The Metric Content of Spatial Views as Represented in the Primate Hippocampus

Alessandro Treves[1], Pierre Georges-Francois[2], Stefano Panzeri[2], Robert G Robertson[2] and Edmund T Rolls[2]

[1]SISSA - Programme in Neuroscience, 34014 Trieste, Italy
[2]Dept. of Exp. Psychology, Univ. of Oxford, Oxford OX1 3UD, UK

Abstract. Coexisting memory representations of the same information may differ in the amount of structure they embody, i.e. in the metric of relationships among individual memory items. Such an amount of structure may be quantified by the metric content index. We extract the metric content of the representation of spatial views in the monkey hippocampus and parahippocampal cortical areas, and find indications of quantitative differences that might be associated with the connectivity pattern in different neural substrates.

Keywords. spatial views, hippocampus, information theory, decoding, ultrametricity, metric structure

1 Quantifying the amount of structure

The structure of neural representations of the outside world has been studied in detail in some simple situations. Typically these are situations in which a well defined correlate of neuronal activity (i.e. a stimulus, a response, or even a behavioural state) is characterized by one or a few parameters that are made to vary continuously or in steps. Examples are the Hubel and Wiesel [2] description of orientation selectivity in cat visual cortex, the O'Keefe [5] finding of place cells in the rat hippocampus, the mitral cell coding of n-aliphatic acid hydrocarbon length in the olfactory system [12], the coding of the direction of movement in 3D-space in the primate motor cortex [1].

In many interesting situations, though, especially in those parts of the brain which are more remote from the periphery, external correlates, or, as we shall refer to them for simplicity, stimuli, do not vary (either continuously, or in steps) along any obvious physical dimension. Often, in experiments, the set of stimuli used is just a small ensemble of a few disparate individual items, arbitrarily selected and difficult to classify systematically. Examples for the ventral visual system are faces [9], simple or complex [4] abstract patterns, or the schematic objects reached with the reduction procedure of Tanaka *et al* [13]. In such situations, the resulting patterns of neuronal activity across populations of cells can still provide useful insight on the structure

of neuronal representations of the outside world, but such insight has to be derived independently of any explicit correlation with a natural, physical structure of the stimulus set.

The only obvious *a priori* metric of the stimulus set, in the general case, is the trivial categorical metric of each element s being equal to itself, and different from any other element in the set. *A posteriori*, the neuronal firing patterns embed the stimulus set into a potentially metric structure defined by the similarities and differences among the patterns corresponding to the various elements. A truly metric structure can be extracted by quantifying such similarities and differences into a notion of distance (among firing patterns) that satisfies the 3 required relations: positivity, symmetry, the triangle inequality. At a more basic level, though, the overall amount of structure, i.e. the overall importance of relations of similarity and difference among firing patterns, can be quantified even independently of any notion of distance, just from a matrix $Q(s|s')$ characterizing the confusability of s' with s, a matrix which need not be symmetrical. It is moreover important to notice that such a matrix $Q(s|s')$ can indeed be derived, as discussed below, from the firing patterns corresponding to each stimulus s, but it can also be derived from other, e.g. behavioural, measures. Behavioural measures of the confusability of s' with s do not access the representation of the two stimuli directly, but indirectly they reflect the multiplicity of neural representations that are important in generating that particular behaviour. If some of these representations are damaged or lost, as in brain-damaged patients, the resulting behavioural measures can be indicative of the structure of the surviving representations [3].

The amount of structure can be quantified by comparing the mutual information in the matrix $Q(s|s')$,

$$I = \sum_{s,s' \in \mathcal{S}} Q(s|s')P(s') \log_2 \frac{Q(s|s')}{\sum_{s"} Q(s|s")P(s")} \tag{1}$$

with its minimum and maximum values I_{min} and I_{max} [14] corresponding to a given percent correct $f_{cor} = \sum_s Q(s|s)P(s)$. The lowest information values compatible with a given f_{cor} are those attained when equal probabilities (or equal frequencies of confusion) result for all wrong stimuli. In this case one finds

$$I_{min} = \log_2 S + f_{cor} \log_2 f_{cor} + (1 - f_{cor}) \log_2[(1 - f_{cor})/(S - 1)]. \tag{2}$$

Conversely, maximum information for a given f_{cor} is contained in the confusion matrix when stimuli are confused only within classes of size $1/f_{cor}$ (for analytical simplicity we assume that each class may contain a non integer number of elements), and the individual stimuli within the class are allocated on a purely random basis. It is easy to see that then

$$I_{max} = \log_2 S + \log_2 f_{cor}. \tag{3}$$

Interpreting the probability of confusion as a monotonically decreasing function of some underlying distance (e.g. as discussed above), the first situation can be taken to correspond to the limit in which the stimuli form an equilateral simplex, or equivalently the stimulus set is drawn from a space of extremely high dimensionality. In the Euclidean $d \to \infty$ limit, points drawn at random from a finite e.g. hyperspherical region tend to be all at the same distance from each other, and from the point of view of the metric of the set this is the *trivial* limit mentioned above. The second situation can be taken to correspond to the *ultrametric* limit, instead, in which all stimuli at distance less than a critical value from each other form clusters such that all distances between members of different classes are above the critical value. This is a non-Euclidean structure (although it could be embedded in a Euclidean space of sufficiently large dimension), and it is a first example of the possible emergence of non-Euclidean aspects from a quantitative analysis that does not rely on *a priori* assumptions.

Intermediate situations between the two extremes are easy to imagine and can be parametrized in a number of different ways. A convenient parameter that simply quantifies the relative amount of information in excess of the minimum, without having to assume any specific parametrization for the confusion matrix, is

$$\lambda = \frac{I - I_{min}}{I_{max} - I_{min}} \tag{4}$$

which ranges from 0 to 1 and can be interpreted as measuring the *metric content* of the matrix. What is quantified by λ can be called the metric content not in the sense that it requires the introduction of a real metric, but simply because it gives the degree to which relationships of being close or different (distant), among stimuli, emerge in the $Q(s|s')$ matrix. For $\lambda = 0$ such relationships are irrelevant, to the point that if confusion occurs, it can be with any (wrong) stimulus. For $\lambda = 1$ close stimuli are so similar as to be fully confused with the correct one, whereas other stimuli are 'maximally distant' and never mistaken for it.

In summary, the metric content index λ quantifies the dispersion in the distribution of errors, from maximal, $\lambda = 0$, to minimal, $\lambda = 1$. The errors may be actual behavioural errors in identifying or categorizing stimuli or in producing appropriate responses, or errors which a neuronal population appears to be making, as an outside observer infers by reading out the spiking activity of the population. We now turn to what is exactly implied by the notion of 'reading out'.

2 Decoding the responses of spatial view cells

Decoding the spike trains emitted by a population of neurons, when a stimulus s from a given set is presented, means applying an algorithm that estimates, given the current spike trains $\vec{r}_s$ and those previously recorded in

response to each stimulus, the likelihoods for each (s') of the possible stimuli to be the current one, $L(s'|\vec{r}_s)$. The stimulus $s' = s_p$ for which this likelihood is maximal can be said to be the stimulus *p*redicted on the basis of the response. In general s_p will not coincide with the true s and the accuracy in the decoding can of course be measured by the percent correct decoding (or the corresponding fraction f_{cor}), but also by the mutual information in the joint probability table $Q(s, s_p)$,

$$I = \sum_{s, s_p \in \mathcal{S}} Q(s, s_p) \log_2 \frac{Q(s, s_p)}{P(s)Q(s_p)}. \tag{5}$$

This is the quantity referred to above. It measures the information in the predictions based on maximum likelihood, and as such it does not only reflect, like percent correct, the number of times the decoding is exact, but also, beyond percent correct, the distribution of wrong decodings. A further quantity which it is sometimes useful to consider is the mutual information

$$I_p = \sum_{s, s' \in \mathcal{S}} P(s, s') \log_2 \frac{P(s, s')}{P(s)P(s')} \tag{6}$$

obtained from the *p*robability $P(s'|s)$ of confusing s with s', which is given by averaging $L(s'|\vec{r}_s)$ over the responses to s. This second information measure reflects, unlike the first, also the degree of certainty with which each single trial has been decoded, and it thus sheds light on a further aspect of the quality attained in decoding. Both information quantities suffer from limited sampling distortions [15, 6] but the second much less than the first, in the sense that, with the limited sampling correction procedures we have developed, I_p can be estimated accurately even with few trials per stimulus, while I requires more trials. Hovever in practice, especially when extracting these measures from limited periods of firing of cortical cells, I is a much better estimate of the actual information contained in the firing (i.e., before decoding) than I_p [7], and because of this fidelity it is preferable to rely on measures of I whenever limited sampling distortions are not the main concern. We note that the metric content index appropriate to I_p would be derived in the same terms, by only replacing f_{cor} with the analogous quantity based on probabilities, $g_{cor} = \sum_s P(s|s)P(s)$.

Decoding algorithms can be optimised to extract as much information as possible, or they can be modelled on the decoding likely to be implemented by real neurons downstream of the recorded populations. Information and percent correct values in the decoding of face cells responses from the primate temporal visual cortex have been reported [10]. There we show that simple, neuronally plausible decoding algorithms, based on dot product operations, perform virtually like optimal decoding algorithms in terms of I, and are only 20-30% inferior in terms of I_p. This is because the simple dot product algorithms are poorer at quantifying likelihoods, even if they order them

correctly and identify correctly the most likely stimulus that can be predicted for each trial.

f_{cor}, I and I_p all depend on the number of cells in the population, as recording the responses of more cells obviously allows better decoding. We have reported the important result [10] that the information decoded from face cells appears to grow linearly with the number of cells in the population, until it begins to saturate at the maximum allowed, which is just the entropy of the stimulus set, $H = -\sum_s P(s) \log_2 P(s)$. This result implies that the different cells in the sample tend to code for different aspects of the stimulus set, so that each contributes an additive term to the information provided by the population. This result appears to hold for the data recorded in a number of experiments, including both the primate inferior temporal cortex face cells [10] mentioned above and the primate hippocampal spatial view cells [11] considered in this report, but also primate orbitofrontal cells coding for odours (Rolls, Treves and Critchley, in preparation), rat hippocampal cells coding for spatial position (e.g. [16]; and also [17]), and rat somatosensory cells coding for whisker deflection [8].

The issue we want to focus on here is not, however, how the accuracy in the decoding depends on the number of cells in the population, but rather how it provides insight on the structure of the stimulus set as encoded in the firing of different populations of cells, and as quantified by the λ metric content index.

3 The metric content in neighbouring areas

The data we consider are the responses of spatial view cells in the primate hippocampus, described by Rolls et al [11], to which we refer for all the details of the experiment and analysis. Briefly, single cells were successively recorded in 2 monkeys while the animals were free to locomote in the lab, and their gaze direction was simultaneously recorded with magnetic coils. For the purpose of the analysis, gaze directions were discretized into 16 'views', which corresponded to an equal number of portions of the lab walls. 20 of the cells used here were recorded in one anymal (monkey *av*) and 6 in another (monkey *az*). Pseudosimultaneous response vectors were constructed by randomly pairing equal numbers of trials in which each cell included in the vectors was recorded in its response to any given view.

Each trial consisted of a 100ms long stretch during which the monkey's gaze was fixed within one of the 16 preset spatial views. At the end of each trial, if the gaze remained fixed for another 100ms period, another trial associated with the same spatial view was constructed, and so on. Decoding was thus based on the number of spikes emitted by each cell considered in the current sample, within one of the 100ms pseudosimultaneous trials. Typically about 80 trials were available for each view.

Note the difference between our response vectors and Georgopoulos' pop-

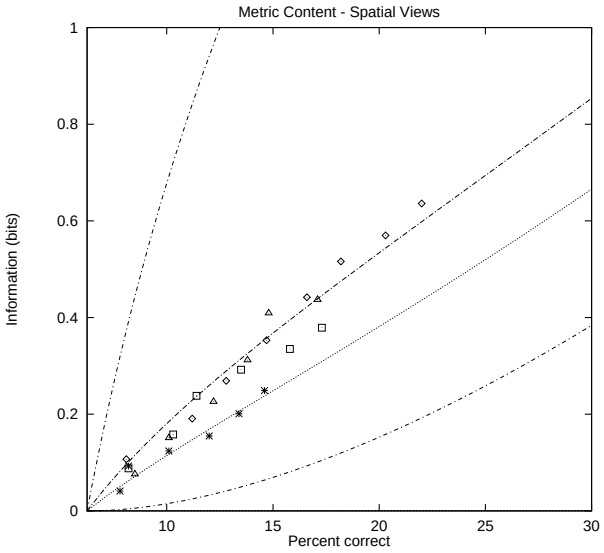

Figure 1: The information decoded from different cell populations *vs.* the corresponding percent correct. I_{min} and I_{max} are indicated with dash-dotted lines, along with the curve $\lambda = 0.25$. The other curve is $\lambda = 0.15$. Datapoints are for populations of CA3 ($*$), CA1 (triangles), parasubiculum (squares) and parahippocampal gyrus cells (diamonds).

ulation vectors [1], which live in the physical 3D or 2D movement space rather than in the space of dimensionality equal to the number of cells included, and which correspond to a continuous rather than a discretized correlate. Vectors were constructed including all 26 cells, only the cells recorded in each animal, only those in a given brain region of both animals, and so on exploring different combinations. Since the metric content index is expected to be relatively constant as the number of cells randomly sampled from an homogeneous population varies, comparisons can be made, to some extent, even among the metric content characterizing vectors of different dimensionality.

One can see from the figure the extent to which metric content, while not being a strictly invariant characterization of the responses, valid for all percent correct values, is still a relatively stable index. For each given cortical area, as more cells are considered, both percent correct and decoded information grow, and the relation between the two, expressed as metric content, varies somewhat in a limited band of values characterizing each cortical area. One should note that the variability as the number of cells varies is limited only because of the extensive averaging we perform, e.g. when decoding from 3 CA1 cells, over nearly all possible triplets of cells from the 6 available from the CA1 area. Obviously, this averaging cannot compensate from the fluc-

tuations induced by the very limited number of cells – 6 – in the total CA1 sample. This is ultimately one of the main limits of this preliminary analysis, which prevents us from drawing definite conclusions.

The figure shows the individual datapoints obtained for the average sample of c cells from each cortical area, and also two representative lines of constant metric content, one for $\lambda = 0.25$ and one for $\lambda = 0.15$. Datapoints from hippocampal area CA1 ($c = 6$), from the parahippocampal gyrus (PHG, $c = 8$) and from the parasubiculum (PSUB, $c = 6$) tend to cluster around the upper metric content curve, while datapoints for hippocampal area CA3 ($c = 6$) tend to cluster around the lower curve. As it happened, 4 out of 6 CA3 cells were recorded in monkey *az*, while all but 2 (1 CA1 and 1 PSUB) of the other cells were from monkey *av*. Extensive testing with subsets of cells taken from both the same area *and* the same monkey failed to clarify conclusively whether the emerging metric content difference is due to the area or to the monkey.

4 Comments and outlook

The data analysed in this paper are not fully adequate, on at least two accounts. First, the number of cells recorded and the number of 100ms trials available for each cell and each spatial view were not sufficiently large to safely avoid limited sampling effects. Second, the recordings should be simultaneous, and from the same monkey, to avoid differences due to slow changes in the representations e.g. with learning or habituation or increasing boredom, or due to individual differences. Both inadequacies can be removed with parallel recording from several cells at once, so the preliminary results of the type of analysis presented here will soon be confirmed or disproved by analysing more adequate data.

Within these limits, one possible interpretation of the different metric content in the CA3 area, with respect to the other 3 areas sampled, lies in the different pattern of connectivity, whereby in CA3 recurrent collateral connections are the numerically dominant source of inputs to pyramidal cells, and travel relatively long distance, to form an extended network connected by intrinsic circuitry. Considerations based on simplified network models suggest that such a connectivity pattern would express memory representations with a different metric structure from those expressed by networks of different types. The difference could be further related to the qualitative nature of the memory representation, which might be characterized as being more *episodic* in CA3 and more *structured* in the other areas. The metric content depends also on the average sparseness of these representations, though, and further analyses are required to dissociate the effects of connectivity (and of representational structure) from those purely due to changes in sparseness.

The present recordings were from neighbouring areas in the temporal lobes, and it is possible that any difference among memory representations

will be more striking when more distant areas are compared. In addition, it is possible that any difference may be more striking when the correlate considered does not have its own intrinsic metric, as with spatial views, but instead lives in a high dimensional space, as e.g. with faces, thereby letting more room for arbitrary metric structures to be induced in the neural representations by the learning process. For both reasons, it will be interesting to extend this analysis to entirely different experiments, sharing with the present one only the generic requirement that different populations of cells are recorded in their response to the same set of stimuli, or in general correlates.

Finally, possible changes in the representations that develop with time could be examined by recording from the *same* populations – not the same cells – over periods during which some behaviourally relevant phenomenon may have occured, such as new learning, forgetting, or a modulation of the existing representations. One specific such modulation of interest for the case of human patients is the one resulting from localized lesions to another cortical area, which may affect the structure of the representations in surviving areas of the cortex.

Acknowledgements

Partial support was from the Medical Research Council of the UK, the Human Capital and Mobility Program of the EU, and the National Research Council of Italy.

References

[1] Georgopoulos, A.P., Kettner, R.E. and Schwartz, A.B.: Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. J. Neurosci. *8*, 2928-2937 (1988)

[2] Hubel, D.H. and Wiesel, T.N.: Sequence regularity and geometry of orientation columns in the monkey striate cortex. J. Comp. Neurol. **158**, 267-294 (1974)

[3] Lauro-Grotto, R., Piccini, C., Borgo, F. and Treves, A.: What remains of memories lost in Alzheimer and herpetic encephalitis. Soc. Neurosci. Abs. **23**, 734.2 (1997)

[4] Miyashita, Y. and Chang, H.S.: Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature **331**, 68-70 (1988)

[5] O'Keefe, J.: A review of the hippocampal place cells. Progr. Neurobiol. **13**, 419-439 (1979)

[6] Panzeri, S. and Treves, A.: Analytical estimates of limited sampling biases in different information measures. Network **7** 87-107 (1996)

[7] Panzeri, S., Treves, A., Schultz, S. and Rolls, E.T.: On decoding the responses of a populations of neurons from short time windows. Oxford U., Dep. of Exp. Psych. preprint (1998)

[8] Petersen, R.S., Treves, A., Lebedev, M. and Diamond, M.: Information theoretic analysis of the responses of rat cortical neurons to vibrissal stimulation. Soc. Neurosci. Abs **23**, 913.17 (1997)

[9] Rolls, E.T.: Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Phil. Trans. Roy. Soc. **B 335**, 11-21 (1992)

[10] Rolls, E.T., Treves, A. and Tovee, M.J.: 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 (1997)

[11] Rolls, E.T., Treves, A., Robertson, R.G., Georges-Francois, P. and Panzeri, S.: Information about spatial view in an ensemble of primate hippocampal cells. J. Neurophysiol. **79**, 1797-1813 (1998)

[12] Sullivan, S.L. and Dryer, L.: Information processing in mammalian olfactory system. J. Neurobiol. **30**, 20-36 (1996)

[13] Tanaka, K.: Neuronal mechanisms of object recognition. Science **262**, 685-688 (1993)

[14] Treves, A.: On the perceptual structure of face space. Biosystems **40**, 189-196 (1997)

[15] Treves, A. and Panzeri, S.: The upward bias in measures of information derived from limited data samples. Neural Comp. **7**, 399-407 (1995)

[16] Treves, A., Skaggs, W.E. and Barnes, C.A.: How much of the hippocampus can be explained by functional constraints? Hippocampus **6**, 666-674 (1996)

[17] Wilson, M. and McNaughton, B.L.: Dynamics of the hippocampal ensemble code for space. Science **261**, 1055-1058 (1993)