

Neuronal Tuning: To Sharpen or Broaden?

Kechen Zhang

Howard Hughes Medical Institute, Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, U.S.A.

Terrence J. Sejnowski

*Department of Biology, University of California, San Diego, La Jolla, CA 92093, U.S.A.
Howard Hughes Medical Institute, Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, U.S.A.*

Sensory and motor variables are typically represented by a population of broadly tuned neurons. A coarser representation with broader tuning can often improve coding accuracy, but sometimes the accuracy may also improve with sharper tuning. The theoretical analysis here shows that the relationship between tuning width and accuracy depends crucially on the dimension of the encoded variable. A general rule is derived for how the Fisher information scales with the tuning width, regardless of the exact shape of the tuning function, the probability distribution of spikes, and allowing some correlated noise between neurons. These results demonstrate a universal dimensionality effect in neural population coding.

1 Introduction ---

Let the activity of a population of neurons represent a continuous D -dimensional vector variable $\mathbf{x} = (x_1, x_2, \dots, x_D)$. Randomness of spike firing implies an inherent inaccuracy, because the numbers of spikes fired by these neurons differ in repeated trials; thus, the true value of \mathbf{x} can never be completely determined, regardless of the method for reading out the information. The Fisher information J provides a good measure on encoding accuracy because its inverse is the Cramér-Rao lower bound on the mean squared error:

$$E \left[\varepsilon^2 \right] \geq \frac{1}{J}, \quad (1.1)$$

which applies to all possible unbiased estimation methods that can read out variable \mathbf{x} from population activity without systematic error (Paradiso, 1988; Seung & Sompolinsky, 1993; Snippe, 1996). The Cramér-Rao bound can sometimes be reached by biologically plausible decoding methods (Pouget,

Zhang, Deneve, & Latham, 1999; Zhang, Ginzburg, McNaughton, & Sejnowski, 1998). Here the square error in a single trial is

$$\varepsilon^2 = \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_D^2, \quad (1.2)$$

with ε_i the error for estimating x_i . The Fisher information J can be defined by

$$\frac{1}{J} = \sum_{i=1}^D \left\{ \mathbb{E} \left[\left(\frac{\partial}{\partial x_i} \ln P(\mathbf{n} | \mathbf{x}, \tau) \right)^2 \right] \right\}^{-1}, \quad (1.3)$$

where the average is over $\mathbf{n} = (n_1, n_2, \dots, n_N)$, the numbers of spikes fired by all the neurons within a time interval τ , with the probability distribution P depending on the value of the encoded variable \mathbf{x} . The definition in equation 1.3 is appropriate if the full Fisher information matrix is diagonal. This is indeed the case in this article because we consider only randomly placed radial symmetric tuning functions for a large population of neurons so that the distributions of estimation errors in different dimensions are always identical, and uncorrelated. A recent introduction to Fisher information can be found in Kay (1993).

2 Scaling Rule for Tuning Width

The problem of how coding accuracy depends on the tuning width of neurons and dimensionality of the space being represented was first studied by Hinton, McClelland, and Rumelhart (1986) and later by Baldi and Heiligenberg (1988), Snippe and Koenderink (1992), Zohary (1992), and Zhang et al. (1998). All of these earlier results involved specific assumptions on the tuning functions, the noise, and the measure of coding accuracy.

Here we consider the general case using Fisher information as a measure and show that there is a universal scaling rule. This rule applies to all methods that can achieve the best performance given by the Cramér-Rao bound, although it cannot constrain the tuning properties of suboptimal methods.

The tuning function refers to the dependence of the mean firing rate $f(\mathbf{x})$ of a neuron on the variable of interest $\mathbf{x} = (x_1, x_2, \dots, x_D)$. We consider only radial symmetric functions

$$f(\mathbf{x}) = F\phi\left(\frac{|\mathbf{x} - \mathbf{c}|^2}{\sigma^2}\right), \quad (2.1)$$

which depend on only the Euclidean distance to the center \mathbf{c} . Here σ is the tuning width, which scales the tuning function without changing its shape, and F is the mean peak firing rate, given that the maximum of ϕ is 1. This general formula includes all radial symmetric functions, of which gaussian tuning is the special case: $\phi(z) = \exp(-z/2)$. Other tuning functions

can sometimes be transformed into a radial symmetric one by scaling or a linear transformation on the variable. If \mathbf{x} is a circular variable, the tuning equation, 2.1, is reasonable only for sufficiently narrow tuning functions, because a broad periodic tuning function cannot be scaled uniformly without changing its shape.

Suppose the probability for n spikes to occur within a time window of length τ is

$$P(n | \mathbf{x}, \tau) = S(n, f(\mathbf{x}), \tau). \quad (2.2)$$

This general formulation requires only that the probability be a function of the mean firing rate $f(\mathbf{x})$ and the time window τ .

These general assumptions are sufficient conditions to prove that the total Fisher information has the form

$$J = \eta \sigma^{D-2} K_\phi(F, \tau, D), \quad (2.3)$$

where η is a number density—the number of neurons whose tuning centers fall into a unit volume in the D -dimensional space of encoded variable, assuming that all neurons have identical tuning parameters and independent activity. We also assume that the centers of the tuning functions are uniformly distributed at least in the local region of interest. Thus, η is proportional to the total number of neurons that are activated. The subscript of K_ϕ implies that it also depends on the shape of function ϕ . Equation 2.3 gives the complete dependence of J on tuning width σ and number density η . The factor σ^{D-2} is consistent with the specific examples considered by Snippe and Koenderink (1992) and Zhang et al. (1998), but the exponent is off by one from the noiseless model considered by Hinton et al. (1986).

More generally, when different neuron groups have different tuning widths σ and peak firing rates F , we have

$$J = \eta \left\langle \sigma^{D-2} K_\phi(F, \tau, D) \right\rangle, \quad (2.4)$$

where the average is over neuron groups, and η is the number density including all groups so that J is still proportional to the total number of contributing neurons. Equation 2.4 follows directly from equation 2.3 because Fisher information is additive for neurons with independent activity.

Equations 2.3 and 2.4 show how the Fisher information scales with the tuning width in arbitrary dimensions D . Sharpening the tuning width helps only when $D = 1$, has no effect when $D = 2$, and reduces information encoded by a fixed set of neurons for $D \geq 3$ (see Figure 1A). Although sharpening makes individual neurons appear more informative, it reduces the number of simultaneously active neurons, a factor that dominates in higher dimensions where neighboring tuning functions overlap more substantially.

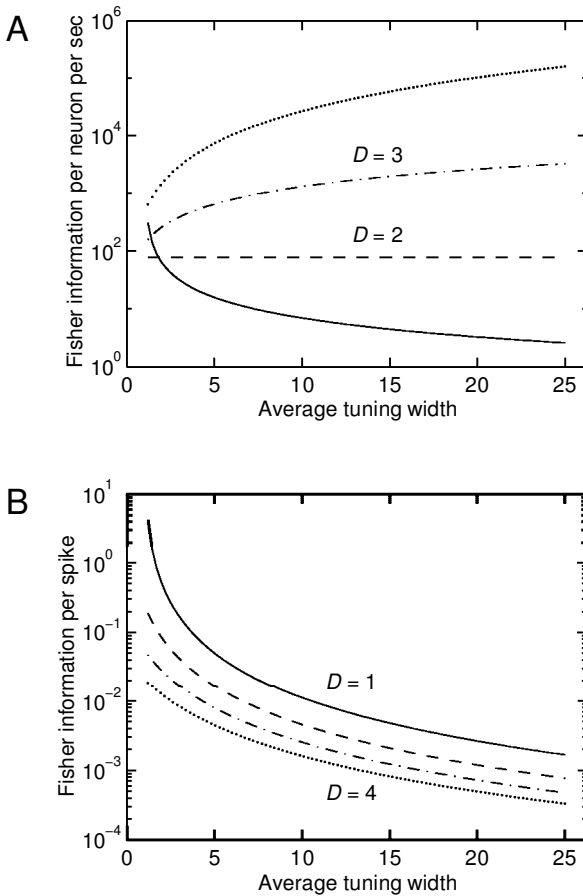


Figure 1: The accuracy of population coding by tuned neurons as a function of tuning width follows a universal scaling rule regardless of the exact shape of the tuning function and the exact probability distribution of spikes. The accuracy depends on the total Fisher information, which is here proportional to the total number of both neurons and spikes. (A) Sharpening the tuning width can increase, decrease, or not change the Fisher information coded per neuron, depending on the dimension D of the encoded variable, but (B) sharpening always improves the Fisher information coded per spike and thus energy efficiency for spike generation. Here the model neurons have gaussian tuning functions with random spacings (average in each dimension taken as unity), independent Poisson spike distributions, and independent gamma distributions for tuning widths and peak firing rates (the average is 25 Hz).

2.1 Derivation. To derive equation 2.3, first consider a single variable, say, x_1 , from $\mathbf{x} = (x_1, x_2, \dots, x_D)$. The Fisher information for x_1 for a single neuron is

$$J_1(\mathbf{x}) = \mathbb{E} \left[\left(\frac{\partial}{\partial x_1} \ln P(n | \mathbf{x}, \tau) \right)^2 \right] \quad (2.5)$$

$$= A_\phi \left(\frac{|\mathbf{x} - \mathbf{c}|^2}{\sigma^2}, F, \tau \right) \frac{(x_1 - c_1)^2}{\sigma^4}, \quad (2.6)$$

where the first step is a definition and the average is over the number of spikes n . It follows from equations 2.1 and 2.2 that

$$\begin{aligned} \frac{\partial}{\partial x_1} \ln P(n | \mathbf{x}, \tau) = \\ T \left(n, F\phi \left(\frac{|\mathbf{x} - \mathbf{c}|^2}{\sigma^2} \right), \tau \right) F\phi' \left(\frac{|\mathbf{x} - \mathbf{c}|^2}{\sigma^2} \right) \frac{2(x_1 - c_1)}{\sigma^2}, \end{aligned} \quad (2.7)$$

where $\phi'(z) = d\phi(z)/dz$ and function T is defined by

$$T(n, z, \tau) = \frac{\partial}{\partial z} \ln S(n, z, \tau). \quad (2.8)$$

Therefore, averaging over n must yield the form in equation 2.6, with function A_ϕ depending on the shape of ϕ . Next, the total Fisher information for x_1 for the whole population is the sum of $J_1(\mathbf{x})$ over all neurons. The sum can be replaced by an integral, assuming that centers of tuning functions are uniformly distributed with density η in the local region of interest:

$$J_1 = \eta \int_{-\infty}^{\infty} J_1(\mathbf{x}) dx_1 \cdots dx_D \quad (2.9)$$

$$= \eta \sigma^{D-2} \int_{-\infty}^{\infty} A_\phi(\xi^2, F, \tau) \xi_1^2 d\xi_1 \cdots d\xi_D \quad (2.10)$$

$$\equiv \eta \sigma^{D-2} K_\phi(F, \tau, D), \quad (2.11)$$

where new variables $\xi_i = (x_i - c_i)/\sigma$ have been introduced so that

$$\frac{|\mathbf{x} - \mathbf{c}|^2}{\sigma^2} = \xi_1^2 + \cdots + \xi_D^2 \equiv \xi^2, \quad (2.12)$$

$$dx_1 \cdots dx_D = \sigma^D d\xi_1 \cdots d\xi_D. \quad (2.13)$$

Finally, the Fisher information for all D dimensions is $J = J_1/D$, because the mean squared error in each dimension is the same. The result is equation 2.3.

2.2 Example: Poisson Spike Model. A Poisson distribution is often used to approximate spike statistics:

$$P(n | \mathbf{x}, \tau) = S(n, f(\mathbf{x}), \tau) = \frac{(\tau f(\mathbf{x}))^n}{n!} \exp(-\tau f(\mathbf{x})). \quad (2.14)$$

Then equation 2.4 becomes

$$J = \eta \left\langle \sigma^{D-2} F \right\rangle \tau k_\phi(D), \quad (2.15)$$

where

$$k_\phi(D) = \frac{4}{D} \int_{-\infty}^{\infty} \frac{(\phi'(\xi^2) \xi_1)^2}{\phi(\xi^2)} d\xi_1 \cdots d\xi_D, \quad (2.16)$$

with $\xi^2 = \xi_1^2 + \cdots + \xi_D^2$. For example, if the tuning function ϕ is gaussian,

$$k_\phi(D) = (2\pi)^{D/2} / D. \quad (2.17)$$

One special feature of Poisson spike model is that Fisher information in equation 2.15 is proportional to the peak firing rate F .

3 Fisher Information per Spike

The energy cost of encoding can be estimated by the Fisher information per spike:

$$J_{\text{spikes}} = J / N_{\text{spikes}}. \quad (3.1)$$

If all neurons have identical tuning parameters, the total number of spikes within time window τ is

$$N_{\text{spikes}} = \eta \int_{-\infty}^{\infty} \tau f(\mathbf{x}) dx_1 \cdots dx_D \quad (3.2)$$

$$= \eta \sigma^D F \tau \int_{-\infty}^{\infty} \phi(\xi^2) d\xi_1 \cdots d\xi_D \quad (3.3)$$

$$\equiv \eta \sigma^D F \tau Q_\phi(D), \quad (3.4)$$

where $f(\mathbf{x})$ is mean firing rate given by equation 2.1 and $\xi^2 = \xi_1^2 + \cdots + \xi_D^2$. More generally, when tuning parameters vary in the population, we have

$$N_{\text{spikes}} = \eta \left\langle \sigma^D F \right\rangle \tau Q_\phi(D). \quad (3.5)$$

For example,

$$J_{\text{spike}} = \frac{\langle \sigma^{D-2} \rangle}{D \langle \sigma^D \rangle} \quad (3.6)$$

holds when the neurons have gaussian tuning functions, independent Poisson spike distributions, and independent distributions of peak firing rates and tuning widths. As shown in Figure 1B, sharpening the tuning saves energy for all dimensions.

4 Scaling Rule Under Noise Correlation

The example in this section shows that scaling law still holds when firing-rate fluctuations of different neurons are weakly correlated, and the difference is a constant factor. Assume a continuous model for spike statistics based on multivariate gaussian distribution, where the average number of spikes n_i for neuron i is

$$\mu_i = E[n_i] = \tau f_i(\mathbf{x}) = \tau F \phi \left(\frac{|\mathbf{x} - \mathbf{c}_i|^2}{\sigma^2} \right), \quad (4.1)$$

and different neurons have identical tuning parameters except for the location of the centers. The noise correlation between neurons i and j is

$$C_{ij} = E[(n_i - \mu_i)(n_j - \mu_j)] = \begin{cases} C_i^2 & \text{if } i = j, \\ q C_i C_j & \text{otherwise,} \end{cases} \quad (4.2)$$

where

$$C_i = \psi(\mu_i) = \psi(\tau f_i(\mathbf{x})), \quad (4.3)$$

with ψ an arbitrary function. For example, $\psi(z) \equiv \text{constant}$ and $\psi(z) = az$ are the additive and multiplicative noises considered by Abbott and Dayan (1999), and $\psi(z) = \sqrt{z}$ corresponds to the limit of a Poisson distribution.

For large population and weak correlation, we obtain the Fisher information

$$J = \eta \sigma^{D-2} \left(\frac{1}{1-q} A_{\phi, \psi}(F, \tau, D) + \left(1 + \frac{1}{1-q} \right) B_{\phi, \psi}(F, \tau, D) \right), \quad (4.4)$$

ignoring contributions from terms slower than linear with respect to the population size. Here

$$A_{\phi, \psi}(F, \tau, D) = \frac{4\tau^2 F^2}{D} \int_{-\infty}^{\infty} \left(\frac{\phi'(\xi^2) \xi_1}{\psi(\zeta)} \right)^2 d\xi_1 \cdots d\xi_D, \quad (4.5)$$

$$B_{\phi, \psi}(F, \tau, D) = \frac{4\tau^2 F^2}{D} \int_{-\infty}^{\infty} \left(\frac{\phi'(\xi^2) \psi'(\zeta) \xi_1}{\psi(\zeta)} \right)^2 d\xi_1 \cdots d\xi_D, \quad (4.6)$$

with $\xi^2 = \xi_1^2 + \dots + \xi_D^2$ and

$$\zeta = \tau F\phi(\xi^2). \quad (4.7)$$

Thus, the only contribution of noise correlation is the constant factor $1/(1 - q)$, which slightly increases the Fisher information when there is positive correlation ($q > 0$). This result is consistent with the conclusion of Abbott and Dayan (1999). Notice that now the scaling rule for tuning width remains the same, and the Fisher information is still proportional to the total number of contributing neurons.

Equation 2.15 for the Poisson spike model can be recovered from equation 4.4 when $\psi(z) = \sqrt{z}$ with a large time window and high firing rates so that the contribution from ψ' or $B_{\phi, \psi}$ can be ignored. The only difference is an additional proportional constant $1/(1 - q)$.

5 Hierarchical Processing

In hierarchical processing, the total Fisher information cannot increase when transmitted from population A to population B (Pouget, Deneve, Ducom, & Latham, 1998, 1999). This is because decoding a variable directly from population B is indirectly decoding from population A and therefore must be subject to the same Cramér-Rao bound. Assuming a Poisson spike model with fixed noise correlation (cf. the end of section 4), we have

$$\frac{N_A}{1 - q_A} \langle \sigma^{D-2F} \rangle_A \geq \frac{N_B}{1 - q_B} \langle \sigma^{D-2F} \rangle_B, \quad (5.1)$$

where the averages are over all neurons of total numbers N_A and N_B in the two populations. This constrains allowable tuning parameters in the hierarchy.

6 Concluding Remarks

The issue of how tuning width affects coding accuracy was raised again recently by the report of progressing sharpening of tuning curves for interaural time difference (ITD) in the auditory pathway (Fitzpatrick, Batra, Stanford, & Kuwada, 1997). In a hierarchical processing system, the total information cannot be increased at a later stage by altering tuning parameters because of additional constraints such as inequality 5.1. (See the more detailed discussion by Pouget et al., 1998.)

For a one-dimensional feature such as ITD, more information can be coded per neuron for a sharper tuning curve, provided that all other factors are fixed, such as peak firing rate and noise correlation. For two-dimensional features, such as the spatial representation by hippocampal place cells, coding accuracy should be insensitive to the tuning width (Zhang et al., 1998).

In three and higher dimensions, such as the multiple visual features represented concurrently in the ventral stream of primate visual system, more information can be coded per neuron by broader tuning.

For energy consumption, narrower tuning improves the information coded per spike, provided that the tuning width stays large enough compared with the spacing of tuning functions. Therefore, it is advantageous to use relatively narrow tuning for one- and two-dimensional features, but there is a trade-off between coding accuracy and energy expenditure for features of three and higher dimensions.

The scaling rule compares different system configurations or the same system under different states, such as attention. For example, contrary to popular intuition, sharpening visual receptive fields should not affect how accurately a small, distant target can be localized by the visual system, because the example here is two-dimensional. The results presented here are sufficiently general to apply to neural populations in a wide range of biological systems.

Acknowledgments

We thank Alexandre Pouget, Richard S. Zemel, and the reviewers for helpful comments and suggestions.

References

- Abbott, L. F., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, *11*, 91–101.
- Baldi, P., & Heiligenberg, W. (1988). How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biological Cybernetics*, *59*, 313–318.
- Fitzpatrick, D. C., Batra, R., Stanford, T. R., & Kuwada, S. (1997). A neuronal population code for sound localization. *Nature*, *388*, 871–874.
- Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 77–109). Cambridge, MA: MIT Press.
- Kay, S. M. (1993). *Fundamentals of statistical signal processing: Estimation theory*. Englewood Cliffs, NJ: Prentice Hall.
- Paradiso, M. A. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics*, *58*, 35–49.
- Pouget, A., Deneve, S., Ducom, J.-C., & Latham, P. E. (1999). Narrow versus wide tuning curves: What's better for a population code?. *Neural Computation*, *11*, 85–90.
- Pouget, A., Zhang, K.-C., Deneve, S., & Latham, P. E. (1998). Statistically efficient estimation using population code. *Neural Computation*, *10*, 373–401.

- Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences USA*, *90*, 10749–10753.
- Snippe, H. P. (1996). Parameter extraction from population codes: A critical assessment. *Neural Computation*, *8*, 511–539.
- Snippe, H. P., & Koenderink, J. J. (1992). Discrimination thresholds for channel-coded systems. *Biological Cybernetics*, *66*, 543–551.
- Zhang, K.-C., Ginzburg, I., McNaughton, B. L., & Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *Journal of Neurophysiology*, *79*, 1017–1044.
- Zohary, E. (1992). Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biological Cybernetics*, *66*, 265–272.

Received January 13, 1998; accepted May 27, 1998.