A theoretical model of neuronal population coding of stimuli with both continuous and discrete dimensions

Valeria Del Prete*and Alessandro Treves
SISSA, Programme in Neuroscience, via Beirut 4, Trieste, Italy

May 15, 2001

Abstract

In a recent study the initial rise of the mutual information between the firing rates of N neurons and a set of p discrete stimuli has been analytically evaluated, under the assumption that neurons fire independently of one another to each stimulus and that each conditional distribution of firing rates is gaussian. Yet real stimuli or behavioural correlates are high-dimensional, with both discrete and continuously varying features. Moreover, the gaussian approximation implies negative firing rates, which is biologically implausible. Here, we generalize the analysis to the case where the stimulus or behavioural correlate has both a discrete and a continuous dimension, like orientation and shape could be in a visual stimulus, or type and direction in a motor action. The functional relationship between the firing patterns and the continuous correlate is expressed through the tuning curve of the neuron, using two different parameters to modulate its width and its flatness. In the case of large noise we evaluate the mutual information up to the quadratic approximation as a function of population size. We also show that in the limit of large N and assuming that neurons can discriminate between continuous values with a resolution $\Delta\vartheta$, the mutual information grows to infinity like $log(1/\Delta\vartheta)$ when $\Delta\vartheta$ goes to zero. Then we consider a more realistic distribution of firing rates, truncated at zero, and we prove that the resulting correction, with respect to the gaussian firing rates, can be expressed simply as a renormalization of the noise parameter. Finally, we demonstrate the effect of averaging the distribution across the discrete dimension, evaluating the mutual information only with respect to the continuously varying correlate.

Introduction

The strategy used by populations of neurons to code for external stimuli or behavioural correlates is a major issue which has been recently investigated both through data analysis and theoretical modelling. The mutual information between external correlates and the spiking activity of the population is one way to assess such coding quantitatively [1]. Several analyses have focused on the coding of a *discrete* set of stimuli ([2], see [3] for a review), which is the paradigm used in many experiments[4, 5, 6, 7]. In this situation the mutual information is bounded by the entropy of the stimulus set. Some theoretical studies have also considered the coding of stimuli varying in a *continuous* domain [8, 9], which is interesting with respect to basic properties like orientation in

*delprete@sissa.it

visual stimuli, frequency in auditory stimuli, velocity and position in motor actions. In particular in [9] the authors have studied the asymptotic (large population) behaviour of the mutual information, with respect to a stimulus with a continuously varying dimension.

Yet no study has been proposed so far considering a mixture of both continuous and discrete features, which is obviously closer to real world stimuli or behavioural correlates. Moreover the initial rise of the mutual information for small but increasing population size is more relevant for a comparison with estimates from real data, at least as far as the possibility of having simultaneous recordings from very large populations of neurons is restricted to very few cases.

We have recently analyzed data recorded in the motor areas of behaving monkeys, in the laboratory of Eilon Vaadia. The monkeys moved a manipulandum in several possible directions (approximating a continuous correlate) and with different combinations of arms (4 *types* of movement, i.e. a discrete correlate). In trying to characterize the neural coding of these movements, we were particularly interested in whether, as it is reasonable to expect, different motor areas differ, at least quantitatively, in their coding properties. The results, obtained from records of activity in areas M1 and SMA, will be reported elsewhere [10]; but they suggest the importance of developing theoretical models of how populations of neurons might code simultaneously discrete and continuous correlates. For example, one clear conclusion has been that type and direction are not *independent* dimensions of the movement, in the sense for example that the information about direction, extracted from the activity recorded with all movement types, is much lower (roughly half) of the average information about direction, obtained with a single movement type.

Can we embody similar properties in a model of the scheme used by neurons to code movements? What would then be the dependence of the mutual information on the number of the possible types of movement? How would it depend on the resolution with which the continuous correlate is sampled? How on the level of noise affecting the firing patterns?

In a recent work [2] some of these questions have been investigated for a set of p discrete stimuli, under the assumption that neurons fire independently of one another to each stimulus and that the distribution of the firing rates is gaussian. The linear and the quadratic approximations to the mutual information as a function of population size were studied analytically, in the limit of large noise, as well as the approach to the ceiling in the case of small noise. We generalize this study considering both a discrete and a continuous dimension in the stimulus, referring specifically to motor actions characterized by a direction and a "type". Nonetheless, our model is equally applicable to other complex correlates. We introduce a more realistic conditional firing rate distribution, than the simple gaussian one, and find a simple resulting correction to the gaussian model: the analytical expression of the mutual information remains the same, except for a renormalization of the expansion parameter.

We then evaluate the information loss when the original activity distribution is averaged across the discrete correlate, as is sometimes the case in the analysis of real data, and the mutual information is evaluated solely with respect the continuously varying feature. Averaging out dimensions in the stimulus corresponds to losing accuracy in its description, and hence the information loss.

Our theoretical analysis allows a direct comparison with real curves; we present one comparison and discuss possible causes for the discrepancy between data and model. In particular, correlations between neurons, that are not included in the model, might play a relevant role, enhancing or decreasing redundancy in population coding [11]. This issue will be the object of future work.

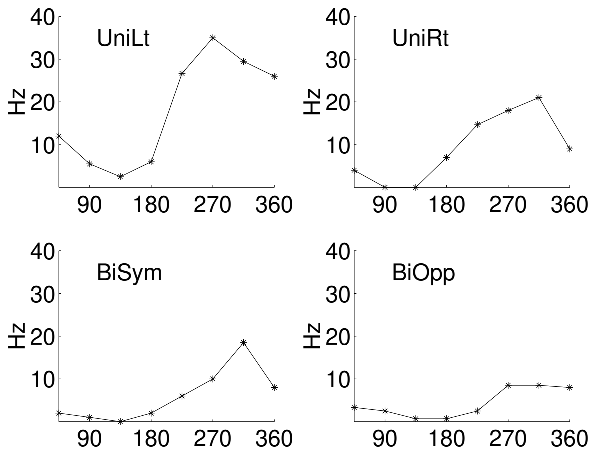

Figure 1: Directional tuning for a cell recorded in the right supplementary motor area of a monkey performing 4 different types of arm movement. UniLt=unimanual left; UniRt=unimanual right; BiSym=bimanual symmetric; BiOpp=bimanual opposite. Notice that the type of movement strongly modulates the amplitude of directional tuning, but not so much its preferred direction.

The gaussian model

First, we consider a coding scheme where the distribution of the firing rates conditional to the movements is gaussian, similarly to the case examined in [2]. This assumption implies that negative rates have a non zero probability to occur, but it allows an easier analytical treatment. We will examine a more realistic scheme later on.

Consider the following distribution:

$$P(\{\eta_i\}|\vartheta, s) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} exp - \left[\left(\eta_i - \tilde{\eta}_i(\vartheta, s)\right)^2 / 2\sigma^2 \right]; \tag{1}$$

η_i is the firing rate of neuron i; ϑ and s parameterize respectively the direction and the type of movement; $\tilde{\eta}_i(\vartheta, s)$ is the average firing rate of the neuron with the movement parameterized by ϑ,s.

In general, the directional tuning of real cells in motor cortices is modulated by the type of movement performed. We show an example of this modulation, with the typical shape of tuning curves, in fig.1 (data kindly provided by Eilon Vaadia). The modulation of the preferred direction looks weaker than the overall amplitude modulation.

For our model we have considered the following function:

$$\tilde{\eta}_i(\vartheta, s) = \varepsilon_s^i \bar{\eta}_i(\vartheta) - (1 - \varepsilon_s^i)\eta_i^f; \tag{2}$$

$$\bar{\eta}_i(\vartheta - \vartheta_i^0) = \eta_i^0 \cos^{2m}\left(\frac{\vartheta - \vartheta_i^0}{2}\right) = \eta_i^0 \left[\frac{1}{2^{2m}} \left(\begin{array}{c} 2m \\ m \end{array} \right) + \frac{1}{2^{2m-1}} \sum_{\nu=0}^{m-1} \left(\begin{array}{c} 2m \\ \nu \end{array} \right) \cos\left\{ (m-\nu)\left(\vartheta - \vartheta_i^0\right) \right\} \right]; \tag{3}$$

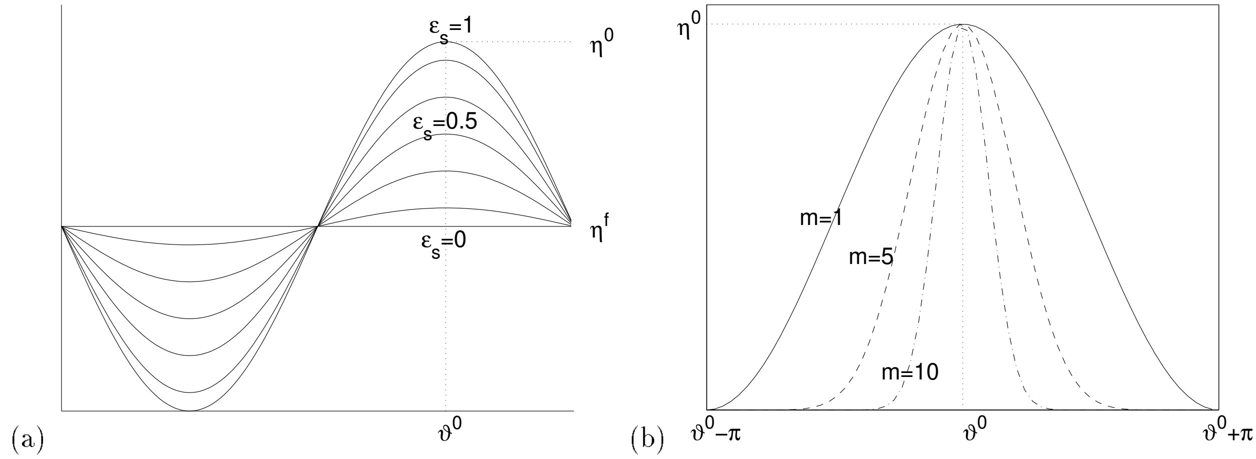

Figure 2: Cosinusoidal tuning curves as in eqs.(2), (3). (a) $m = 1$; modulation for different values of ε_s. (b) $\varepsilon_s = 1$; modulation for different values of m.

ε_s^i is a quenched random variable distributed between 0 and 1. The meaning of eq.(1),(2),(3) is that the firing rate of neuron i given the movement parameterized by (ϑ,s) follows a gaussian distribution centered around a tuning curve $\bar{\eta}_i(\vartheta)$ whose flatness is modulated through the parameters ε_s^i. If ε_s^i is zero for some particular s the firing of the cell does not depend, for that movement type, on the direction of the movement. On the other hand if ε_s^i assumes a fixed value for all s the directional tuning does not modulate with the type of movement. Tuning curves with a cosinusoidal shape have already been considered to model the directional selectivity of sensory neurons [8]. Fig.2 shows the amplitude of the tuning curve.

We neglect the modulation of the preferred direction with the type, as it would burden the analytical calculations; it will be the object of future analyses.

Evaluation of the mutual information for the gaussian model

We are interested in the mutual information [12] between the neuronal firing rates and the movements:

$$I(\{\eta_i\}, \vartheta \otimes s) = \left\langle \sum_{s=1}^{p} \int d\vartheta \int \prod_i d\eta_i P(\vartheta, s) P(\{\eta_i\}|\vartheta, s) \log_2 \frac{P(\{\eta_i\}|\vartheta, s)}{P(\{\eta_i\})} \right\rangle_{\varepsilon, \vartheta^0} ; \tag{4}$$

where the distribution $P(\eta_i|\vartheta, s)$ is given in eq.(1) and $\langle .. \rangle_{\varepsilon,\vartheta^0}$ is a short notation for the average across the quenched variables $\{\varepsilon_s^i\}$,$\{\vartheta_i^0\}$. In fact we are not interested in a particular realization of the tuning, but in the average across all its possible realizations.

Eq.(4) can be written as:

$$I(\{\eta_i\}, \vartheta \otimes s) = H(\{\eta_i\}) - \langle H(\{\eta_i\}|\vartheta, s) \rangle_{\vartheta,s} ; \tag{5}$$

$$\langle H(\{\eta_i\}|\vartheta, s) \rangle_{\vartheta,s} = \left\langle \sum_{s=1}^{p} \int d\vartheta \int \prod_i d\eta_i P(\vartheta, s) P(\{\eta_i\}|\vartheta, s) \log_2 P(\{\eta_i\}|\vartheta, s) \right\rangle_{\varepsilon, \vartheta^0} ; \tag{6}$$

$$H(\{\eta_i\}) = \left\langle \sum_{s=1}^{p} \int d\vartheta \int \prod_i d\eta_i P(\vartheta, s) P(\{\eta_i\}|\vartheta, s) \log_2 \left[\sum_{s'=1}^{p} \int d\vartheta' P(s', \vartheta') P(\{\eta_i\}|\vartheta', s') \right] \right\rangle_{\varepsilon, \vartheta^0} ; \quad (7)$$

The calculation of the equivocation $\langle H(\{\eta_i\}|\vartheta, s)\rangle_{\vartheta,s}$ is straightforward and the result is additive in the population size:

$$\langle H(\{\eta_i\}|\vartheta, s)\rangle_{\vartheta,s} = \frac{N}{2\ln 2}[1 + \ln(2\pi\sigma^2)]; \quad (8)$$

The linearity in N is standard whenever the conditional distribution of firing rates can be factorized across neurons, as in eq.(1).

The evaluation of the rate entropy $H(\{\eta_i\})$ can be carried out introducing n replicas [13, 14] for both the discrete and the continuous dimensions ϑ and s, which allows to get rid of the logarithm in eq.(7):

$$H(\{\eta_i\}) = -\frac{1}{\ln 2} \lim_{n\to 0} \frac{1}{n} \left(\sum_{s_1..s_{n+1}=1}^{p} \int d\vartheta_1..d\vartheta_{n+1} \int \prod_i d\eta_i \frac{1}{(2\pi p)^{n+1}} \right.$$
$$\left. \text{x} \left\langle \prod_{i=1}^{N} \prod_{k=1}^{n+1} \frac{1}{\sqrt{2\pi\sigma^2}} exp - \left[\left(\eta_i - \varepsilon^i_{s_k}\bar{\eta}_i(\vartheta_k) - (1-\varepsilon^i_{s_k})\eta^f_i\right)^2 / 2\sigma^2 \right] \right\rangle_{\varepsilon,\vartheta^0} - 1 \right); \quad (9)$$

Integrating over $\{\eta_i\}$ and rearranging terms one obtains:

$$H(\{\eta_i\}) = -\frac{1}{\ln 2} \lim_{n\to 0} \frac{1}{n} \left(\sum_{s_1..s_{n+1}=1}^{p} \int d\vartheta_1..d\vartheta_{n+1} \frac{(n+1)^{-\frac{N}{2}}}{(2\pi p)^{n+1}} \frac{1}{(\sqrt{2\pi\sigma^2})^{nN}} \left\langle \prod_{i=1}^{N} exp(-R_i) \right\rangle_{\varepsilon,\vartheta^0} - 1 \right); \quad (10)$$

$$R_i = \sum_{k,l} \left(\varepsilon^i_{s_k}(\bar{\eta}_i(\vartheta_k) - \eta^f_i) - \varepsilon^i_{s_l}(\bar{\eta}_i(\vartheta_l) - \eta^f_i) \right)^2 / (4\sigma^2(n+1)); \quad (11)$$

Large σ limit

An exact analytical evaluation of eq.(10) is not possible without resorting to some approximation. In line with the analysis performed in [2] we assume now that the quenched randomness is uncorrelated and identically distributed across neurons:

$$\varrho(\{\varepsilon^i_s\}) = \left(\prod_s \varrho(\varepsilon_s) \right)^N ;$$
$$\varrho(\{\vartheta^0_i\}) = \left[\varrho(\vartheta^0) \right]^N = \frac{1}{(2\pi)^N};$$

We assume also that $\eta^f_i = \eta^f \ \forall i$. Then one can write:

$$\left\langle \prod_i exp(-R_i) \right\rangle_{\varepsilon,\vartheta^0} = \langle exp(-R)\rangle^N_{\varepsilon,\vartheta^0} \;;$$

We consider now the limit of large noise σ; in this case, since $R \simeq 1/\sigma^2$ we can expand $exp(-R)$; keeping only terms of order $(N/\sigma^k)^l$, with $k \leq 2$ and $l = 1, 2$ we obtain:

$$\langle exp(-R)\rangle^N_{\varepsilon,\vartheta^0} \simeq 1 - N\langle R\rangle_{\varepsilon,\vartheta^0} + \frac{N^2}{2}\langle R\rangle^2_{\varepsilon,\vartheta^0} =$$
$$1 - \frac{N}{4\sigma^2(n+1)}\left\langle \sum_{k,l\neq k}\left(\varepsilon_{s_k}(\bar\eta(\vartheta_k)-\eta^f) - \varepsilon_{s_l}(\bar\eta(\vartheta_l)-\eta^f)\right)^2\right\rangle_{\varepsilon,\vartheta^0} + \frac{1}{2}\frac{N^2}{(4\sigma^2(n+1))^2}$$
$$\sum_{k,l\neq k}\sum_{\varrho,\mu\neq\varrho}\left\langle\left(\varepsilon_{s_k}(\bar\eta(\vartheta_k)-\eta^f)-\varepsilon_{s_l}(\bar\eta(\vartheta_l)-\eta^f)\right)^2\right\rangle_{\varepsilon,\vartheta^0}\left\langle\left(\varepsilon_{s_\varrho}(\bar\eta(\vartheta_\varrho)-\eta^f)-\varepsilon_{s_\mu}(\bar\eta(\vartheta_\mu)-\eta^f)\right)^2\right\rangle_{\varepsilon,\vartheta^0} \tag{12}$$

To first order in N/σ^2 we obtain:

$$H(\{\eta_i\}) = -\frac{1}{\ln 2}\lim_{n\to 0}\frac{1}{n}\left(\sum_{s_1..s_{n+1}=1}^{p}\int d\vartheta_1..d\vartheta_{n+1}\frac{(n+1)^{-\frac{N}{2}}}{(2\pi p)^{n+1}}\frac{1}{(\sqrt{2\pi\sigma^2})^{nN}}\right.$$
$$\left.\left(1 - \frac{N}{4\sigma^2(n+1)}\left\langle\sum_{k,l\neq k}\left(\varepsilon_{s_k}(\bar\eta(\vartheta_k)-\eta^f) - \varepsilon_{s_l}(\bar\eta(\vartheta_l)-\eta^f)\right)^2\right\rangle_{\varepsilon,\vartheta^0}\right) - 1\right); \tag{13}$$

This result is valid for a generic directional tuning curve $\bar\eta(\vartheta)$. We consider now our specific choice, eq.(3), examining the simplest case $m = 1$ first. After averaging across direction selectivities $\{\vartheta^0\}$ and integrating over continuous replicas $\vartheta_1..\vartheta_{n+1}$ we obtain

$$H(\{\eta_i\}) = -\frac{1}{\ln 2}\lim_{n\to 0}\frac{1}{n}\left(\sum_{s_1..s_{n+1}=1}^{p}\frac{(n+1)^{-\frac{N}{2}}}{(p)^{n+1}}\frac{1}{(\sqrt{2\pi\sigma^2})^{nN}}\right.$$
$$\left.\text{x}\left(1 - \frac{N(\eta^0)^2}{4\sigma^2(n+1)}\sum_{k,l\neq k}\left[\left(\alpha(\alpha-1)+\frac{1}{4}\right)\left\langle(\varepsilon_{s_k}-\varepsilon_{s_l})^2\right\rangle_\varepsilon + \frac{1}{8}\left\langle\varepsilon^2_{s_k}+\varepsilon^2_{s_l}\right\rangle_\varepsilon\right]\right) - 1\right);$$

where we have defined $\alpha = \eta^f/\eta^0$. To perform the average across the quenched variables $\{\varepsilon_s\}$ we assume that they are equally distributed across p movement types, namely:

$$\varrho(\{\varepsilon_s\}) = \prod_s \varrho(\varepsilon_s) = [\varrho(\varepsilon)]^p;$$

In this case the sum over indexes k and l generates $n(n+1)$ identical terms. The summation on discrete replicas yields a factor p^{n+1} multiplying the term $\left\langle(\varepsilon_{s_k})^2\right\rangle_\varepsilon$ and a factor $p^n(p-1)$ multiplying the term $\left\langle(\varepsilon_{s_k}-\varepsilon_{s_l})^2\right\rangle_\varepsilon$, since this last term is non zero only when $s_k \neq s_l$.

Taking the limit $n \to 0$ yields

$$H(\{\eta_i\}) \simeq \frac{N}{2\ln 2}[1+\ln(2\pi\sigma^2)]$$
$$+\frac{1}{\ln 2}\frac{N(\eta^0)^2}{4\sigma^2}\left[\frac{p-1}{p}2\left(\alpha(\alpha-1)+\frac{1}{4}\right)\lambda_1+\frac{1}{4}\lambda_2\right]; \quad (14)$$

$$\lambda_1 = \int d\varepsilon\varrho(\varepsilon)\varepsilon^2 - \left[\int d\varepsilon\varrho(\varepsilon)\varepsilon\right]^2; \quad (15)$$

$$\lambda_2 = \int d\varepsilon\varrho(\varepsilon)\varepsilon^2; \quad (16)$$

From eqs.(14) and (8) the final expression for the mutual information can be written

$$I(\{\eta_i\},\vartheta\otimes s) \simeq \frac{1}{\ln 2}\frac{N(\eta^0)^2}{4\sigma^2}\left[\frac{p-1}{p}2\left(\alpha(\alpha-1)+\frac{1}{4}\right)\lambda_1+\frac{1}{4}\lambda_2\right]; \quad (17)$$

In the more general case of a power $2m$ of the cosine, in eq.(3), it is easy to show that the final result can be expressed as

$$I(\{\eta_i\},\vartheta\otimes s) \simeq \frac{1}{\ln 2}\frac{N(\eta^0)^2}{4\sigma^2}\left[\frac{p-1}{p}2\left(\alpha-A_1\right)^2\lambda_1+2\left(A_2-(A_1)^2\right)\lambda_2\right]; \quad (18)$$

where:

$$\eta^0 A_1 = \langle\bar\eta(\vartheta_k)\rangle_{\vartheta^0} = \eta^0\frac{1}{2^{2m}}\begin{pmatrix} 2m \\ m \end{pmatrix}; \quad (19)$$

$$\left(\eta^0\right)^2 A_2 = \left\langle\bar\eta^2(\vartheta_k)\right\rangle_{\vartheta^0} = \left(\eta^0\right)^2\frac{1}{2^{4m}}\begin{pmatrix} 4m \\ 2m \end{pmatrix}; \quad (20)$$

The calculation of the coefficient of the second order term, which multiplies N^2/σ^4 in eq.(12), is slightly more complex and implies integration of terms with 4-replica interaction. The detailed analytical evaluation is given in the appendix. The final result up to the quadratic approximation reads:

$$I(\{\eta_i\},\vartheta\otimes s) \simeq \frac{1}{\ln 2}\left\{\frac{N(\eta^0)^2}{4\sigma^2}\left[\frac{p-1}{p}2\left(\alpha-A_1\right)^2\lambda_1+2\left(A_2-(A_1)^2\right)\lambda_2\right]-\frac{N^2(\eta^0)^4}{2(4\sigma^2)^2}\right. \quad (21)$$
$$\left.\left\{\frac{p-1}{p^2}2\left(2\left(\alpha-A_1\right)^2\lambda_1\right)^2+\left[\frac{p-1}{p}\left(\lambda_1-\lambda_2\right)^2+\frac{(\lambda_2)^2}{p}\right]\left[\left(\frac{1}{2^{2m-1}}\right)^4\sum_{\nu=0}^{m-1}\left[\begin{pmatrix} 2m \\ \nu \end{pmatrix}\right]^4\right]\right\}\right\};$$

where the expressions of λ_1,λ_2,A_1,A_2 are given respectively in eq.(15),(16),(19),(20) In the limit of large p we have

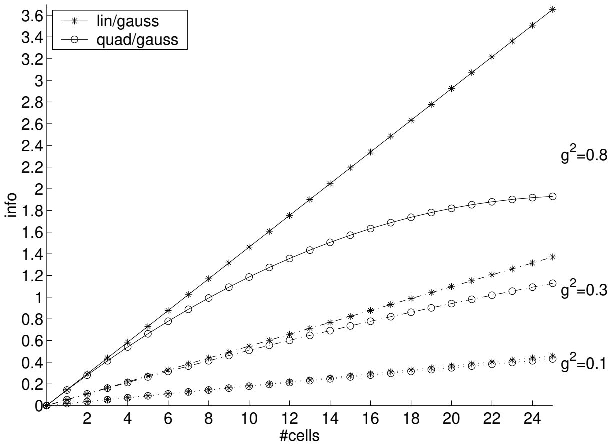

Figure 3: Information rise, from eq.(22), for different values of the expansion parameter $g^2 = (\eta^0/2\sigma)^2$; $m = 1$; $p = 4$; the distribution $\varrho(\varepsilon)$ in eqs.(15), (16) is just equal to 1/3 for each of the 3 allowed ε values of 0,1/2 and 1.

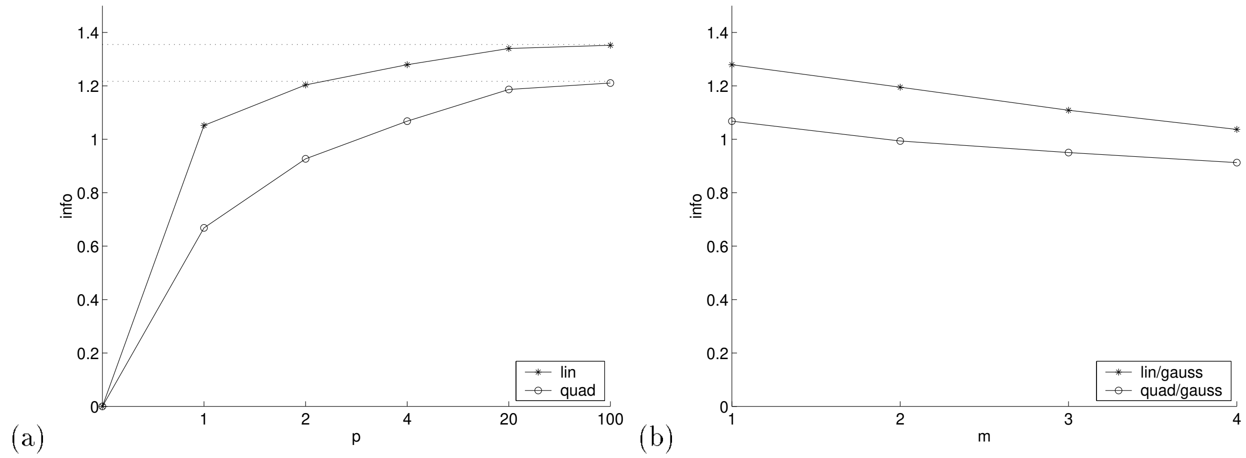

Figure 4: Mutual information in linear and quadratic approximation as in eq.(22), for a sample of 10 cells; $g^2 = 0.7$; the distribution $\varrho(\varepsilon)$ in eqs.(15), (16) is just equal to 1/3 for each of the three allowed a ε values of 0,1/2 and 1. (a)Dependence on the number of movement types p. Dotted lines are for the asymptotes, eq.(22); m=1. (b)Dependence on the power m of the cosine in eq.(3); p=4.

$$I(\{\eta_i\},\vartheta\otimes s)\simeq\frac{1}{\ln 2}\left\{\frac{N(\eta^0)^2}{4\sigma^2}\left[2\left(\alpha-A_1\right)^2\lambda_1+2\left(A_2-(A_1)^2\right)\lambda_2\right]\right.$$
$$\left.-\frac{N^2(\eta^0)^4}{2(4\sigma^2)^2}\left\{(\lambda_1-\lambda_2)^2\left[\left(\frac{1}{2^{2m-1}}\right)^4\sum_{\nu=0}^{m-1}\left[\left(\begin{array}{c}2m\\ \nu\end{array}\right)\right]^4\right]\right\}\right\};\tag{22}$$

Fig.3 shows the linear and the quadratic approximations of eq.(22) for different values of the expansion parameter $g^2=(\eta^0/2\sigma)^2$. It is easy to see that, for very small values of g^2, linear and quadratic approximations roughly coincide, while when $g^2>0.8$ the quadratic approximation begins to fail and one should add higher orders in perturbation theory.

Fig.4(a) shows the dependence of the mutual information on the number of types p. The dependence on p is weak for the linear approximation and somewhat stronger for the quadratic one. In both cases an increase in the number of discrete correlates p produces an increase in the mutual information. The distance between the linear and the quadratic approximation remains asymptotically finite (for $p\to\infty$), contrary to what happens in the case of discrete stimuli alone [2].

Fig.4(b) shows the dependence of the mutual information on the width of the directional tuning (see fig.2,(b)). Since we are considering the case when the noise σ is large, a very narrow tuning in ϑ corresponds to a larger overlap in the conditional probabilities $p(\eta|\vartheta,s)$, for most angles ϑ. A consequence is that, especially in the linear approximation, the mutual information is a (slowly) decreasing function of m.

The limit of large N

We consider now the case when the number of neurons is large. Since we deal with an infinite number of stimuli the mutual information is unbounded. Thus we expect that when the number of neurons becomes large and the noise is finite the mutual information tends asymptotically to infinity.

In order to study this limit we discretize the $\{\vartheta\}$ space into a finite set of $M=2\pi/\Delta\vartheta$ angles $\vartheta_1..\vartheta_M$, and then we take the limit $\Delta\vartheta\to 0$.

The entropy of the responses $H(\{\eta_i\})$ can be written

$$H(\{\eta_i\})=\left\langle\sum_{s=1}^{p}\sum_{k=1}^{M}\int\prod_i d\eta_i P(\vartheta_k,s)P(\{\eta_i\}|\vartheta_k,s)\log_2\left[\sum_{s'=1}^{p}\sum_{k'=1}^{M}P(s',\vartheta_{k'})P(\{\eta_i\}|\vartheta_{k'},s')\right]\right\rangle_{\varepsilon,\vartheta^0}\tag{23}$$

where in analogy with eq.(1) we define

$$P(\{\eta_i\}|\vartheta_k,s)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}exp-\left[\left(\eta_i-\varepsilon_s^i\bar{\eta}_i(\vartheta_k)-(1-\varepsilon_s^i)\eta_i^f\right)^2/2\sigma^2\right];\tag{24}$$

and we discretize the average across the directional selectivities $\{\vartheta_i^0\}$ as well

$$\int d\vartheta^0\varrho(\vartheta^0)\longrightarrow\sum_{k=1}^{M}\varrho(\vartheta_k^0).\tag{25}$$

This situation corresponds to the case when each neuron can discriminate across different angles $\vartheta_1..\vartheta_M$ with a resolution $\Delta\vartheta$.

The calculation can be carried out introducing replicas as in the previous case. One gets

$$H(\{\eta_i\}) = -\frac{1}{\ln 2}\lim_{n\to 0}\frac{1}{n}\left(\sum_{s_1..s_{n+1}=1}^{p}\sum_{k_1...k_{n+1}=1}^{M}\frac{(n+1)^{-\frac{N}{2}}}{(Mp)^{n+1}}\frac{1}{(\sqrt{2\pi\sigma^2})^{nN}}\left\langle exp(-R)^N\right\rangle_{\varepsilon,\vartheta^0}-1\right) \quad (26)$$

where

$$R = \left[\sum_{l,m}\left(\varepsilon_{s_l}(\bar\eta(\vartheta_{k_l})-\eta^f)-\varepsilon_{s_m}(\bar\eta(\vartheta_{k_m})-\eta^f)\right)^2/(4\sigma^2(n+1))\right]; \quad (27)$$

and we have assumed symmetry across neurons in the quenched randomness and in the parameters characterizing the conditional distribution $P(\eta_i|s,\vartheta_k)$.

Now we take the limit $N\to\infty$. As it is evident from eq.(27), $exp(-R)\leq 1$ and $exp(-R)=1$ when $s_l=s_m$ and $k_m=k_l$ for each pair of indexes (m,l). Thus when $N\to\infty$ the only terms which survive in the sum on replicas are the ones with $s_1=s_2..=s_{n+1}$ and $k_1=k_2..=k_{n+1}$. Since we have p stimuli s and M stimuli ϑ_k the total number of terms is Mp. Substituting this value in the sums over replicas in eq.(26) and putting $exp(-R)=1$ one obtains an expression for the entropy of the responses $H(\{\eta_i\})$, which summed to the equivocation as in eq.(8) gives the final result for the mutual information:

$$I(\{\eta_i\},\vartheta\otimes s) = \log_2(p)+\log_2(M); \quad (28)$$

Now we remember that $M=2\pi/\Delta\vartheta$. Taking the limit to continuous angles, $\Delta\vartheta\to 0$, it is easy to see that asymptotically the mutual information tends logarithmically to infinity.

Beyond the gaussian assumption: the *tg* model

So far we have considered the case where the rate distribution for each neuron is normal. This assumption implies that negative rates have a non zero probability to occur; the more the average rate is small and close to zero, the more this probability becomes large. The bias introduced by the inclusion of negative rates in the space of possible states might be even more serious since we have considered the limit of large noise, where the tail of the distribution in the domain $\eta<0$ acquires a significant weight.

Cutting the distribution at zero is not enough to assign the proper weight to under-threshold activity: each time the summation of the inputs coming from other units is lower than threshold the neuron remains silent, and this occurs with a well defined probability.

A natural choice for the rate distribution $P(\eta_i|\vartheta,s)$ is a thresholded gaussian plus a δ peak in zero (*tg* model):

$$P(\eta_i|\vartheta,s) = \frac{1}{\sqrt{2\pi\sigma^2}}exp-\left[(\eta_i-\tilde\eta_i(\vartheta,s))^2/2\sigma^2\right]\Theta(\eta_i)+2(1-\mathrm{erf}(\tilde\eta_i(\vartheta,s)/\sigma)\delta(\eta_i)\Theta(-\eta_i) \quad (29)$$

where $\Theta(x)$ is the Heaviside step function, $\tilde\eta_i(\vartheta,s)$ is the same as defined in eq.(2) and $\mathrm{erf}(x)$ is the error function:

$$\mathrm{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} dt e^{-t^2/2}. \tag{30}$$

The factor multiplying the δ function ensures a correct normalization and it assigns the proper weight to the peak in zero, which is larger the more the average rate $\tilde{\eta}_i(\vartheta, s)$ is close to zero. A similar distribution has already been considered in networks of threshold linear neurons [15, 16].

The analytical evaluation of the mutual information is obviously more difficult than in the case of the simple gaussian, because of the presence of the error function, which cannot be integrated exactly. Nonetheless in the limit of large σ it is possible to evaluate both the linear and the quadratic approximation in N and thus quantify the impact of the correction with respect to the gaussian case, eq.(22).

The limit of large σ for the tg model: the equivocation

We remind the expression of the equivocation, eq.(6):

$$\langle H(\{\eta_i\}|\vartheta, s)\rangle_{\vartheta,s} = \left\langle \sum_{s=1}^{p} \int d\vartheta \int \prod_i d\eta_i P(\vartheta, s) P(\{\eta_i\}|\vartheta, s) \log_2 P(\{\eta_i\}|\vartheta, s) \right\rangle_{\varepsilon,\vartheta^0}. \tag{31}$$

Assuming independence among neurons in the conditional probability $P(\{\eta_i\}|\vartheta, s)$, eq.(31) can be written

$$\langle H(\{\eta_i\}|\vartheta, s)\rangle_{\vartheta,s} = \frac{N}{\ln 2} \sum_{s=1}^{p} \int d\vartheta P(\vartheta, s) \left\langle \int d\eta P(\eta|\vartheta, s) \ln P(\eta|\vartheta, s) \right\rangle_{\varepsilon,\vartheta^0}. \tag{32}$$

In the specific case of the distribution (29) it is easy to show that

$$\begin{aligned} &\left\langle \int d\eta P(\eta|\vartheta, s) \ln P(\eta|\vartheta, s) \right\rangle_{\varepsilon,\vartheta^0} = \\ &\frac{1}{\sqrt{2\pi}} \left\langle \frac{\tilde{\eta}(\vartheta, s)}{2\sigma} e^{-[\tilde{\eta}(\vartheta,s)]^2/2\sigma^2} \right\rangle_{\varepsilon,\vartheta^0} - \frac{1}{2}\Big(1 + \ln(2\pi\sigma^2)\Big) \langle \mathrm{erf}(\tilde{\eta}(\vartheta, s)/\sigma)\rangle_{\varepsilon,\vartheta^0} \\ &+ \left\langle \int_0^{\infty} d\eta \delta(\eta) \left(1 - \mathrm{erf}(\tilde{\eta}(\vartheta, s)/\sigma)\right) \left[\ln \delta(\eta) + \ln 2 + \ln\left(1 - \mathrm{erf}(\tilde{\eta}(\vartheta, s)/\sigma)\right)\right] \right\rangle_{\varepsilon,\vartheta^0}. \end{aligned} \tag{33}$$

To proceed with the calculation we have to be careful with the integration of the delta function. In fact it is easy to show that the integration of the product $\delta(x) \ln \delta(x)$ yields a logarithmic divergence. Since the mutual information must remain finite with a finite number of neurons N, we expect this divergence to cancel exactly with an analogous term in the rate entropy and, in fact, in the next section, we will show that this is the case. For the moment we use the equality

$$\int_{-\infty}^{+\infty} dx \delta(x) F(x) = \lim_{\epsilon \to 0} \int_{-\epsilon/2}^{\epsilon/2} dx \frac{1}{\epsilon} F(x). \tag{34}$$

Assuming as usual that the quenched disorder is identically distributed across neurons and stimuli and that $\tilde{\eta}(\vartheta, s)$ is like in eq.(1) we can write:

$$\langle H(\{\eta_i\}|\vartheta,s)\rangle_{\vartheta,s} = \frac{N}{2\ln 2}\left\{\left(1+\ln(2\pi\sigma^2)\right)\langle \mathrm{erf}(\tilde\eta(\vartheta,s)/\sigma)\rangle_{\varepsilon,\vartheta^0} - \left\langle \frac{\tilde\eta(\vartheta,s)}{\sqrt{2\pi}\sigma} e^{-[\tilde\eta(\vartheta,s)]^2/2\sigma^2}\right\rangle_{\varepsilon,\vartheta^0}\right.$$

$$\left. +2\langle[1-\mathrm{erf}(\tilde\eta(\vartheta,s)/\sigma)]\rangle_{\varepsilon,\vartheta^0}\ln\frac{\epsilon}{2} - 2\langle[1-\mathrm{erf}(\tilde\eta(\vartheta,s)/\sigma)]\ln[1-\mathrm{erf}(\tilde\eta(\vartheta,s)/\sigma)]\rangle_{\varepsilon,\vartheta^0}\right\}. \qquad (35)$$

The average across quenched disorder cannot be performed if we do not resort to some approximation. Since we have already focused on the limit of large σ it is natural to consider an expansion of the error function in eq. (30) for a small value of its argument:

$$\mathrm{erf}(x) \simeq \frac{1}{2} + \frac{1}{\sqrt{2\pi}}x + o(x^2); \qquad (36)$$

Approximating all the error functions in eq.(35) we obtain:

$$\langle H(\{\eta_i\}|\vartheta,s)\rangle_{\vartheta,s} = \frac{N}{2\ln 2}\left\{\frac{1}{2}\left(1+\ln(2\pi\sigma^2)\right)+\ln(\epsilon)\right.$$

$$\left. +\left\langle\frac{\tilde\eta(\vartheta,s)}{\sqrt{2\pi}\sigma}\right\rangle_{\varepsilon,\vartheta^0}\left(2+\ln(2\pi\sigma^2)-2\ln\epsilon\right) - \frac{1}{\pi}\left\langle\frac{\tilde\eta(\vartheta,s)^2}{\sigma^2}\right\rangle_{\varepsilon,\vartheta^0}\right\} + o(1/\sigma^2) \qquad (37)$$

where in line with the approximation used in the case of the simple gaussian distribution we have omitted terms of order N/σ^k with $k>2$.

Evaluation of the mutual information

We reconsider eq.(7). Using replicas and assuming that the quenched randomness is identically distributed across neurons we obtain

$$H(\{\eta_i\}) = -\frac{1}{\ln 2}\lim_{n\to 0}\frac{1}{n}\left(\sum_{s_1..s_{n+1}=1}^{p}\int d\vartheta_1..d\vartheta_{n+1}\frac{1}{(2\pi p)^{n+1}}\left\langle\int d\eta\prod_{k=1}^{n+1}P(\eta|\vartheta_k,s_k)\right\rangle_{\varepsilon,\vartheta^0}^{N} - 1\right), \quad (38)$$

where $P(\eta|\vartheta,s)$ is given in eq.(29). Integrating over $d\eta$ yields

$$\left\langle\int d\eta\prod_{k=1}^{n+1}P(\eta|\vartheta_k,s_k)\right\rangle_{\varepsilon,\vartheta^0}^{N} = \left\langle (n+1)^{-\frac{1}{2}}\frac{1}{(\sqrt{2\pi\sigma^2})^n}exp(-R)\mathrm{erf}\left(\frac{1}{\sqrt{n+1}}\sum_k\tilde\eta(\vartheta_k,s_k)/\sigma\right)\right.$$

$$\left. +\left(\frac{2}{\epsilon}\right)^n\prod_k\left[1-\mathrm{erf}\left(\frac{\tilde\eta(\vartheta_k,s_k)}{\sigma}\right)\right]\right\rangle_{\varepsilon,\vartheta^0}^{N} \qquad (39)$$

where we have used eq.(34) to integrate the δ function in eq.(29), and the expression of R is like in eq.(11). Using the approximation (36) for the error function and considering the expansion for small n

$$a^n \simeq 1 + n\ln a \qquad (40)$$

we obtain

$$H(\{\eta_i\}) = -\frac{1}{\ln 2}\lim_{n\to 0}\frac{1}{n}\left(\sum_{s_1..s_{n+1}=1}^{p}\int d\vartheta_1..d\vartheta_{n+1}\frac{1}{(2\pi p)^{n+1}}\left(1-nC-\sum_{k,l\neq k}G_{kl}\right)^N - 1\right); \quad (41)$$

where

$$C = \frac{1}{2}\left[\frac{1}{2}\left(1+\ln(2\pi\sigma^2)\right)+\ln(\epsilon)+\sum_k\left\langle\frac{\tilde{\eta}(\vartheta_k,s_k)}{\sqrt{2\pi}\sigma}\right\rangle_{\varepsilon,\vartheta^0}\left(2+\ln(2\pi\sigma^2)-2\ln\epsilon\right)\right]; \quad (42)$$

$$G_{kl} = \frac{1}{8(n+1)\sigma^2}\left[\left\langle\left(\tilde{\eta}(\vartheta_k,s_k)-\tilde{\eta}(\vartheta_l,s_l)\right)^2\right\rangle_{\varepsilon,\vartheta^0}-\frac{1}{2\pi}\left\langle\tilde{\eta}(\vartheta_k,s_k)\tilde{\eta}(\vartheta_l,s_l)\right\rangle_{\varepsilon,\vartheta^0}\right]. \quad (43)$$

It is simple to verify that C remains finite when n goes to zero. Now we expand in powers of N up to the second order:

$$\left(1-nC-\sum_{k,l\neq k}G_{kl}\right)^N \simeq 1-N\left(nC+\sum_{k,l\neq k}G_{kl}\right)+\frac{N^2}{2}\sum_{k,l\neq k}\sum_{\varrho,\mu\neq\varrho}G_{kl}G_{\varrho\mu}; \quad (44)$$

where we have omitted terms that are $o(n)$ when $n\to 0$. This quantity has to be summed over continuous and discrete replicas, after having explicitly performed the average across quenched disorder in eqs.(42) and (43). It is easy to show that C cancels exactly with analogous terms in the equivocation, eq.(37). The evaluation of the linear and quadratic term in the mutual information can be performed with a similar technique to the one used in the case of the gaussian distribution and it involves averaging terms with $2-$ and $4-$replica interactions.

The final expression for the mutual information in the linear and quadratic approximation reads

$$\begin{aligned} &I(\{\eta_i\},\vartheta\otimes s)\simeq \\ &\frac{1}{\ln 2}\left\{\frac{N(\eta^0)^2}{4\sigma^2}\left(\frac{1}{2}+\frac{1}{\pi}\right)\left[\frac{p-1}{p}2\left(\alpha-A_1\right)^2\lambda_1+2\left(A_2-(A_1)^2\right)\lambda_2\right]-\frac{N^2(\eta^0)^4}{2(4\sigma^2)^2}\left(\frac{1}{2}+\frac{1}{\pi}\right)^2\right. \quad (45)\\ &\left.\left\{\frac{p-1}{p^2}2\left(2\left(\alpha-A_1\right)^2\lambda_1\right)^2+\left[\frac{p-1}{p}\left(\lambda_1-\lambda_2\right)^2+\frac{(\lambda_2)^2}{p}\right]\left[\left(\frac{1}{2^{2m-1}}\right)^4\sum_{\nu=0}^{m-1}\left[\binom{2m}{\nu}\right]^4\right]\right\}\right\}. \end{aligned}$$

Comparing eq.(22) and (46) it is evident that modifying the gaussian model into the more realistic *tg* model has no effect on the analytical expression of the mutual information, except for a renormalization of the expansion parameter $g^2 = (\eta^0/2\sigma)^2$:

$$g' = g\sqrt{\frac{1}{2}+\frac{1}{\pi}}. \quad (46)$$

Fig.5 shows the effect of the renormalization for different values of g^2. The mutual information is lower in the *tg* model than in the gaussian approximation, as expected.

We have explored whether the two models can fit the information rise estimated from real data. Since the analytical expression of the mutual information is the same in both cases the fit does not change between the two models.

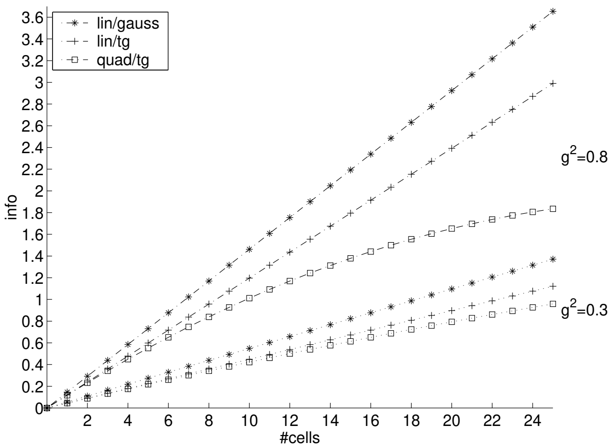

Figure 5: Information rise as in eqs.(22), (46) for different values of the expansion parameter $g^2 = (\eta^0/2\sigma)^2$; $m = 1$; $p = 4$; the distribution $\varrho(\varepsilon)$ in eqs.(15), (16) is just equal to 1/3 for each of the 3 allowed ε values of 0,1/2 and 1.

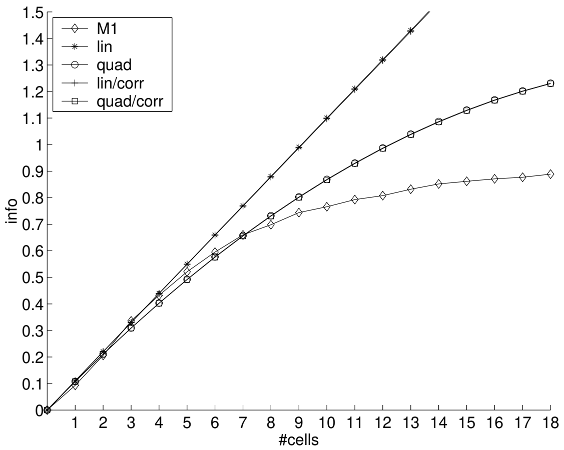

Figure 6: Comparison between the theoretical curves, eq.(22), (46) and the information estimated from a sample of cells recorded in the right primary motor cortex [10]; m=1;p=2; the distribution $\varrho(\varepsilon)$ in eqs.(15), (16) is just equal to 1/3 for each of the three allowed ε of 0,1/2,1; the values of $g^2 = (\eta^0/2\sigma)^2$ used for the fit are 0.64 for eq.(22) and 0.78 for eq.(46)

Fig.6 shows the comparison between the information estimated from a sample of cells recorded in the right primary motor cortex [10] and the prediction given by either of the two theoretical distributions. In the limit of large σ the gaussian model fails to provide a good fit, but we can conclude that this failure is not due to the inclusion of negative rates in the distribution.

Information loss in averaging activity distributions

Fig.1 suggests that the directional tuning of real cells is modulated, albeit moderately, by the type of movement. In fact the analysis of real data has proved that the coding of the direction is not unique, but it is specific to the complex correlate which is being considered [10] (here, which arm moves).

More in general, distinct features characterizing a complex stimulus are not expected to be coded independently of one another. This raises the question of how central representations of external correlates are constructed and which are the basic featural components of these representations. Of course the categorization of natural stimuli is arbitrary and the more accurate a description is provided, the higher the dimensionality of the stimulus set. Since an infinite number of different descriptors could be chosen to characterize a stimulus, any (finite) categorization has the effect of emphasizing some *relevant* features and averaging out other *irrelevant* features. An obvious consequence is that we end up, even involuntarily, evaluating how some features are coded *on average*, with respect to the dimension we have chosen, explicitly or implicitly, to neglect.

Thus, with correlates which have one continuous and one discrete dimension, one might wonder which are the relationships among the information carried about the total number of continuous+discrete dimensions, the information carried about the continuous dimension, disregarding the discrete dimension and, finally, the information carried about the continuous dimension, if a single value of the discrete dimension had been fixed when recording neural activity. In other words, suppose that we investigate how the direction is coded *on average* across different types of movement. This corresponds to averaging the full distribution $P(\{\eta_i\}|\vartheta, s)$ on s:

$$P(\{\eta_i\}|\vartheta, s) \to \sum_s P(s)P(\{\eta_i\}|\vartheta, s) = P_s(\{\eta_i\}|\vartheta). \tag{47}$$

The resulting expression of the mutual information is:

$$I(\{\eta_i\}, \langle\vartheta\rangle_s) = \left\langle \int d\vartheta \int \prod_i d\eta_i P(\vartheta) P_s(\{\eta_i\}|\vartheta) \log_2 \frac{P_s(\{\eta_i\}|\vartheta)}{P(\{\eta_i\})} \right\rangle_{\varepsilon,\vartheta^0}. \tag{48}$$

The analytical evaluation is very similar to the cases already discussed.

As usual, the mutual information can be expressed as the difference between the entropy of the responses $H(\{\eta_i\})$ and the equivocation $\langle H(\{\eta_i\}|\vartheta)\rangle_\vartheta$, where

$$\langle H(\{\eta_i\}|\vartheta)\rangle_\vartheta = \left\langle \int d\vartheta \int \prod_i d\eta_i P(\vartheta) P_s(\{\eta_i\}|\vartheta) \log_2 P_s(\{\eta_i\}|\vartheta) \right\rangle_{\varepsilon,\vartheta^0}; \tag{49}$$

$$H(\{\eta_i\}) = \left\langle \int d\vartheta \int \prod_i d\eta_i P(\vartheta) P_s(\{\eta_i\}|\vartheta) \log_2 \left[\int d\vartheta' P(\vartheta') P_s(\{\eta_i\}|\vartheta') \right] \right\rangle_{\varepsilon,\vartheta^0}. \tag{50}$$

We focus on the more realistic *tg* model, eq.(29); the entropy of the responses is obviously independent on the chosen categorization of the stimuli; in the limit of large σ the procedure is precisely the one followed in the previous section.

The difference with respect to the cases already discussed is in the evaluation of the equivocation: since $P_s(\{\eta_i\}|\vartheta)$ is obtained averaging the distribution (29) across s, we need to introduce discrete replicas $s_1..s_n$, as when evaluating the entropy of the responses. Then the calculation is straightforward and the basic steps are given in the previous section and in the appendix.

Since in one case (for the entropy of the responses) we sum both over discrete and over continuous replicas and in the other case (for the equivocation) we sum only on discrete replicas, it is clear that all terms that do not involve 2 or more replica interacting cancel out.

More in detail, if the evaluation of the entropy of the responses requires the analytical calculation of averages like $\langle\bar\eta(\vartheta_k)\bar\eta(\vartheta_l)\rangle_{\vartheta^0}$ (see the appendix), these averages disappear in the evaluation of the equivocation, since replica indexes are only for the discrete variable s.

The final result for the mutual information up to the quadratic approximation reads:

$$
\begin{aligned}
I(\{\eta_i\},\langle\vartheta\rangle_s) \simeq \frac{1}{\ln 2}&\left\{\frac{N(\eta^0)^2}{4\sigma^2}\left(1+\frac{2}{\pi}\right)\left(A_2-(A_1)^2\right)\left[\lambda_2-\frac{p-1}{p}\lambda_1\right]\right.\\
-\frac{N^2(\eta^0)^4}{2(4\sigma^2)^2}&\left(\frac{1}{2}+\frac{1}{\pi}\right)^2\left\{\frac{p-1}{p^2}8\left[(\alpha-A_1)^4-\left(A_2-2\alpha A_1+\alpha^2\right)^2\right](\lambda_1)^2\right.\\
+&\left.\left.\left[\frac{p-1}{p}(\lambda_1-\lambda_2)^2+\frac{(\lambda_2)^2}{p}\right]\left[\left(\frac{1}{2^{2m-1}}\right)^4\sum_{\nu=0}^{m-1}\left[\binom{2m}{\nu}\right]^4\right]\right\}\right\}.
\end{aligned}
\tag{51}
$$

Fig.7(a) compares the *averaged* information, $I(\{\eta_i\},\langle\vartheta\rangle_s)$, with the *full* information $I(\{\eta_i\},\vartheta\otimes s)$, where in the case of $I(\{\eta_i\},\vartheta\otimes s)$ we have put $p=1$. The full information calculated with $p=1$ in fact gives the curve one would obtain, on average, by considering only one movement type (or value of the discrete correlate) at a time. As one could expect, averaging the distribution across the discrete correlates results in an information loss. Moreover, the the *full* information with $p=4$ movement types is obviously above the *specific* one, obtained by setting $p=1$ (compare fig.5 and 7(a)).

Fig.7(b) shows the dependence of the full and averaged information on the number p of discrete correlates. Contrary to the *full* information, the *averaged* information decreases monotonically with p, both in the linear and in quadratic approximation.

As one would expect, averaging the distribution across a large number p of correlates is equivalent to a regularization of the activity distribution, which results in a lower mutual information.

Discussion

We have studied a model of the coding of discrete+continuous stimuli by a population of N neurons, referring to the specific case of movements categorized according to a direction and a type. We have shown that, asymptotically in the limit of large populations of neurons, the mutual information tends to infinity logarithmically with the resolution in the continuous dimension. This result aside, we have focused on the initial rise of the mutual information with the population size, which may offer a more direct comparison with the analysis of real data.

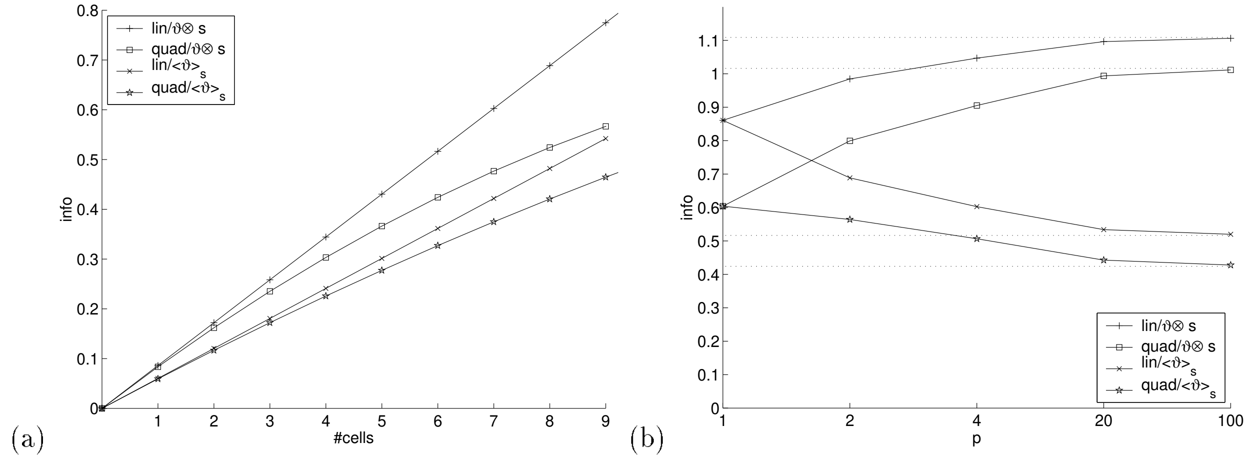

Figure 7: Mutual information in the linear and quadratic approximations, as in eqs.(46), (51), for a sample of 10 cells; $g^2 = 0.7$; $m = 1$; the distribution $\varrho(\varepsilon)$ in eqs.(15), (16) is just equal to 1/3 for each of the 3 allowed ε values of 0,1/2 and 1. (a) Full curve as a function of the population size; $p = 1$ for the *full* information $I(\{\eta_i\}, \vartheta \otimes s)$ and $p = 4$ for the *averaged* information $I(\{\eta_i\}, \langle\vartheta\rangle_s)$. (b) Dependence on the number of movement types p. Dotted lines are for the asymptotes, $p \to \infty$

In the limit of large noise we have therefore derived an analytical formula for such initial rise of the information, up to the quadratic approximation. We have examined the dependence on the number of discrete correlates and on the width of the directional tuning. A comparison with the information estimated from real data, in which the linear term is used as a fit coefficient, has shown that the quadratic approximation fails to capture the deviation from linearity.

We have then considered a more realistic model for the conditional firing distribution, a thresholded gaussian with a δ-peak. We have shown that this more realistic distribution simply renormalizes the expansion parameter applicable to the gaussian distribution. Therefore, the discrepancy in the fit to real data does not originate in the firing rate distribution. There are several possible reasons for this discrepancy:

- the value of the expansion parameter $g^2 = (\eta^0/\sigma)^2$ corresponding to the best fit in fig.(6) is quite high ($g^2 = 0.78$). This value is in the range where the quadratic approximation is expected to fail on its own (see fig.(3),(5)). Adding higher orders in perturbation theory might improve the fit. Moreover, we have neglected terms of order N/σ^k with $k > 2$, which might be non negligible when g^2 becomes large and N is not too small.

- information estimates from real data are often biased because of poor sampling. Several procedures have been proposed to correct the bias [17], but the improvement given by the correction is not precisely quantifiable, and sampling biases cannot be discarded for good.

- both with the *tg* model and in the gaussian approximation, we have assumed that neurons fire independently of one another to each movement, but the analysis of extracellular recordings has shown that correlations may play a non-negligible role in the coding [11, 18].

Finally, we have examined the effect of averaging the distribution across the discrete correlates, evaluating the mutual information with respect to the continuously varying dimension alone. As expected, averaging the distribution across s results in an information loss, which is more serious the larger the number p of discrete correlates.

Further developments of this work include: the introduction of weak correlations in the signal of different neurons and the analysis of the information transfer to other stages of processing, given this as the coding scheme in the input layer of the network. In the specific case of movement coding these research directions might help model how information is transmitted down the motor system, in the planning and execution of motor tasks.

A Detailed evaluation of the second order coefficient

We show here how to evaluate the coefficient of the quadratic term in eq.(12), that we write again:

$$\sum_{k,l\neq k}\sum_{\varrho,\mu\neq\varrho}\left\langle\left(\varepsilon_{s_k}(\bar\eta(\vartheta_k)-\eta^f)-\varepsilon_{s_l}(\bar\eta(\vartheta_l)-\eta^f)\right)^2\right\rangle_{\varepsilon,\vartheta^0}\left\langle\left(\varepsilon_{s_\varrho}(\bar\eta(\vartheta_\varrho)-\eta^f)-\varepsilon_{s_\mu}(\bar\eta(\vartheta_\mu)-\eta^f)\right)^2\right\rangle_{\varepsilon,\vartheta^0}. \tag{52}$$

This quantity has to be integrated over continuous and discrete replicas. First we perform the average across the quenched variable ϑ_0. Expanding the products it is easy to see that the quantities to be averaged are all like $\langle\bar\eta(\vartheta_k)\rangle_{\vartheta^0}$ and $\langle\bar\eta^2(\vartheta_k)\rangle_{\vartheta^0}$, that we have already calculated in eq.(19),(20). Another average that we need is $\langle\bar\eta(\vartheta_k)\bar\eta(\vartheta_l)\rangle_{\vartheta^0}$, with $k\neq l$:

$$\langle\bar\eta(\vartheta_k)\bar\eta(\vartheta_l)\rangle_{\vartheta^0}=\left(\eta^0\right)^2\left\{\left(\frac{1}{2^{2m}}\right)^2\left[\left(\begin{array}{c}2m\\ m\end{array}\right)\right]^2+\frac{1}{2}\left(\frac{1}{2^{2m-1}}\right)^2\sum_{\nu=0}^{m-1}\left[\left(\begin{array}{c}2m\\ \nu\end{array}\right)\right]^2\cos\left\{(m-\nu)\left(\vartheta_k-\vartheta_l\right)\right\}\right\}. \tag{53}$$

Thus only terms like $\langle\bar\eta(\vartheta_k)\bar\eta(\vartheta_l)\rangle_{\vartheta^0}$ still depend on continuous replicas.

Since terms like $\cos(m-\nu)(\vartheta_k-\vartheta_l)$, with $k\neq l$ are zero when integrated on ϑ_k,ϑ_l the only term which requires a careful evaluation in performing the integration on continuous replicas is the product $\langle\bar\eta(\vartheta_k)\bar\eta(\vartheta_l)\rangle_{\vartheta^0}\langle\bar\eta(\vartheta_\varrho)\bar\eta(\vartheta_\nu)\rangle_{\vartheta^0}$.

After integration on continuous replicas this term yields

$$\left(\eta^0\right)^4\left\{\left(\frac{1}{2^{2m}}\right)^4\left[\left(\begin{array}{c}2m\\ m\end{array}\right)\right]^4+\frac{1}{4}\left(\frac{1}{2^{2m-1}}\right)^4\sum_{\nu=0}^{m-1}\left[\left(\begin{array}{c}2m\\ \nu\end{array}\right)\right]^4\left(\delta_{k\varrho}\delta_{\mu l}+\delta_{k\mu}\delta_{\varrho l}\right)\right\}.$$

Rearranging all terms one has, finally

$$\begin{aligned}&\sum_{k,l\neq k}\sum_{\varrho,\mu\neq\varrho}\left[(\alpha-A_1)^2\left\langle(\varepsilon_{s_k}-\varepsilon_{s_l})^2\right\rangle_\varepsilon+\left(A_2-(A_1)^2\right)\left\langle\varepsilon_{s_k}^2+\varepsilon_{s_l}^2\right\rangle_\varepsilon\right]x\\ &x\left[(\alpha-A_1)^2\left\langle(\varepsilon_{s_\varrho}-\varepsilon_{s_\mu})^2\right\rangle_\varepsilon+\left(A_2-(A_1)^2\right)\left\langle\varepsilon_{s_\varrho}^2+\varepsilon_{s_\mu}^2\right\rangle_\varepsilon\right]\\ &+\langle\varepsilon_{s_k}\varepsilon_{s_l}\rangle_\varepsilon\langle\varepsilon_{s_\varrho}\varepsilon_{s_\mu}\rangle_\varepsilon\left(\frac{1}{2^{2m-1}}\right)^4\sum_{\nu=0}^{m-1}\left[\left(\begin{array}{c}2m\\ \nu\end{array}\right)\right]^4\left(\delta_{k\varrho}\delta_{\mu l}+\delta_{k\mu}\delta_{\varrho l}\right)\end{aligned} \tag{54}$$

where A_1 and A_2 are defined in eq.(19) and (20). To correctly perform the summation across the discrete replicas and to keep only terms of order n in the limit $n\to 0$, we consider separately the different contributions to the sum:

$$\sum_{k,l\neq k}\sum_{\varrho,\mu\neq\varrho} \Rightarrow \sum_{k\neq l\neq\varrho\neq\mu} + \sum_{\varrho,k\neq l\neq\mu} (\delta_{\varrho k}+\delta_{\varrho l}) + \sum_{\mu,k\neq l\neq\varrho} (\delta_{\mu k}+\delta_{\mu l}) + \sum_{\mu,\varrho,k\neq l} (\delta_{\mu k}\delta_{\varrho l}+\delta_{\varrho k}\delta_{\mu l})\,. \tag{55}$$

Since in eq.(54) terms like $\left\langle(\varepsilon_{s_k}-\varepsilon_{s_l})^2\right\rangle_\varepsilon$ are zero if $s_k = s_l$, we have to distinguish between different cases for each term in eq.(55) in performing the summation on discrete replicas:

$$\begin{aligned}\sum_{s_1}..\sum_{s_{n+1}} \Rightarrow & \sum_{s_1}..\sum_{s_k\neq s_l}\sum_{s_\varrho\neq s_m}..\sum_{s_{n+1}} + \sum_{s_1}..\sum_{s_k=s_l}\sum_{s_\varrho\neq s_m}..\sum_{s_{n+1}} \\ & + \sum_{s_1}..\sum_{s_k\neq s_l}\sum_{s_\varrho=s_m}..\sum_{s_{n+1}} + \sum_{s_1}..\sum_{s_k=s_l}\sum_{s_\varrho=s_m}..\sum_{s_{n+1}}\,;\end{aligned} \tag{56}$$

With a bit of combinatorics and keeping only the terms order n, which give a finite contribution in the limit $n \to 0$, we obtain the final result for the second order term

$$\frac{N^2(\eta^0)^4}{2(4\sigma^2)^2}\left\{\frac{p-1}{p^2}2\left(2\left(\alpha-A_1\right)^2\lambda_1\right)^2+\left[\frac{p-1}{p}\left(\lambda_1-\lambda_2\right)^2+\frac{\lambda_2^2}{p}\right]\left[\left(\frac{1}{2^{2m-1}}\right)^4\sum_{\nu=0}^{m-1}\left[\left(\begin{array}{c}2m\\ \nu\end{array}\right)\right]^4\right]\right\};$$

where the expressions of λ_1,λ_2 are given in eqs.(15), (16). Considering the result obtained at the first order, eq.(18), it is easy to derive the expression for the mutual information up to the second order in N/σ^2:

$$\begin{aligned}I(\{\eta_i\},\vartheta\otimes s) \simeq \frac{1}{\ln 2}&\left\{\frac{N(\eta^0)^2}{4\sigma^2}\left[\frac{p-1}{p}2\left(\alpha-A_1\right)^2\lambda_1+2\left(A_2-(A_1)^2\right)\lambda_2\right]\right.\\ &\left.-\frac{N^2(\eta^0)^4}{2(4\sigma^2)^2}\left\{\frac{p-1}{p^2}2\left(2\left(\alpha-A_1\right)^2\lambda_1\right)^2+\left[\frac{p-1}{p}\left(\lambda_1-\lambda_2\right)^2+\frac{\lambda_2^2}{p}\right]\left[\left(\frac{1}{2^{2m-1}}\right)^4\sum_{\nu=0}^{m-1}\left[\left(\begin{array}{c}2m\\ \nu\end{array}\right)\right]^4\right]\right\}\right\}.\end{aligned}$$

Acknowledgements

We have enjoyed extensive discussions with Inés Samengo, Eilon Vaadia, Edmund Rolls and Elka Korutcheva. Partial support from Human Frontier Science Programme grant RG 0110/1998-B.

References

[1] W. Bialek, F. Rieke, R. R. de Ruyter van Steveninck, and D. Warland. Reading a neural code. *Science*, 252:1854–1857, 1991.

[2] Inés Samengo and Alessandro Treves. Representational capacity of a set of independent neurons. *Physical Review*, E 63:art. 011910, 2001.

[3] Alessandro Treves. Information coding in higher sensory and memory areas. In F. Moss and S. Gielen, editors, *Neuro-informatics and Neural Modelling*, volume 4 of *Handbook of Biological Physics*, pages 803–829, Amsterdam, 2000. Elsevier.

[4] L. M. Optican and B. J. Richmond. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex:iii.information theoretic analysis. *J. Neurophysiol.*, 57:162–178, 1987.

[5] L. M. Optican and B. J. Richmond. Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex:ii.information transmission. *J. Neurophysiol.*, 64:351–369, 1990.

[6] Edmund T Rolls, Hugo D Critchley, and Alessandro Treves. Representation of olfactory information in the primate orbitofrontal cortex. *J. Neurophysiol.*, 75:1982–1996, 1996.

[7] E. T. Rolls, A. Treves, and M. J. Tovée. The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. *Exp. Brain Res.*, 114:149–162, 1997a.

[8] H. S. Seung and H. Sompolinsky. Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences of the USA*, 90:10749–10753, 1993.

[9] K. Kang and H. Sompolinsky. The mutual information of population codes and distance measures in probability space. *cond-mat/0101161*, 2001.

[10] V. Del Prete, O. Steinberg, A. Treves, and E. Vaadia. *in preparation*, 2000.

[11] S. Panzeri, S.R. Schultz, A. Treves, and E.T.Rolls. Correlations and the encoding of information in the nervous system. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 266:1001–1012, 1999.

[12] C.E. Shannon. A mathematical theory of communication. *AT&T Bell laboratories Technical journal*, 27:379–423, 1948.

[13] M. Mezard, G. Parisi, and M. Virasoro. *Spin glass theory and beyond.* World Scientific Publishing Co Pte Ltd., Singapore, 1987.

[14] A. Treves. Quantitative estimate of the information relayed by the schaffer collaterals. *J. Comp. Neurosci.*, 2:259–272, 1995.

[15] A. Treves. Threshold linear formal neurons in auto-associative nets. *J. Phys.A: Math. Gen.*, 23:2631–2650, 1990.

[16] A. Treves. Graded-response neurons and information encodings in autoassociative memories. *Physical Review A*, 742:2418–2430, 1990.

[17] S. Panzeri and A. Treves. Analytical estimates of limited sampling biases in different information measures. *Network*, 7:87–107, 1996b.

[18] S. Panzeri, S.R. Schultz, A. Treves, and E.T.Rolls. On decoding the responses of a population of neurons from short time windows. *Neural Comp.*, 11:1553–1577, 1999.