Neural Computation. 11:85-90. 1999.

Narrow vs wide tuning curves: what's best for a population code?

Alexandre Pouget[1], Sophie Deneve[1], Jean-Christophe Ducom[1] and Peter E. Latham[2]

[1] Georgetown Institute for Computational and Cognitive Sciences
Georgetown University
Washington, DC 20007-2197
and
[2] Department of Neurobiology
University of California at Los Angeles
Los Angeles, CA 90095-1763

To whom correspondence should be addressed:
Alexandre Pouget
Georgetown Institute for Cognitive and Computational Sciences
New Research Building. Room EP04
3970 Reservoir Road NW
Washington, DC 20007-2197
Fax: (202) 687 0617
Email: alex@giccs.georgetown.edu

Acknowledgments: We thank Rich Zemel, Peter Dayan and Kechen Zhang for their comments on an earlier version of this manuscript.

Abstract

Neurophysiologists are often faced with the problem of evaluating the quality of a code for a sensory or motor variable, either to relate it to the performance of the animal in a simple discrimination task, or to compare the codes at various stages along the neuronal pathway. One common belief that has emerged from such studies is that sharpening of tuning curves improves the quality of the code, although only to a certain point beyond which further sharpening is believed to be harmful. We show that this belief relies on either problematic technical analysis or improper assumptions about the noise. We conclude that one cannot tell, in the general case, whether narrow tuning curves are better than wide ones; the answer depends critically on the covariance of the noise. The same conclusion applies to other manipulations of the tuning curve profiles such as gain increase.

It is widely assumed that sharpening tuning curves, up to a certain point, can improve the quality of a coarse code. For instance, attention is believed to improve the code for orientation by sharpening the tuning curves to orientation in the visual area V4 (Spitzer *et al.* , 1988). This belief comes partly from a seminal paper by Hinton, McClelland and Rumelhart (1986) which showed that there exists an optimal width for which the accuracy of a population code is maximized, suggesting that sharpening is beneficial when the tuning curves have a width larger than the optimal one. This result, however, was derived for binary units and does not readily generalize to continuous units.

A recent attempt to show experimentally that, for continuous tuning curves, sharper is better relied on the center-of-mass estimator to evaluate the quality of the code (Fitzpatrick *et al.* , 1997). These authors measured the tuning curves of auditory neurons to interaural time difference (ITD), a cue for localizing auditory stimuli. They argued that narrow tuning curves are better than wide ones —in the range they observed experimentally— in the sense that the minimum detectable change (MDC) in ITD is smaller with narrow tuning curves when using a center-of-mass estimator.

Their analysis, however, suffered from two problems: i) they did not consider a biologically plausible model of the noise and, ii) the MDC obtained with a center-of-mass is not, in the general case, an objective measure of the information content of a representation, because center-of-mass is not an optimal readout method (Snippe, 1996).

A better way to proceed is to use Fisher information, the square root of which is inversely proportional to the smallest achievable MDC independent of the readout method (Paradiso, 1988; Seung & Sompolinsky, 1993; Pouget *et al.* , 1998) (Shannon information would be another natural choice but it is simply, and monotonically, related to Fisher information in the case of population coding with a large number of units; see Brunel and Nadal, 1998. It thus yields identical results when comparing codes). To determine whether sharp tuning curves are indeed better than wide ones, one can simply plot the MDC obtained from Fisher information as a function of the width of the tuning curves. Fisher information is defined as:

$$I = E\left[-\frac{\partial^2}{\partial\theta^2}\log P(\mathbf{A}|\theta)\right], \tag{1}$$

where $P(\mathbf{A}|\theta)$ is the distribution of the activity conditioned on the encoded variable θ and $E[\cdot]$ is the expected value over the distribution $P(\mathbf{A}|\theta)$.

As we show next, sharpening increases Fisher information when the noise distribution is fixed, but sharpening can also have the opposite effect, i.e., it can decrease information, when the distribution of the noise changes with the width. The latter case, which happens when sharpening is the result of computation in a network, is the most relevant for neurophysiologists.

Consider first the case in which the noise distribution is fixed. For instance, for a population of N neurons with Gaussian tuning curves and *independent* Gaussian noise with variance σ^2, Fisher information reduces to:

$$I = \sum_{i=1}^{N}\frac{f_i'(\theta)^2}{\sigma^2}, \tag{2}$$

where $f_i(\theta)$ is the mean activity of unit i in response to the presentation angle, θ,

and $f_i'(\theta)$ is its derivative with respect to θ. Therefore, as the width of the tuning curve decreases, the derivative increases, resulting in an increase of information. This implies that the smallest achievable MDC goes up with the width of tuning as shown in figure 1-A, because the MDC is inversely proportional to the square root of the Fisher information. This is a case where narrow tuning curves are better than wide ones. Note, however, that the optimal tuning curve for Fisher information has zero width (or more precisely, a width on the order of $1/N$ where N is the number of neurons), unlike what Hinton *et al.* found for binary tuning curves. Note also that for the same kind of noise, the MDC measured with center-of-mass shows the opposite trend, i.e., wide is better, confirming that the MDC obtained with the center-of-mass does not reflect the information content of the representation [1].

Consider now a case in which the noise distribution is no longer fixed, such as in the two layer network illustrated in figure 1-B. The network has the same number of units in both layers and the output layer contains lateral connections which sharpen the tuning curves. This case is particularly relevant for neurophysiologists since this type of circuit is quite common in the cortex. In fact, some evidence suggests that a similar network is involved in tuning curve sharpening in the primary visual cortex for orientation selectivity (Ringach *et al.* , 1997).

Do the output neurons contain more information than the input neurons, just because they have narrower tuning curves? The answer is no, regardless of the details of the implementation, because processing and transmission *cannot* increase information in a closed system (Shannon & Weaver, 1963). Sharpening is done at the cost of introducing correlated noise among neurons, and the loss of information in the output layer can be traced to those correlations (Pouget & Zhang, 1996; Pouget *et al.* , 1998). This is a case where wide tuning curves (the ones in the input layer) are better than narrow ones (the ones in the output

[1] Fitzpatrick et al. (1997) reported the opposite result; they found sharp tuning curves to be better than wide ones when using a center-of-mass estimator. This is because the noise model they used is different from ours and biologically implausible.

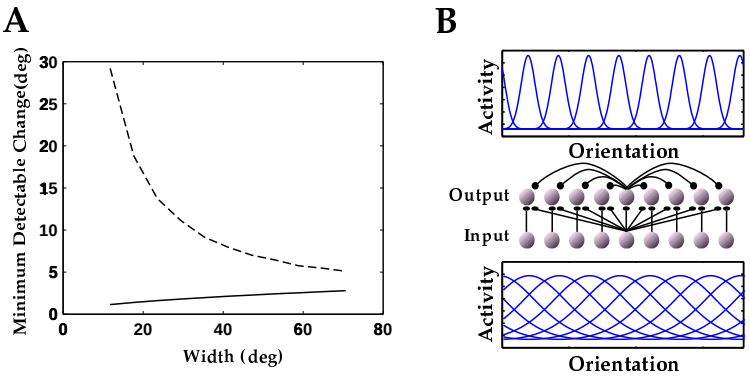

Figure 1: A- For a *fixed* noise distribution, the minimum detectable change (MDC) obtained from Fisher information (solid line) increases as the width increases. Therefore, in this case, narrow tuning curves are best, in the sense that they transmit more information about the presentation angle. Note that using a center-of-mass estimator (dashed line) to compute the MDC leads to the opposite conclusion, that wide tuning curves are best. This is a compelling demonstration that the center of mass is not a proper way to evaluate information content. B- A neural network with 10 input units and 10 output units, fully connected with feedforward connections between layers and lateral connections in the output layer. We show only one representative set of connections for each layer. The lateral weights can be set in such a way that the tuning curves in the output layer are narrower than in the input layer (see Pouget *et al.*, 1998, for details). Because the information in the output layer cannot be greater than the information in the input layer, sharpening tuning curves in the output layer can only decrease (or at best preserve) the information. Therefore, the wide tuning curves in the input layer contain more information about the stimulus than the sharp tuning curves in the output layer. In this case, *wide* tuning curves are best.

layer).

That wide tuning curves contain more information than narrow ones in this particular architecture, can be easily missed if one assumes the wrong noise distribution. Unfortunately, it is difficult to measure precisely the joint distribution of the noise or, even, its covariance matrix. It is therefore often assumed that the noise is independent among neurons when dealing with real data. Let's examine what happens if we assume independent noise for the output units of the network depicted in figure 1-B. We consider the case in which the output units are deterministic, i.e., the only source of noise is in the input activities, and the output tuning curves have the same width as the input tuning curves. We have shown in a previous paper (Pouget *et al.* , 1998) that, in this case, the network performs a close approximation to maximum likelihood and the noise in the output units is Gaussian with variance $f_i'(\theta)^2/I_1$, where I_1 is the Fisher information in the input layer. Using equation 2 for independent Gaussian noise we find that the information in the output layer, denoted I_2, is given by:

$$I_2 = \sum_{i=1}^{N} \frac{f_i'(\theta)^2}{f_i'(\theta)^2/I_1} = \sum_{i=1}^{N} I_1 = NI_1.$$

The independence assumption would therefore lead us to conclude that the information in the output layer is *much* larger than in the input layer, which is clearly wrong.

These simple examples demonstrate that a proper characterization of the information content of a representation must rely on an objective measure of information, such as Fisher information, and detailed knowledge of the noise distribution and its covariance matrix. (The number of variables being encoded is also critical, as shown by Zhang et al., in press). Using estimators such as the center-of-mass, or assuming independent noise, are not guaranteed to lead to the right answer. Therefore, attention may sharpen (Spitzer *et al.* , 1988) tuning curves (and/or increase their gain, McAdams and Maunsell, 1996) but whether this results in a better code is *impossible* to tell without knowledge of the covariance of the noise across conditions. Hopefully, the emergence of

multielectrode recordings will soon make it possible to measure these covariance matrices.

References

Brunel, N., & Nadal, J.P. 1998. Mutual information, Fisher information and population coding. *Neural Computation*, **In press**.

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. *Chap. 3, pages 318–362 of:* Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (eds), *Parallel Distributed Processing*, vol. 1. Cambridge, MA: MIT Press.

McAdams, C.J., & Maunsell, J.R.H. 1996. Attention enhances neuronal responses without altering orientation selectivity in macaque area V4. *In: Society For Neuroscience Abstracts*, vol. 22.

Paradiso, M.A. 1988. A theory of the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics*, **58**, 35–49.

Pouget, A., & Zhang, K. 1996. A statistical perspective on orientation selectivity in primary visual cortex. *In: Society For Neuroscience Abstracts*, vol. 22.

Pouget, A., Zhang, K., Deneve, S., & Latham, P.E. 1998. Statistically efficient estimation using population coding. *Neural Computation*, **10**, 373–401.

Ringach, D.L., Hawken, M.J., & R., Shapley. 1997. Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, **387**, 281–284.

Seung, H.S., & Sompolinsky, H. 1993. Simple model for reading neuronal population codes. *Proceedings of National Academy of Sciences. USA.*, **90**, 10749–10753.

Shannon, E., & Weaver, W. 1963. *The mathematical theory of communication.* Urbana and Chicago: University of Illinois Press.

Snippe, H.P. 1996. Parameter extraction from population codes: a critical assessment. *Neural Computation*, **8(3)**, 511–530.

Spitzer, H., Desimone, R., & Moran, J. 1988. Increased attention enhances both behavioral and neuronal performance. *Science*, **240(4850)**, 338–340.

Zhang, K., & Sejnowski, T.J. 1999. Neuronal tuning: to sharpen or broaden? *Neural Computation*, **11**.