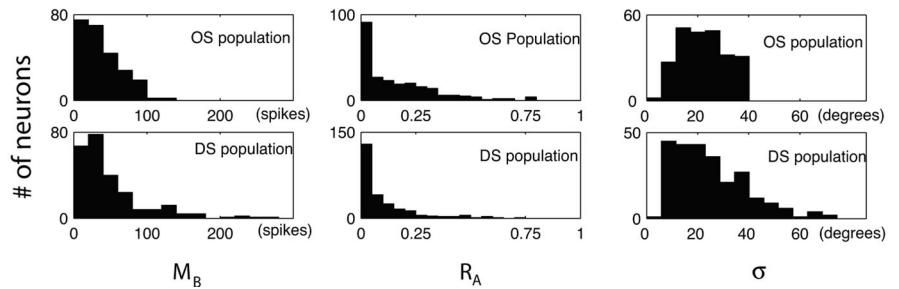


Figure 11. Histograms of peak response M_B , baseline R_A , and response tuning width σ for OS population (top graphs) and for DS population (bottom graphs). M_B is $A + \max\{B_1, B_2\}$. R_A is A/M_B , where A is the baseline of the tuning curve and M_B is the peak response. See Figure 4 for the description of the model of the tuning curve.

Table 1. Mean and median values of relative baseline R_A , peak response M_B , and tuning width σ

	R_A	M_B	σ
OS	0.16 (0.11)	38.6 (31.1)	22.2° (22.2°)
DS	0.10 (0.048)	49.7 (35.6)	24.8° (22.1°)

Table 2. Correlation coefficients between relative baseline R_A , peak response M_B , and tuning width σ

	R_A and M_B	M_B and σ	σ and R_A
OS	0.24 (<0.1%)	0.09 (0.14%)	0.39 (<0.1%)
DS	-0.1 (11.2%)	0.28 (<0.1%)	0.12 (5.7%)

Comparison with CV and Chernoff distance

For a given orientation tuning curve $\lambda(\theta)$, CV is defined in the following way:

$$CV = 1 - |f_2|/|f_0|, \quad (15)$$

f_n is $\int d\theta e^{in\theta} \lambda(\theta)$.

For a flat tuning curve, the CV is 1, and for a very narrow tuning curve with zero baseline, the CV is 0. Therefore, a bigger (smaller) CV is interpreted as a sign of lower (higher) orientation selectivity.

We found that the CV showed a very strong correlation with $D_C(\delta\theta)$ when $\delta\theta$ is smaller than 90° . It has strongest correlation with $D_C(\delta\theta)$ for $\delta\theta = 45^\circ$. Figure 13 shows three scatter plots between the CV and $D_C(\delta\theta)$ for $\delta\theta = 3, 45$, and 180° , respectively. The relationship between the CV and $D_C(\delta\theta)$ is very linear.

Our result shows that the CV is a good measure of orientation selectivity. But we also find that the CV behaves in a qualitatively opposite way to Chernoff distance sometimes. For example, we can calculate the CV and $D_C(\delta\theta)$ for our model tuning curve shown in Figure 4. For one case, we fixed relative baseline R_A to be 0 and changed σ from 8 to 40° . For another calculation, we fixed σ to be 20° and changed R_A from 0 to 0.2. Figure 14 illustrates the results. Because the smaller CV (larger $D_C(\delta\theta)$) represents higher orientation selectivity, a plot of the CV and $D_C(\delta\theta)$ should have a negative slope to be qualitatively correct. Figure 14 shows that, however, there are cases when the CV and $D_C(\delta\theta)$ are positively correlated. When tuning width σ is small, the orientation selectivity for $\delta\theta = 90^\circ$ increases, as we increase σ . The CV tells us, however, that orientation selectivity decreases. For smaller $\delta\theta$, the part of line (a) with positive slope is shorter so that this problem disappears. When σ is fixed to be 20° and R_A is changed, the line of the CV and $D_C(90^\circ)$ has a negative slope.

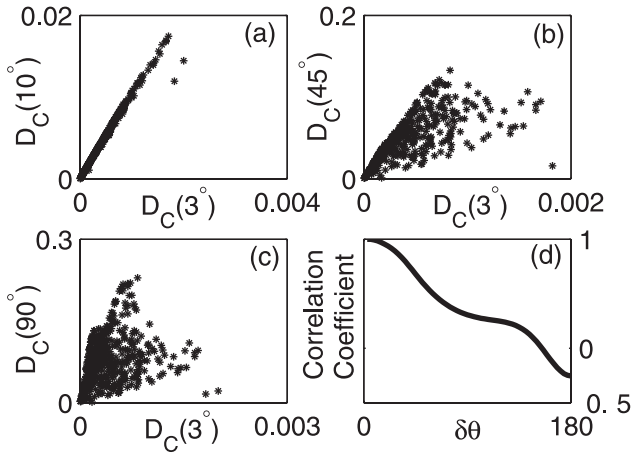


Figure 12. Scatter plots of $D_C(\delta\theta_1)$ and $D_C(\delta\theta_2)$ for 490 neurons in V1. For the three scatter plots, $\delta\theta_1 = 3^\circ$ and $\delta\theta_2 = 10, 45,$ and 90° , respectively. *d* is a plot of correlation coefficients between $D_C(3^\circ)$ and $D_C(\delta\theta)$ as a function of $\delta\theta$.

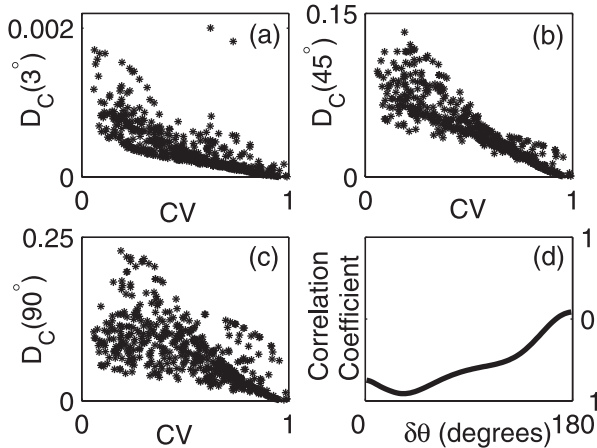


Figure 13. Correlation between CV and $D_C(\delta\theta)$. *a–c* are scatter plots for $\delta\theta = 3, 45,$ and 90° , respectively. *d* is a plot of correlation coefficients between R_A and $D_C(\delta\theta)$.

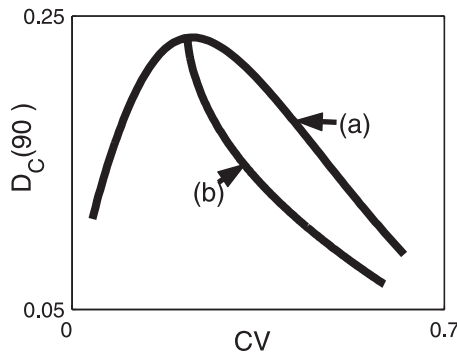


Figure 14. Plot of CV and $D_C(90^\circ)$. For the model of tuning curve, see Figure 4. For line (a), relative baseline $R_A = 0$ and tuning width σ is from 8 to 40° (left side is for smaller σ). For line (b), σ is 20° , and R_A is from 0 to 0.2 (left side is for smaller R_A).

Relative baseline and tuning width

In a previous section, we showed how $D_C(\delta\theta)$ depends on R_A and σ for idealized OS tuning curves. Here, we show correlations between $D_C(\delta\theta)$ and these quantities calculated for 490 neurons in V1.

Relative baseline R_A , the response to orthogonal orientation

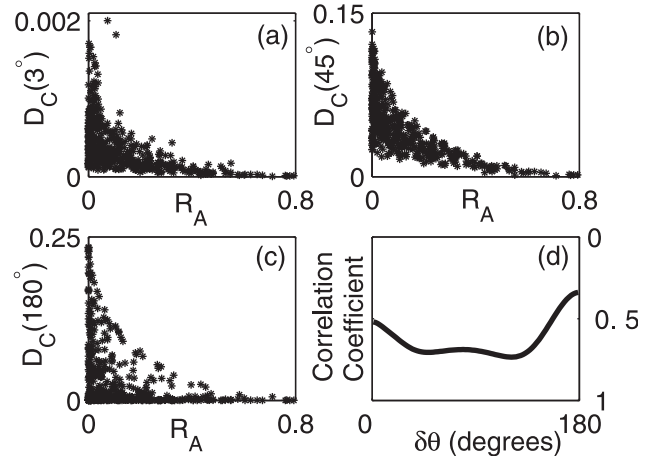


Figure 15. Correlation between R_A and $D_C(\delta\theta)$ in the V1 population. *a–c* are scatter plots for $\delta\theta = 3, 45,$ and 180° , respectively. *d* is a plot of correlation coefficients between R_A and $D_C(\delta\theta)$.

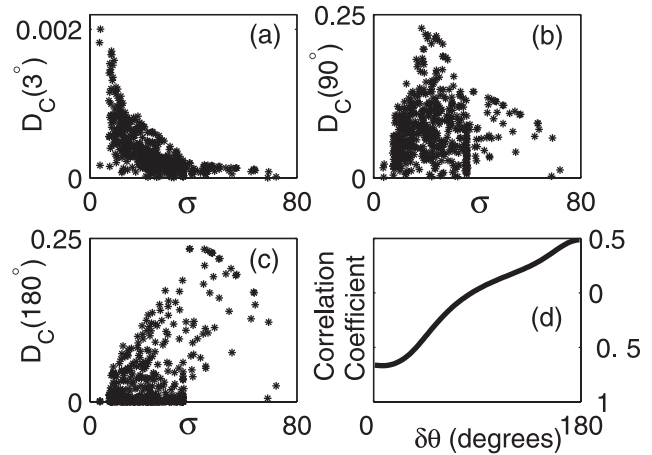


Figure 16. Correlation between σ and $D_C(\delta\theta)$ in the V1 population. *a–c* are scatter plots for $\delta\theta = 3, 90,$ and 180° , respectively. *d* is a plot of correlation coefficients between σ and $D_C(\delta\theta)$.

divided by the response to preferred orientation, is strongly correlated with $D_C(\delta\theta)$ for intermediate values of $\delta\theta$ (Fig. 15). R_A is weakly correlated with $D_C(\delta\theta)$ for small $\delta\theta$ because $D_C(\delta\theta)$ depends on σ more sensitively for smaller $\delta\theta$ (Fig. 8). When $\delta\theta$ is close to 180° , whether a neuron is DS or OS is a decisive factor for the discrimination capability. This makes R_A relatively less important in determining $D_C(\delta\theta)$.

Scatter plots between tuning width σ and $D_C(\delta\theta)$ have a bigger dispersion than for the CV or R_A versus $D_C(\delta\theta)$. Figure 16 shows that tuning width σ is strongly correlated with $D_C(\delta\theta)$ only for small $\delta\theta$ and for $\delta\theta$ close to 180° . This is partly because for a fixed R_A and small $\delta\theta$, $D_C(\delta\theta)$ decreases monotonically as σ increases. For a fixed R_A and $\delta\theta$ close to 180° , $D_C(\delta\theta)$ increases monotonically as σ increases. Because the relationship between $D_C(\delta\theta)$ and σ is linear, the correlation is strong there. For a fixed R_A and intermediate values of $\delta\theta$, $D_C(\delta\theta)$ maximizes at an optimal tuning width σ^* , and the relationship between $D_C(\delta\theta)$ and σ is convex. This makes the correlation coefficient small, but the small correlation coefficient is also because $D_C(\delta\theta)$ depends on R_A more sensitively than σ .

Discussion

Information measure for a population of neurons

Because many neurons in V1 have receptive fields at the same place or nearby places, it is natural to assess their discrimination capability in terms of population coding. However, it has been difficult to study population coding partly because it is difficult to calculate an information measure such as mutual information (Rolls et al. 1997; Panzeri et al., 1999) and the error of maximum-likelihood discriminator for a population of neurons. Chernoff distance often can be calculated when these measures are impossible to calculate. It is because sum of log is difficult when the log of a sum is tractable. Chernoff distance has analytical expressions for several important cases such as Poisson and Gaussian distributions. When the responses of neurons to given stimuli are independent of each other, the computational cost to calculate Chernoff distance increases linearly, not exponentially as the size of the neuronal population increases. Chernoff distance provides a clear interpretation through its relationships with mutual information, Fisher information, and the error of maximum-likelihood discrimination. Here, we calculated Chernoff distance for a population of neurons with tuning curves that are the same, except for preferred orientation. We considered homogeneous populations of neurons because we wanted to study how much contribution comes from such a population to the total discrimination power of the whole population of neurons in V1. Neurons in V1 have various shapes of tuning curves. The Chernoff distance for the whole population in V1 will be a sum of the Chernoff distances calculated for many homogeneous populations.

Information tuning curves

When we studied how the activities of a population of neurons represent a set of stimuli, tuning curves separately drawn for each neuron did not give much intuition. One natural idea may be to make a table of “distances” between pairs of stimuli in representation space of the population of neurons. This table may play the role of the tuning curve for a population of neurons. We used the Chernoff distance as a measure of the distance. For a population of neurons with preferred orientations that are distributed isotropically, this table of distances can be summarized by a curve. This information tuning curve helps us to study the relationship between the discrimination capability of a population of neurons and the shape of response tuning curves. Our method does not assume that it is for nearby angles, or for a small population of neurons, or for a readout with a specific form. Therefore, this method is more general than previous studies of population coding.

Discrimination capability and the shape of the response tuning curve

We introduce a Gaussian model of a response tuning curves of neurons in V1 to study the relationship between the discrimination capability of a population of neurons and the shape of response tuning curves. The discrimination capability of a neuron is very sensitive to its baseline activity R_A . A response tuning curve with a relative baseline R_A as large as 0.1 has significantly smaller discrimination capability than a tuning curve with no baseline. This result shows that it could be very wrong to subtract spontaneous activity level from evoked activity level in studying the discrimination capability of neurons. We found that the optimal tuning width σ^* is about $0.3 \delta\theta$ for small $\delta\theta$ and that σ^* has a value from 0 to 20° for any $\delta\theta$. Discrimination capability is more sensitive to σ for smaller $\delta\theta$.

Specialization and optimization of neurons in V1

We fit our model to the tuning curves of neurons in V1 and studied how these parameters of tuning curves are distributed in V1. The degradation of discrimination capability attributable to relative baseline R_A is small for most of the neurons in V1. OS neurons tend to have a bigger baseline relative to their peak response than DS cells. We found that the distribution of tuning width σ is relatively flat between 10 and 40° . This may suggest that different neurons are specialized for discriminations with different $\delta\theta$. But it also means that neurons with tuning width optimal for discrimination with $\delta\theta < 20^\circ$ do not exist in V1 because the optimal tuning width, σ^* , is $\sim 0.3 \delta\theta$. This means neurons in V1 are not optimized to discriminate nearby angles.

Relationship with other measures

We show the relationships between Chernoff distance with other measures of orientation selectivity. Several measures of orientation selectivity have been used without a theoretical background. Examples of such measures are CV, tuning width, and the ratio of the response to orthogonal orientation divided by the response to preferred orientation. For 490 neurons in V1, we calculated these values and compared them with the Chernoff distance for normalized tuning curves. It turns out that the CV showed an almost linear relationship with Chernoff distance. The CV shows the strongest correlation with $D_C(45^\circ)$. The ratio of the response to orthogonal orientation divided by the response to preferred orientation is relative baseline R_A . The Chernoff distance strongly correlates with it. Tuning width shows the weakest correlation with Chernoff distance among the three measures. It is mainly because the Chernoff distance is most sensitive to tuning width when tuning width and $\delta\theta$ are small. Such small tuning width does not exist in V1 (Rolls et al., 1997; Panzeri et al., 1999).

Applications to other sensory areas

It is natural to believe that population coding is being used in many different areas of the cortex because the same or similar information is often delivered by many neurons. But a satisfying measure of efficiency of population coding has been lacking. Many sensory stimuli such as sound patterns and odors are either complex or discrete by nature. For such cases, Chernoff distance can be useful to study the neuronal representation of various kinds of sensory information.

Appendix

Proof of Equation 11

Because we assumed that the statistics of the spike counts are Poisson, the mean spike count generated determines $P(\vec{r}|\theta)$ the probability distribution of spike count for a given direction of the stimulus. $\vec{r} = \{r^{1,1}, r^{2,1}, \dots, r^{N,1}, r^{1,2}, \dots, r^{N,2}\}$ is a vector of spike counts of $2N$ neurons, the mean of which value is $\vec{r} = \{\lambda^{1,1}, \lambda^{2,1}, \dots, \lambda^{N,1}, \lambda^{1,2}, \dots, \lambda^{N,2}\}$. $\lambda^{k,a}$ for θ is $\lambda[(\theta - \theta^k)(-1)^a]$, where $\theta^k = 360^\circ k/N$, $k = 0 \dots N - 1$, and $a = 1$ or 2 . k is an index for rotation of the tuning curve, and a is an index for reflection of the tuning curve.

$P(\vec{r}|\theta)$ is a product of $2N$ Poisson distributions:

$$\begin{aligned} P(\vec{r}|\theta) &= \prod_{a=1,2,k=1 \dots N} P(r^{k,a}|\theta) \\ &= \prod_{a=1,2,k=1 \dots N} \frac{(\lambda^{k,a}(\theta))^{r^{k,a}}}{r^{k,a}!} e^{-\lambda^{k,a}(\theta)}. \end{aligned} \quad (\text{A-1})$$

$Tr_{\vec{r}}$ in Equation 4 was summation over all possible values of \vec{r} . For this population of neurons, it has the following form:

$$Tr_{\vec{r}}f(\vec{r}) = \left[\prod_{a=1,2,k=1,\dots,N} \sum_{r^{k,a}=0}^{\infty} \right] f(\vec{r}). \quad (\text{A-2})$$

Inserting Equation A-2 into Equation 4 gives the following result:

$$D_{\alpha}(\theta_1, \theta_2) = -\log \left[\prod_{a=1,2,k=1,\dots,N} \sum_{r^{k,a}=0}^{\infty} P^{\alpha}(\vec{r}|\theta_1) P^{1-\alpha}(\vec{r}|\theta_2) \right] \quad (\text{A-3})$$

$$= - \sum_{a=1,k=0}^{a=2,k=N-1} \log \sum_{r^{k,a}=0} P^{\alpha}(r^{k,a}|\theta_1) P^{1-\alpha}(r^{k,a}|\theta_2). \quad (\text{A-4})$$

Remember that $D_{\alpha}(\theta_1, \theta_2)$ should be maximized in terms of α to get $D_C(\theta_1, \theta_2)$. Here is short proof that α^* , the value of α maximizing $D_{\alpha}(\theta_1, \theta_2)$, is 0.5 in this case because of orientation symmetry of the neuronal population. For each term in the summation in Equation A-4 with index k and $a = 1$, there exists another term with index k' and $a = 2$ such that $\lambda^{k,1}(\theta_1) = \lambda^{k',2}(\theta_2)$, and $\lambda^{k,1}(\theta_2) = \lambda^{k',2}(\theta_1)$. This means that $P(r^{k,1}|\theta_1) = P(r^{k',2}|\theta_2)$ and $P(r^{k,1}|\theta_2) = P(r^{k',2}|\theta_1)$, because mean values determine Poisson distributions. Now note that $D_{\alpha}(\theta_1, \theta_2)$ has the same value when we replace α with $1 - \alpha$ because $P(r^{k,1}|\theta_1)^{1-\alpha} P(r^{k,1}|\theta_2)^{\alpha} = P(r^{k',2}|\theta_2)^{1-\alpha} P(r^{k',2}|\theta_1)^{\alpha}$. Therefore, $\alpha^* = 1 - \alpha^*$ and α^* is 0.5.

We get the following result by inserting Equation A-1 into A-4:

$$D_{\alpha=0.5}(\theta_1, \theta_2) = - \sum_{a=1,k=0}^{a=2,k=N-1} \log \sum_{r^{k,a}=0} \sqrt{P(r^{k,a}|\theta_1) P(r^{k,a}|\theta_2)} \quad (\text{A-5})$$

$$= \sum_{a=1,k=0}^{a=2,k=N-1} \frac{1}{2} \left(\sqrt{\lambda^{k,a}(\theta_1)} - \sqrt{\lambda^{k,a}(\theta_2)} \right)^2. \quad (\text{A-6})$$

This is the derivation of Equation 11 in the text.

Relationship with Fisher information

Fisher information (Cover and Thomas, 1991; Seung and Sompolinsky, 1993; Abbott and Dayan, 1999; Sompolinsky et al., 2001) measures the estimation error of a continuous variable. For two separated angles, the error of the maximum-likelihood discriminator is determined by Fisher information when these two angles are very close to each other.

When $P(\vec{r}|\theta)$ is defined for a continuous variable, θ , and $\delta\theta = \theta_1 - \theta_2$ is much smaller than the width of the tuning curve, the Chernoff distance $D_C(\theta_1, \theta_2)$ is proportional to Fisher information, J (Cover and Thomas, 1991):

$$D_C(\theta_1, \theta_2) \approx \frac{J(\theta_1)}{8} \delta\theta^2 \quad (\text{A-7})$$

$$J(\theta) = Tr_{\vec{r}} P(\vec{r}|\theta) \left(\frac{\partial}{\partial \theta} \log P(\vec{r}|\theta) \right)^2. \quad (\text{A-8})$$

Relationship with mutual information

To measure the discrimination capability for any pair of orientations, we may calculate mutual information (Cover and Thomas, 1991; Rieke et al., 1997). Mutual information, I from information theory (Cover and Thomas, 1991), is defined in the following way:

$$I = \sum_{i=1,2} P(\theta_i) Tr_{\vec{r}} P(\vec{r}|\theta_i) \log \frac{P(\vec{r}|\theta_i)}{P(\vec{r})}. \quad (\text{A-9})$$

$P(\theta_i)$ is a priori probability of θ_i . $P(\vec{r}) = P(\theta_1)P(\vec{r}|\theta_1) + P(\theta_2)P(\vec{r}|\theta_2)$. As the difference between $P(\vec{r}|\theta_1)$ and $P(\vec{r}|\theta_2)$ increases, I converges to its maximum value [i.e., the entropy of stimuli, $H(\theta)$]:

$$H(\theta) = - \sum_{i=1,2} P(\theta_i) \log P(\theta_i). \quad (\text{A-10})$$

When I is close to $H(\theta)$ or $D_C(\theta_1, \theta_2) \gg 1$, there is an exponential relationship between mutual information I and Chernoff distance $D_C(\theta_1, \theta_2)$ (Kang and Sompolinsky, 2001):

$$H(\theta) - I \propto \exp[-D_C(\theta_1, \theta_2)]. \quad (\text{A-11})$$

References

- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.
- Bradley A, Skottun BC, Ohzawa I, Sclar G, Freeman RD (1987) Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior. *J Neurophysiol* 57:755–786.
- Cam LL, Yang GL (2000) *Asymptotics in statistics: some basic concepts*. New York: Springer.
- Cover TM, Thomas JA (1991) *Elements of information theory*. New York: Wiley.
- Gegenfurtner KR, Kiper DC, Fenstemaker SB (1996) Processing of color, form and motion in macaque V2. *Vis Neurosci* 13:161–172.
- Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. *Science* 233:1416–1419.
- Hawken MJ, Parker AJ (1990) Detection and discrimination mechanisms in the striate cortex of the Old-World monkey. In: *Vision: coding and efficiency* (Blakemore C, ed). New York: Cambridge UP.
- Hawken MJ, Parker AJ, Lund JS (1988) Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the Old World monkey. *J Neurosci* 8:3541–3548.
- Kang K, Sompolinsky H (2001) Mutual information of population codes and distance measures in probability space. *Phys Rev Lett* 86:4958–4961.
- Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21:227–277.
- Panzeri S, Treves A, Shultz S, Rolls ET (1999) On decoding the response of a population of neurons from short time windows. *Neural Comp* 11:1533–1577.
- Rieke F, Warland D, Steveninck RR, Bialek W (1997) *Spikes: exploring the neural code*. Cambridge, MA: MIT.
- Ringach DL, Shapley RM, Hawken MJ (2002) Orientation selectivity in macaque V1: diversity and laminar dependence. *J Neurosci* 22:5639–5651.
- Rolls ET, Treves A, Tovee MJ (1997) The representational capacity of the distributed encoding of information provided by population of neurons in primate temporal visual cortex. *Exp Brain Res* 114:149–162.
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci USA* 90:10749–10753.
- Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Phys Rev E* 64:051904.
- Swindale NV (1998) Orientation tuning curves: empirical description and estimation of parameters. *Biol Cybern* 78:45–56.
- Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey striate cortex. *Vision Res* 23:775–785.
- Vogels R, Orban GA (1990) How well do response changes of striate neurons signal differences in orientation: a study in the discriminating monkey. *J Neurosci* 10:3543–3558.