Optimal adaptation of neural codes: An account of repetition suppression

Michael C. Mozer‡, **Richard S. Zemel**†, **Mathew Hungerford**‡

‡Department of Computer Science
University of Colorado
Boulder, CO 80309–0430
{mozer,hungerfo}@colorado.edu

†Department of Computer Science
University of Toronto
Toronto, Ontario M5S 1A1
zemel@cs.toronto.edu

Abstract

An important source of evidence concerning rapid adaptation and learning in the brain is the phenomenon of *repetition suppression*—the longlasting and item-specific decrease in neural responsivity with repeated exposure to an item, yielding a sparser representation. Existing accounts are informal and do little more than describe the phenomenon. In this paper, explore the hypothesis that repetition suppression can be characterized as the result of adaptation of the neural code with the rational goal of maximizing information transmission. Our work is based on two assumptions. First, in naturalistic environments, when an organism encounters some object, the object is more likely to be encountered again in the near future; consequently a probabilistic model of the environment must be updated following each experience. Second, neural codes are continually optimized to transmit information efficiently. Through simulation studies, we show that our model can explain the key phenomena of repetition suppression, and therefore serves as a framework for interpreting changes in cortical information processing with experience.

1 Introduction

A widespread and robust finding in primate electrophysiology is that neural responses decrease over repeated exposure to an item (e.g., [4, 10, 14]). Decreased activation is also observed in human imaging studies (e.g., [7]) and a reduction of waveform amplitude is observed in ERP studies [11]. This *repetition suppression* effect likely mediates the psychological phenomenon of repetition priming [14, 7]. The effect has been observed throughout cortex, from inferior temporal to prefrontal cortex, and the areas showing suppression are task specific. Roughly one quarter of inferior temporal neurons show repetition suppression. Some key findings to be explained concerning suppression include [4, 10, 14]: (1) Suppression is item specific, not general habituation. (2) Suppression does not depend on retinal size or location of an object, suggesting that it operates on a relatively abstract representation. (3) Suppression is observed even when the two repetitions are separated by 150 intervening items and delays up to 24 hours. (4) Suppression is graded, showing a continual reduction in firing rate with each presentation, plateauing at 40% of the initial firing rate. (5) Suppression depends merely on repetition, not behavioral significance.

(6) Suppression involves sharpening the neural representation—reducing the overall neural activity, decreasing the number of neurons involved in representing the item, and narrowing the tuning of neural receptive fields. (7) Suppression results in representations more resistant to stimulus degradation [9].

Existing accounts of repetition suppression are little more than descriptions of the data. Desimone suggests that cells unnecessary for identifying an item are suppressed, yielding a sparser and more selective representation, which presumably leads to a more efficient or rapid response [4]. Ringo proposes that suppression of familiar items may contribute to automatic orientation to novel items [10].

In this paper, we propose a preliminary account of repetition suppression based on the hypothesis that neural codes are optimally adapted to the statistical structure of the environment. Specifically, we conjecture that an encounter with an item leads to the expectation that the item will be more prevalent in the future. Because the item has a higher probability of being encountered in the future, it must be encoded more efficiently. Efficiency is characterized in terms of the rate at which information about the item is communicated.

2 Formal model

We formalize our analysis in a neural model. The model consists of a pool of n spiking neurons which encode one of m distinct items or *symbols* at a time, where m and n are roughly of the same order. Assuming a rate coding scheme, the neural activity for each symbol is specified by the vector of mean firing rates over the neurons; the n-dimensional firing rate vector for symbol s is denoted $\mathbf{f}(s)$, and the *codebook* $\mathbf{F}$ is defined as the set of firing rate vectors, $\mathbf{F} = \{\mathbf{f}(s_1), \mathbf{f}(s_2), \ldots, \mathbf{f}(s_m)\}$.

We assume that the pool of neurons is activated by an external environment that randomly selects symbols from a prior distribution, $\mathbf{P}(s) = \{P(s_1), P(s_2), \ldots, P(s_m)\}$, where $P(s)$ is the probability of symbol s. The chosen symbol is instantiated on the pool by setting the mean firing rate of each neuron i to $f_i(s)$, causing the neuron to generate spikes according to a noise model, e.g., independent Poisson noise with mean interspike interval $1/f_i(s)$.

In this paper, we argue that the repetition suppression phenomena fall out of two additional assumptions. These assumptions are *rational* in that they characterize the cognitive system as being well-designed, efficient, and adaptive.

(1) *Optimality of representation.* We assume that neural learning mechanisms produce a codebook that is in some sense optimized given the symbol probabilities, $\mathbf{p}$. One might define an optimization criterion in terms of how effectively the pool of neurons triggers an appropriate response in other pools of neurons, but this approach requires additional explicit assumptions concerning the mechanisms of neural signal propagation and transformation. Instead, we formulate a simple optimization criterion in terms of how well the neural spike train conveys information about the symbol being represented; formally, the measure is the mutual information between the symbol identity and the spike count observed in a given time window for each of the n neurons. Optimization is achieved via search in the $m \times n$ firing-rate space of the codebook. The optimality assumption is not meant to imply that the search finds a global optimizer, or even a local optimizer; for the sake of our arguments, we can rely on a weaker assumption that codebooks with higher mutual information are more likely to result than codebooks with lower mutual information. For example, consistent with this assumption is a search that produces codebooks with probability given by a Boltzmann distribution in which the negative mutual information serves as the free energy. Because our weak assumption suffices for this work, we need not propose specific search algorithms or speculate on their underlying neurobiological implementation.

(2) *Characteristics of environment.* We assume that in naturalistic environments, when an organism encounters some object, the object is more likely to be encountered again in the near future. This characteristic of the environment is reflected in an update rule that increases the probability p_s of symbol s, following experience with s. In accordance with assumption (1), when the symbol distribution is altered, the codebook is reoptimized.

To summarize, our hypothesis for explaining repetition suppression is as follows: perceptual experience leads to revision of a probabilistic model of the environment, which in turn necessitates adaptation of the neural code to preserve its efficiency. Intuitively, the optimization criterion yields a codebook in which the highest frequency symbols are transmitted most efficiently. A symbol is transmitted efficiently when its neural code is distinct from other symbols. One way to ensure distinctiveness is with a more localist representation, the sort of representation implied by the phenomena of repetition suppression. Although distinctiveness does not directly imply local representations, we will show that standard coding assumptions naturally produce this relationship.

3 Related research

Coding efficiency has long been suggested as a first principle governing neural representations.Past theoretical and experimental research have used this principle to produce parsimonious explanations of neural response properties. Wainwright also starts with the rational assumption that neural response properties are adjusted to the changing statistics of the environment to optimize information transmission; via this assumption, he explains phenomena of visual adaptation in orientation, spatial frequency, and motion [13]. Optimal coding accounts have also explained response properties of neurons in primary sensory cortex [12] and tuning function widths [5, 8, 15]. The tuning-function accounts have been based on Fisher information, which has limited validity for measuring the precision of a population code [2, 3].

Our work is distinguished from earlier efforts in several respects. First, prior research has not addressed repetition suppression. Second, because repetition suppression is obtained in higher regions of the brain (inferior temporal cortex and association areas), we are concerned with representing abstract object-identity information; in contrast, previous work has focused on primitive visual features of the sort encountered in primary visual cortex. Third, the objects with which we are concerned may be quite complex, and thus must be modeled in a high-dimensional, discrete representational space; in contrast, most previous research has focused on low-dimensional, continuous spaces (e.g., line orientation). Fourth, we require neurons whose response functions are arbitrary and symbol dependent, as embodied in the codebook; neurons with unimodal tuning functions are inadequate for our domain. Given these properties of our domain, Fisher information is not a useful measure because it requires a neural response function that is differentiable w.r.t. a stimulus (feature or symbol) space. Instead, we optimize mutual information directly.

4 Details of approach

We formulate the problem of optimizing codes in terms of the mutual information (MI) between the population responses $\mathbf{r}$ and the symbols s:

$$MI = \sum_{\mathbf{r},s} P(\mathbf{r},s) \log \frac{P(\mathbf{r}|s)}{P(\mathbf{r})} \tag{1}$$

MI is difficult to calculate exactly, first because knowledge of $P(\mathbf{r}|s)$ is required for each symbol s. We assume knowledge of the codebook, which describes the expected values of these conditional firing rate distributions. We also assume that the noise model, which describes the variability in $\mathbf{r}$ for a given s, is also known. Standard forms for this noise model are independent Poisson noise, and independent or correlated Gaussian noise. We discuss these further below.

Even if we have estimates for the population firing rate distributions, computing MI entails summing over all possible combinations of $\mathbf{r}$ and s, which quickly becomes intractable for reasonable-sized populations with large dynamic ranges encoding several symbols. We have explored two approaches to optimizing the codebook with respect to MI.

The hypothesis underlying the first approach is that not all values of $\mathbf{r}$ need be considered, but instead that MI can be approximated using samples of $\mathbf{r}$. We adopt this approach to approximate MI with samples, and optimize the codebook by hill-climbing in this quantity. Under a standard independent Poisson noise model, where a neuron's response variance equals its mean response to a symbol, the gradients for the codebook are:

$$\frac{\partial MI}{\partial \mathbf{f}(s)} = \frac{P(s)}{\mathbf{f}(s)} \Big(\sum_{\mathbf{r}} P(\mathbf{r}|s) \log \frac{P(\mathbf{r}|s)}{P(\mathbf{r})} (\mathbf{r} - \mathbf{f}(s)) \Big)$$

Intuitively, hill-climbing in MI using these gradients moves the codebook towards each population response in proportion to the information provided by that response about s.

We also explored a second approach to optimizing the codebook, which entails optimizing an upper bound to MI, which we term MIU. Applying Jensen's inequality ($\log E[P(\mathbf{r}|s)] \geq E[\log P(\mathbf{r}|s)]$) to the denominator in MI (Equation 1), we obtain:

$$\begin{aligned} MI \leq MIU &= \sum_{s,s'} P(s)P(s') \sum_{\mathbf{r}} P(\mathbf{r}|s) \log \frac{P(\mathbf{r}|s)}{P(\mathbf{r}|s')} \\ &= \sum_{s,s'} P(s)P(s') KL[P(\mathbf{r}|s); P(\mathbf{r}|s')] \end{aligned}$$

Under the standard Poisson noise model, MIU simplifies to this form: $\sum_{s,s'<s} P(s)P(s')[\sum_i (f_i(s) - f_i(s'))(\log f_i(s) - \log f_i(s'))]$. Optimizing MI under this bound and noise model clearly involves simply separating the codebook entries.

We have also explored a more complicated noise model in which the noise between neurons is correlated. A multiplicative correlated Gaussian noise model produces not only variances that increase as a function of firing rate (as in the Poisson noise model), but also larger correlations for neurons with similar tuning functions, as is seen in neurophysiological data [6]. Under this model, where $Q_{ij}(s) = \sigma^2[\delta_{ij} + c(1 - \delta_{ij})] f_i(s) f_j(s)$ describes the noise covariance in neuron i and j's response to s, and $0 \leq c \leq 1$ is a correlation coefficient,

$$\begin{aligned} MIU &= \sum_{s,s'<s} P(s)P(s') \Big(N - 1/2 \sum_{i,j=1}^{N} K_{ij}(s,s') \Big) \\ K_{ij}(s,s') &= Q_{ij}^{-1}(s') Q_{ij}(s) + Q_{ij}^{-1}(s) Q_{ij}(s') + \\ &\qquad (Q_{ij}^{-1}(s) + Q_{ij}^{-1}(s'))(f_i(s') - f_i(s))(f_j(s') - f_j(s)) \end{aligned}$$

Under both of these noise models, independent Poisson and correlated Gaussian, the codebook can be optimized by hill-climbing in MIU.

5 Simulation experiments

5.1 MI versus MIU

To provide intuitions concerning our approach, we explore a small domain consisting of two neurons encoding three symbols. The firing rate of each neuron is allowed to range

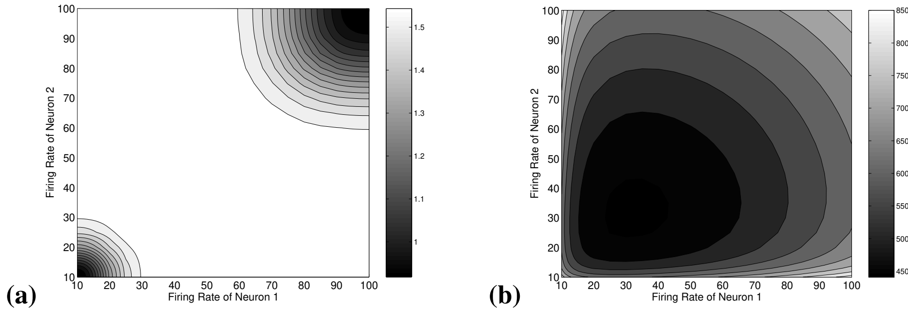

Figure 1: (a) Contour map of MI as a function of x and y for a 3-symbol codebook with uniform priors and codes at (10, 10), (100, 100), and (x, y); (b) the upper bound on mutual information (MIU) for the same system.

from 10 to 100 spikes/second. We assume the three symbols have equal priors. We assign codes to two symbols, $\mathbf{f}(s_1) = (10, 10)$ and $\mathbf{f}(s_2) = (100, 100)$, and vary the encoding of the third symbol. Figure 1a shows a contour plot of the codebook mutual information as a function of the third symbol's firing rate vector, $\mathbf{f}(s_3)$. Placing s_3 near s_1 or s_2 results in low MI. In the constrained space, the MI maxima are obtained in the corners, $\mathbf{f}(s_3) = (100, 10)$ and $\mathbf{f}(s_3) = (10, 100)$, where the MI—to four significant digits—is 1.5850. The maximum possible MI, the entropy of the prior distribution, is also 1.5850; thus, the symbols lie enough apart in the space so as to have little overlap in their firing rate distributions. Although placing s_3 in one of the empty corners yields maximum MI, MI barely varies along the upper-left to lower-right diagonal. For example, the MI for $\mathbf{f}(s_3) = (55, 55)$ is 1.5849. Thus, by our weak optimality-of-representation assumption, all points along the diagonal are likely codes for s_3.

The contour map hints at a key property of our model: MI drops when s_3 is placed in the neighborhood around either (10, 10) or (100, 100), but the (100, 100) neighborhood is larger than the (10, 10) neighborhood. This difference reflects the fact that the variance of a Poisson neuron is equal to its mean. Because a neuron with a higher firing rate will have greater variance, a coding cost is incurred to assigning high firing rates in the codebook. Thus, the pressure for coding efficiency results in sparse codes—codes that minimize the number of neurons active and the activity of neurons. Van Vreeswijk established this consequence of the Poisson noise model analytically for short time windows [12], although his argument was framed specifically in terms of V1 cortical neurons: MI was measured between input images and V1 neurons, where a particular nonlinear response function of V1 neurons was assumed.

Figure 1b shows the same example as Figure 1a panel with the MIU measure instead of MI. Although the optima of the two functions are the same—the (10, 100) and (100, 10) corners— MIU is not a good proxy for MI in that the two functions have little in common, and the MIU has local optima at (10, 10) and (100, 100). Based on many experiments utilizing MIU, we have unfortunately concluded that MIU is not a useful measure. In retrospect, it is not altogether surprising that using an upper bound to search for maxima would lead to failure.

5.2 Integration time window

Computation of MI involves a summation over every possible neural response vector. A neural response vector consists of spike counts for each neuron. Spikes counts are obtained by sampling the pool for a time window of ω msec. In Figure 1, $\omega = 1000$. Figure 2

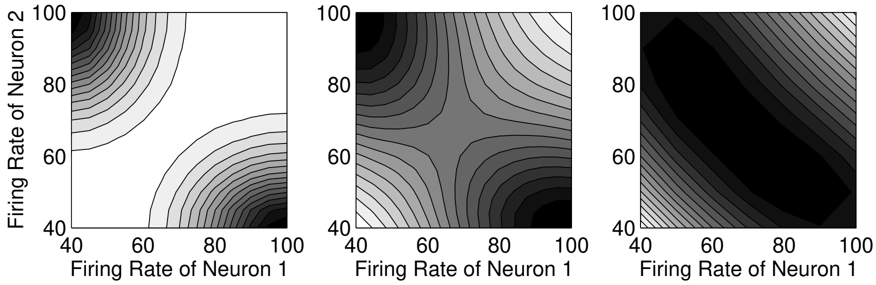

Figure 2: Contour maps of MI based on a sampling time window ω of 1000 msec (left panel), 250 msec (middle panel), and 100 msec (right panel). Each map depicts the MI as a function of x and y for a 3-symbol codebook with uniform priors and codes at (40, 100), (100, 40), and (x, y). White indicates high MI values.

shows the dependence of MI on ω. The example is similar to the previous one: s_1 is fixed at (40, 100) and s_2 is fixed at (100, 40), and the location of s_3 is allowed to vary in the (40, 40) to (100, 100) region.[1] Each contour map depicts MI for the codebook as a function of $\mathbf{f}(s_3)$, with black and white being low and high MI, respectively. The three contour maps are for $\omega = 1000$, $\omega = 250$, and $\omega = 100$, from left to right. With a long integration time, s_3 can be placed along the main diagonal and not be confused with s_1 and s_2. However, as the integration time is reduced, the coarser spike count information makes it difficult to discriminate s_3 from s_1 and s_2 (as well as s_1 and s_2 from one another). Short integration times are the most informative when the aim is to achieve efficient information transmission [12]; at asymptotically long integration times, any two symbols with distinct codebook entries can be discriminated. Thus, the remainder of our simulations will explore a short (100 msec) integration time.

5.3 Varying symbol priors

We now move on to simulations that address the core of our hypothesis: that repetition suppression can be explained in terms of updating symbol priors and the reoptimization of the codebook. We continue with a simple 3 symbol example in a two-dimensional neural activation space. Figure 3 shows the MI contour map for the cases where the three symbols have uniform priors (left panel), and where the prior of s_3 is 15 times that of s_1 and s_2, respectively (right panel). For the uniform priors, $\mathrm{MI}_{(40,\ 40)}$—that is, MI with s_3 at (40, 40)—is about the same as $\mathrm{MI}_{(100,\ 100)}$, whereas when s_3 has a higher prior, $\mathrm{MI}_{(40,\ 40)} > \mathrm{MI}_{(100,\ 100)}$. Figure 4 shows this effect more quantitatively. Note that the MI ratio is always greater than 1, indicating that the code with low firing rates is preferred to the code with high firing rates, and that this preference increases with the prior. Thus, we have illustrated the essential feature of repetition suppression: that repetition of an item, through the increase in its prior and reoptimization of the codebook, leads to a sparser representation. Another central feature of repetition suppression is that the effect diminishes with each repetition and reaches a plateau. We also find this feature in Figure 4, as the

[1] The lower bound of 40 spikes/sec on the mean firing rate may seem uncharacteristic of cortical neurons, but the effects we obtained did not qualitatively depend on either the lower or upper bound. By limiting the range, we observe more interactions among symbols in the codebook, which reduces the effective capacity of each neuron and allows us to explore smaller ensembles of neurons—important given the computation intensive nature of computing MI. Further, we are interested in rapid information transmission, say in a time window of 50 or 100 msec. With low firing rates (say 10 Hz) and a narrow time window, the variability of the neural response is limited due to quantization effects (a neuron will most likely produce 0 or 1 spikes).

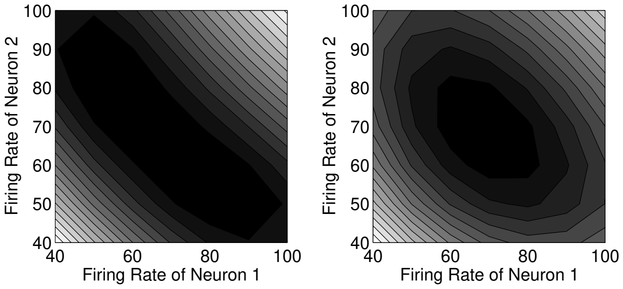

Figure 3: Each contour map depicts the MI as a function of x and y for a 3-symbol codebook with codes at (40, 100), (100, 40), and (x, y). In the left panel, the priors are uniform; in the right panel the prior of the symbol 3 is 15 times that of symbols 1 and 2.

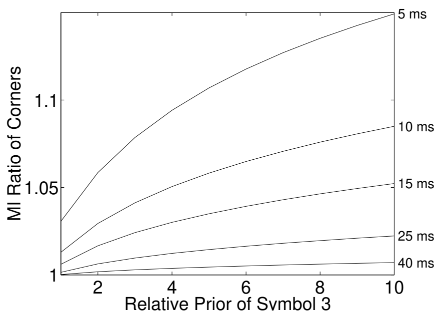

Figure 4: The ratio $\mathrm{MI}_{(40,\ 40)}/\mathrm{MI}_{(100,\ 100)}$ as a function of the relative prior of s_3 (along the x-axis) and the time window ω (the distinct curves).

curves are negatively accelerated.

We are confident that our results can be extrapolated to larger symbol sets and and higher dimensional neural activity spaces, for the following reason. We can view our simple example as being embedded in a much larger domain, where the other symbols are sufficiently distant in neural activity space from the three we consider that interactions do not arise. That is, in the subspace we consider, MI is not affected by the exact locations of the other symbols, nor do small changes to their locations affect MI.

The notion of a large virtual symbol set also addresses the question of what is a plausible change to the priors as a result of a single experience. With three symbols, altering a uniform prior distribution to achieve one in which one symbol is 15 times more likely than the others (as in Figure 3) would involve massive changes to the distribution, i.e., increasing the experienced symbol's prior from .33 to .88. However, with a large symbol set, the magnitude of the change to the prior distribution is much smaller, e.g., with 1000 symbols the change would be from .001 to .014. One can envision many simple learning rules that produce small incremental updates to the prior distribution as items are experienced.

In a final and slightly larger simulation, we searched for the optimal codebook for a three neuron, three symbol problem with a biased distribution that makes ones symbol ten times as likely as the other two—the *primed* symbol. The search began with a random initialization of the codebook entries to values in [40, 100]. A gradient search was conducted for twenty steps, which pushed the symbols apart, but not necessarily to corners of the space. Four thousand replications of the search were run, and the mean firing rate of neurons was computed for the primed and unprimed symbols. Neurons encoding the primed symbol had reliably lower firing rates than neurons encoding the unprimed symbol (by an analysis of variance, $F(1, 3999) = 6.48, p < .02$).

6 Discussion and conclusions

In this paper, we explored the hypothesis that repetition suppression can be characterized as the result of adaptation of the neural code with the rational goal of maximizing information transmission. We demonstrated the plausibility of this hypothesis via simulation studies. The hypothesis and our simulation studies are consistent with the first six properties of repetition suppression listed in the Introduction. Although the present work is neutral with regard to the seventh property, we plan to explore a model in which similarity structure is imposed in the symbol space, allowing us to examine sharpening of tuning functions and greater noise robustness due to repetition.

Although repetition suppression is a widespread and robust phenomenon in neurophysiology, it has not been explained in a coherent formal framework. Our work is a first cut at establishing such a framework, one that characterizes the efficiency of information transmission in neocortex and its change with experience.

Acknowledgements

We thank Joshua Snyder for helpful discussions and assistance at the outset of this research. This research was supported by Grant 97–18 from the McDonnell-Pew Program in Cognitive Neuroscience, NSF award IBN–9873492, and NIH/IFOPAL R01 MH61549–01A1.

References

[1] M. Bethge, D. Rotermund, and K. Pawelzik. Optimal short-term population coding: when Fisher information fails. *Neural Computation*, in press, 2002.

[2] N. Brunel and J.-P. Nadal. Mutual information, Fisher information and population coding. *Neural Computation*, *10*, 1731–1757, 1998.

[3] R. Desimone. Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Science*, *93*, 13494–13499, 1996.

[4] C.W. Eurich and S.D. Wilke. Multi-dimensional encoding strategy of spiking neurons. *Neural Computation*, *12*, 1519–1529, 2000.

[5] D. Lee, N.L. Port, W. Kruse, and A.P. Georgopoulos. Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *Journal of Neuroscience, 18*, 1161-1170, 1998.

[6] R.A. Poldrack and J.D. Gabrieli. Characterizing the neural mechanisms of skill learning and repetition priming. Evidence from mirror reading. *Brain*, *124*, 67–82, 2001.

[7] A. Pouget, S. Denever, J.C. Ducom, and P. Latham. Narrow vs wide tuning curves: what's best for a population code? *Neural Computation*, *11*, 85–90, 1999.

[8] G. Rainer and E.K. Miller. Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, *27*, 179–189, 2000.

[9] J.L. Ringo. Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. *Behavioral Brain Research*, *40*, 191–197, 1996.

[10] M.D. Rugg, M. Soardi, and M.C. Doyle. Modulation of event–related potentials by the repetition of drawings of novel objects. *Cognitive Brain Research*, *3*, 17–24, 1995

[11] C. van Vreeswijk. Whence sparseness? In T. K. Leen, T. Dietterich, & V. Tresp (Eds.), *NIPS 13* (pp 180–186). Cambridge, MA: MIT Press, 2002.

[12] M.J. Wainwright. Visual adaptation as optimal information transmission. *Vision Research*, *39*, 3960–3974, 1999.

[13] C.L. Wiggs and A. Martin. Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*, 227–233, 1998.

[14] K. Zhang and T.J. Sejnowski. Neural tuning: To sharpen or broaden? *Neural Computation*, *11*, 75–84, 1999.