



Contents lists available at [SciVerse ScienceDirect](#)

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth



A novel test to determine the significance of neural selectivity to single and multiple potentially correlated stimulus features

Robin A.A. Ince^{a,1}, Alberto Mazzoni^{b,1}, Andreas Bartels^{a,d}, Nikos K. Logothetis^{a,e}, Stefano Panzeri^{b,c,f,*}

^a Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany

^b Robotics, Brain and Cognitive Sciences Department, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy

^c Bernstein Center for Computational Neuroscience, Tübingen, Germany

^d Vision and Cognition Lab, Centre for Integrative Neuroscience, University of Tübingen, 72076, Tübingen, Germany

^e Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, United Kingdom

^f Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom

ARTICLE INFO

Article history:

Received 30 June 2011

Received in revised form 7 November 2011

Accepted 8 November 2011

Keywords:

Vision

LFP

Neural coding

Natural movies

Mutual information

Oscillations

Gamma band

ABSTRACT

Mutual information is a principled non-linear measure of dependence between stochastic variables, which is widely used to study the selectivity of neural responses to external stimuli. Here we define and develop a set of novel statistical independence tests based on mutual information, which quantify the significance of neural selectivity to either single features or to multiple, potentially correlated stimulus features like those often present in naturalistic stimuli. If the values of different features are correlated during stimulus presentation, it is difficult to establish if one feature is genuinely encoded by the response, or if it instead appears to be encoded only as a side effect of its correlation with another genuinely represented feature. Our tests provide a way to disambiguate between these two possibilities. We use realistic simulations of neural responses tuned to one or more correlated stimulus features to investigate how limited sampling bias correction procedures affect the statistical power of such independence tests, and we characterize the regimes in which the distribution of information values under the null hypothesis can be approximated by simple distributions (Chi-square or Gaussian). Finally, we apply these tests to experimental data to determine the significance of tuning of the band limited power (BLP) of the gamma [30–100 Hz] frequency range of the primary visual cortical local field potential to multiple correlated features during presentation of naturalistic movies. We show that gamma BLP carries significant, genuine information about orientation, space contrast and time contrast, despite the strong correlations between these features.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Mutual information (Cover and Thomas, 1991) (hereafter referred to as information) is a measure of dependence between stochastic variables. Information quantifies the reduction in uncertainty about one variable (for example, a sensory stimulus) after observation of another (for example, a neural response). Information is a popular metric to quantify, on a scale that allows meaningful comparisons across different systems, how much a given type of a neural response is tuned to one or more stimulus parameters (Borst and Theunissen, 1999; Ince et al., 2010; Optican

and Richmond, 1987; Strong et al., 1998). Its main advantage is that it is based on the full probability distributions of the considered variables and can therefore quantify the effect of any non-linear or arbitrarily complicated dependency.

The calculation of neural information values and the evaluation of their significance is made difficult by the fact that in neurophysiological experiments the joint probabilities of stimulus and neural responses have to be estimated empirically from a limited number of data samples (“trials”). Limited sampling induces both a systematic error (bias) and a variance in the estimation of the neural information values. While the limited sampling bias has been studied extensively in the neural literature (Paninski, 2003; Panzeri et al., 2007; Victor, 2006), the evaluation of the statistical significance of an information value computed from limited experimental data has been less thoroughly investigated. The statistical significance of an information value tells us whether we can, at a particular confidence level or *p*-value, reject the null hypothesis that the stimulus and response are independent. The

* Corresponding author at: Robotics, Brain and Cognitive Sciences Department, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy.

Tel.: +39 331 6933490; fax: +39 010 7170817.

E-mail address: stefano.panzeri@iit.it (S. Panzeri).

¹ These authors contributed equally to this work.

significance of information values is traditionally calculated by employing a form of bootstrap test (Optican et al., 1991). The stimuli and response are shuffled across experimental trials to remove any dependence, and the information value is calculated for many repetitions of this shuffling procedure to obtain the distribution of values under the null hypothesis of independence. However, several problems related to the estimation of significance have not yet been considered. The first problem is whether it is possible to employ parametric models for the null hypothesis distribution. The second problem regards the effect of sampling bias corrections on the statistical power of such tests. The third problem is to determine which of the many stimulus features characterizing naturalistic stimuli actually influence the considered neural response. This is difficult because in dynamic and complex stimuli, like the ones found in a natural environment and mimicked by movie stimuli, different stimulus features are often strongly correlated (Felsen and Dan, 2005; Geisler, 2008; Simoncelli, 2003; Simoncelli and Olshausen, 2001). An abrupt movement of an observer, for instance, which leads to sudden appearance or disappearance of objects, will necessarily produce concurrent, and hence correlated, changes in a large number of features (Bartels et al., 2008). If the neural response carries significant information about each of two different but correlated features, it does not follow necessarily that the neural response is genuinely influenced by both of them. The response could instead be influenced by only one of the features, with the dependence on the other feature appearing because of the correlations between them. Understanding how to determine true neuronal selectivity in these cases is essential to allow meaningful investigation of the neural representation of natural stimuli.

Here we investigate these issues by first comparing bootstrap based significance test results for single feature information values including different bias corrections, and presenting comparisons with an analytic expression for the null hypothesis distribution as well as parametric fits of simple distributions to small bootstrap samples. This lays the basis to understand how best to perform a statistical test of independence based on mutual information. We then investigate the extent to which dependencies between variables complicate the estimation of genuine tuning to each feature. We construct statistical tests for inferring true selectivity in the presence of multiple co-varying features. We validate these methods by testing them on realistically simulated neuronal responses to multiple correlated features, and finally we illustrate them by investigating the extent to which the gamma band limited power (BLP) of local field potentials (LFPs) recorded from the V1 area of anesthetized macaques presented with naturalistic movies is genuinely tuned to one or more of several visual features characterizing the movies. We found that orientation, space contrast and time contrast are each genuinely encoded by the gamma BLP, in the sense that tuning to each of these features cannot be accounted for by tuning with respect to the other, correlated features. Moreover, we found that correlation between these different stimulus features across different frames in the movie increased the overall amount of information that the gamma BLP carried about the visual stimuli.

2. Testing the significance of mutual information values

2.1. Mutual information and limited sampling bias

Suppose that a sensory stimulus specified by multiple features is presented to a subject while recording neural responses. One way to quantify whether a neural response, R , is selective to a stimulus S , is to measure the information between the set of stimuli and the set of associated neural responses (de Ruyter van Steveninck et al.,

1997; Fairhall et al., 2001; Panzeri et al., 2003), which is defined as follows:

$$I(R; S) = \sum_s P(s) \sum_r P(r|s) \log_2 \frac{P(r|s)}{P(r)} \quad (1)$$

where $P(s)$ is the probability of presentation of stimulus s , $P(r|s)$ is the probability of observing response r when stimulus s is presented, and $P(r)$ is the probability of observing a response r across all stimulus presentations. In the following we assume that the values of both the neural response R and the stimuli S are discrete, or alternatively are discretized into a finite number of bins to facilitate the experimental sampling of probabilities. Information is non-negative and quantifies the average reduction of the uncertainty about the stimulus that can be gained from observing a single-trial neural response. Here we measure it in units of bits, one bit corresponding to a reduction of uncertainty by a factor of two. If information is larger than zero, then in principle the stimulus feature and neural response are not independent, indicating that the response reflects the particular stimulus presented.

Direct evaluation of the information defined in Eq. (1) from the experimental histograms of stimulus-response occurrences (usually called the “plugin” estimate, see Nemenman et al. (2004)) results in a systematic upward bias in the information estimate. This occurs because of finite sampling; the variability of the distributions will be systematically under-estimated because of the limited amount of data available. It can be shown that in sufficiently well sampled situations (more precisely, in the so called asymptotic regime in which each response is observed many times during the experiment) the bias of the plug-in information estimate can be approximated by the following expression (Panzeri and Treves, 1996):

$$\text{BIAS}[I(R; S)] = \frac{1}{2N \ln 2} \left\{ \sum_s (\hat{R}_s - 1) - (\hat{R} - 1) \right\} \quad (2)$$

where N is the total number of trials, \hat{R}_s is the number of possible responses upon presentation of stimulus s and \hat{R} is the number of possible responses across all stimuli presentations. The sum in the right-hand side is over all stimuli presented. Eq. (2) is useful to build an intuition about the bias problem. Assuming that \hat{R}_s is approximately equal to \hat{R} , and constant across stimuli, then the bias (Eq. (2)) is proportional to $(\hat{S} - 1)(\hat{R} - 1)/N$, where \hat{S} is the number of possible stimuli. If we define $N_s = (N/\hat{S})$ to be the average number of trials per stimulus then we can see the bias of the information is proportional to $((\hat{R} - 1)/N_s) - ((\hat{R} - 1)/N)$ and hence approximately proportional to \hat{R}/N_s . We therefore consider N_s/\hat{R} as the key parameter, which affects the magnitude of the bias; if this ratio is large, the bias will be small and vice versa.

There are several methods to estimate the bias (Montani et al., 2007; Nemenman et al., 2004; Paninski, 2003; Panzeri et al., 2007). For example, the bias can be estimated from Eq. (2) (and then subtracted out). This requires estimating parameters such as \hat{R}_s , the number of possible responses to stimulus s . This can be estimated either from simply counting the number of observed responses (irrespective of the probability of any particular response) (Miller, 1955). However, in cases of limited data this simple counting is likely to underestimate the number of possible responses, since possible responses with low probability may not appear. To address this issue a Bayesian counting procedure can be used (see Panzeri and Treves (1996), referred to in the following as the PT method). Alternatively, the quadratic extrapolation (QE) procedure (Strong et al., 1998) calculates information from the full data set as well as from subsets of the data consisting of $N/2$ and $N/4$ trials, and then estimates the true asymptotic value of information by fitting a quadratic curve to these data points. While these methods

have been shown to be greatly effective in reducing the systematic error due to limited sampling in information estimation (in particular, they are very effective if N_s/\hat{R} is larger than 1, see Panzeri et al. (2007) for a review), the effect of these bias corrections on estimating the significance of information values has been largely unexplored so far.

2.2. The statistical significance of mutual information

To conclude that a neural response is selective for a specific feature it is not sufficient to measure that the information between the two variables is larger than zero. To rule out the possibility that this information value results only from fluctuations due to limited sampling rather than from true dependency, one must also assess the statistical significance of this measure.

The statistical significance of information values is traditionally evaluated through the method of surrogate data, which is referred to here as the bootstrap method (Optican and Richmond, 1987). The available stimuli and response data are shuffled to remove any dependence between the variables. The distribution of information values computed across different instantiations of the shuffled dataset follows the distribution of the information value under the null hypothesis that the variables are statistically independent, since any dependence has been removed by the shuffling procedure. After selecting the required significance level (p -value), the corresponding percentile of the bootstrapped null hypothesis distribution is calculated and the value obtained from the un-shuffled data is compared to this to determine statistical significance.

While the bootstrap procedure is straightforward to perform, it is problematic when small p -values are required, since accurate characterization of the extremes of the null distribution requires high numbers of bootstrap repetitions. Besides being computationally demanding, obtaining a high number of independent shuffled combinations of the same data set requires a very large amount of data, which can in some cases be difficult to collect. It is therefore interesting to consider whether parametric models can be fit to smaller bootstrap sample populations, without compromising the statistical power of the test. In fact, it is a long established, although apparently little known result that for large numbers of samples the distribution of plugin (not bias corrected) information estimates under the null hypothesis of independence, when multiplied by a factor of $2N \ln 2$, follows a chi-square distribution (Fan et al., 2000; Wilks, 1938):

$$2N \ln 2 \cdot I(R; S) \sim \chi^2((\hat{R} - 1)(\hat{S} - 1)) \quad (3)$$

If the stimuli presented are drawn from a set containing \hat{S} different stimuli, and the response is represented with \hat{R} discrete symbols, then the degrees of freedom for this chi-square null distribution are $(\hat{R} - 1)(\hat{S} - 1)$. Note the relationship with the first order bias estimate in the previous section. The mean of this chi-square distribution, when converted back into the information scale by dividing by $2N \ln 2$, is precisely the first order estimate of the bias.

This null hypothesis distribution is the same as the asymptotic null distribution for Pearson's chi-square test of independence (Kullback, 1968). In fact, the scaled mutual information value described above is equal to the test statistic for a likelihood ratio test of independence. The chi-square test statistic can be derived as a quadratic approximation to this likelihood ratio test, by taking a Taylor series expansion (Kullback, 1968). However, with the availability of modern computing resources calculating logarithms is no longer a substantial challenge and so there is less of an advantage to using the approximate quadratic form. An additional point of interest is that the mutual information statistic is itself the dominant term in a full Bayesian test with uniform priors (Wolf, 1995).

However, not much is known yet about the extent to which the null hypothesis distribution can be approximated by parametric

distributions in conditions of moderate sampling, and about the effect of bias corrections upon the effectiveness of the bootstrap test. These issues are of high practical importance. In the following we investigate them by means of realistic numerical simulations of neural responses.

2.3. Investigation of performance of information significance tests using simulations

To investigate the statistics of information values and the efficiency of tests of information significance, we conducted a series of numerical experiments with models of simulated neural responses, which are both realistic and relevant to the neurophysiological analysis that we report in Section 5. In brief, the simulation is constructed to match the first and second order statistics of the gamma-range BLP of LFPs recorded in primary visual cortex (V1) in response to variations of Michelson contrast during the presentation of a naturalistic movie.

Asymptotic theorems of spectral estimation predict that spectral powers are approximately chi-square distributed (Percival and Walden, 1993; Reich et al., 2000). Hence, the cube root of the BLP is approximately Gaussian (Wilson and Hilferty, 1931), a property that was confirmed empirically (not shown) for the data we analyzed in Section 5 (see Magri et al. (2009) for an explicit test of this assumption). Therefore, our model generated LFP responses by assuming that their cubic root followed a Gaussian distribution with the same mean and variance as the cubic root of BLP power from an example recording site (see Section 5.1) in macaque V1 in response to 8 quantized levels of the Michelson contrast in the receptive field of the considered recording site during the presentation of a Hollywood movie. For computing information, the simulated BLP gamma responses were discretized into $R=4$ equipopulated bins (exactly as we do for real data in Section 5). We refer to the Appendix A for full details about the simulation.

Intuition suggests that how difficult it is to determine whether an information value is significant depends both on the number of trials available (the more data there are, the easier it is to sample the probabilities) and also on the amount of information in the neural responses (the less information there is in the responses, the more trials may be needed to determining significance). To investigate these effects, we introduced a parameter α that modulated the information in the simulated responses by transforming the mean and standard deviation parameters as follows:

$$p_s^{\text{model}} = p_s^{\text{data}} + (1 - \alpha)(\overline{p_s^{\text{data}}} - p_s^{\text{data}}) \quad (4)$$

where p_s^{model} is the stimulus conditional parameter (mean or standard deviation) of the model for stimulus s , p_s^{data} is the corresponding parameter measured from the data, $\overline{p_s^{\text{data}}}$ is the average of the parameter across stimuli and α controls the modulation of the information in the simulated responses. If $\alpha = 1$, the simulated responses have the same properties (and information) as the measured data. Decreasing α decreases the information in the model responses, until for $\alpha = 0$ the model responses are not stimulus modulated any more, and hence they carry no information. We show results for $\alpha = 1$ (full information), $\alpha = 0.5$ and $\alpha = 0$ (no information).

We first investigated the statistical power of the bootstrap test in detecting significant information in cases where the model responses had true information about the simulated stimuli. We used 100 random bootstraps to estimate the null hypothesis distribution, which corresponds to relatively large sampling compared to what can be done feasibly in analyses of real datasets, and thus represent a good benchmark. Fig. 1A shows how the percentage of correct detections of significant information, across different realizations of the simulation, depends on the number of trials per stimulus. For $\alpha = 1$ (information in simulation matching the

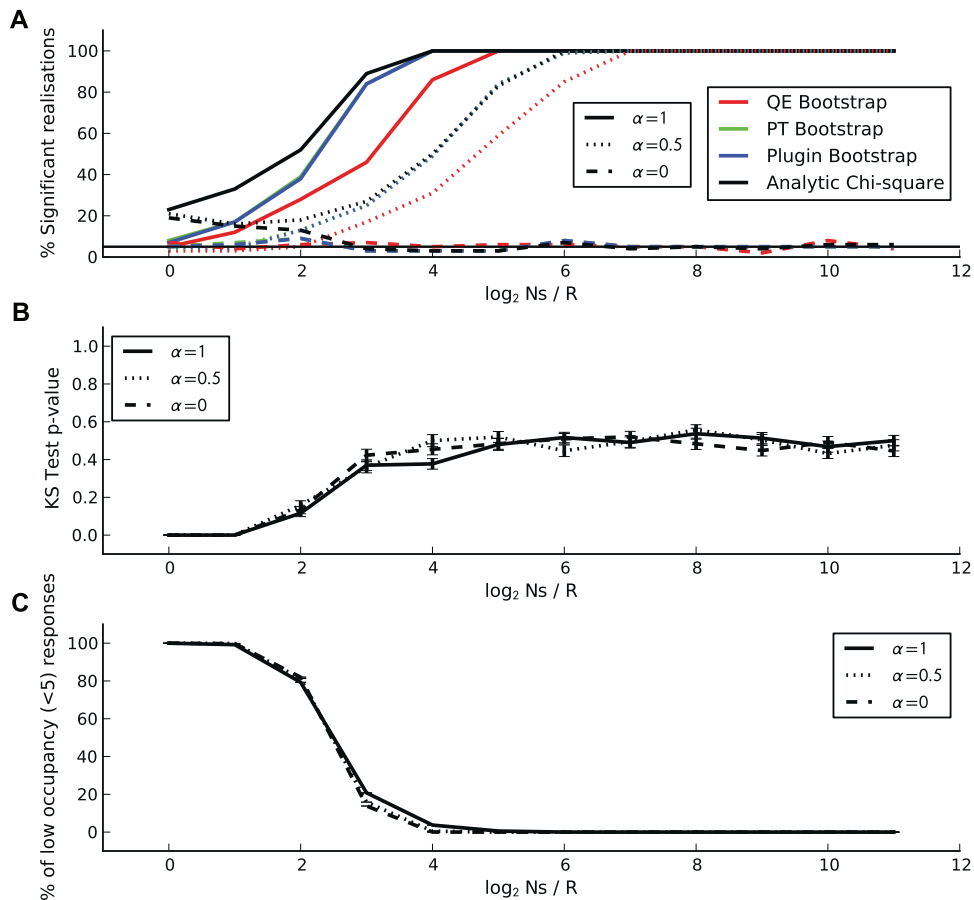


Fig. 1. Determining the significance of mutual information. Model LFP gamma BLP responses to eight different values of Michelson contrast were simulated (see text for details). The number of trials generated per stimulus was varied; for each number of trials 100 realizations of the simulation were generated and the calculations were performed on each realization. The x-axes show the N_s/R ratio, the critical parameter for information estimates (see text, here $R=4$). Panel (A) shows the percentage of realizations that were classed as significant using a variety of methods. Solid lines show the full information model ($\alpha=1$); dotted lines show the $\alpha=0.5$ model; dashed lines show the no information ($\alpha=0$) model. For the QE (red lines) and PT (green lines; partially obscured by overlapping blue curve) bias correction methods the corrected information value was compared to the bootstrap distribution of corrected values calculated from 100 shuffling repetitions. For the chi-square method (black lines) the analytical null hypothesis distribution was used and no bootstrap shuffling was performed. Panel (B) shows the mean (errorbars show standard deviation) over 100 realizations of the p-value from the Kolmogorov–Smirnov test of the goodness of fit of the 100 bootstrap samples to the analytic chi-square distribution. Panel (C) shows the percentage of stimulus conditional response bins with less than 5 observations.

one in real responses), the bootstrap method reliably detected significance down to 32 trials per stimulus (which corresponds to $N_s/R=8$). This shows that detecting significance requires more trials than would be normally required for bias free information estimation (bias techniques are shown to be effective if N_s/R is larger than one, see Panzeri et al. (2007)). When the information content was reduced ($\alpha=0.5$), more samples were required to reliably detect significance. This matches the intuition that with enough trials an arbitrarily small dependence could be reliably detected, while a very strong dependence could be detected with few samples.

The analysis of the model with no information ($\alpha=0$, Fig. 1A, dashed lines) shows the false positive rate of the tests. We found that for all methods this was stable and close to the specified significance level ($p=0.05$).

Across all information levels simulated, the use of bias corrections was never helpful in the bootstrap calculation of significance: it reduced (QE) or did not affect (PT) the statistical power of the test. The reduction in statistical power is likely to occur because of the increased variance of the bias corrected estimates; most methods trade off reduced bias in the estimate for an increased variance of the estimator. In the following, we will therefore concentrate on using uncorrected plugin estimators to determine significance.

We then examined the effectiveness of using the parametric analytical chi-square distribution of values under the

null hypothesis, rather than the empirical non-parametric bootstrap distribution. Fig. 1B shows that the analytic chi-square approximation, with degrees of freedom equal to $(\hat{R}-1)(\hat{S}-1)$, described the bootstrap distribution of information values well (Kolmogorov–Smirnov (KS) test, $p>0.3$) provided that there were at least 32 trials per stimulus, i.e. $N_s/R>8$. In this sampling regime, since the chi-square distribution fits the bootstrap distribution well, the chi-square can be used to both reduce computational time and to compute significance for small p-values which could be difficult to test with the bootstrap method. We verified that, in this sampling regime, the analytic chi-square test was as good as the bootstrap test in both determining true positives and avoiding false negatives.

With fewer trials ($N_s/R<8$) the analytical chi-square no longer modeled the obtained bootstrap distributions accurately. In these cases, the responses were under-sampled, with a large number of bins having occupancy of less than 5 (Fig. 1C). These results are in agreement with conventional statistical practice regarding the application of chi-square tests (Cochran, 1954; Larntz, 1978), which suggests using the chi-square approximation only when the number of expected observations under the null hypothesis of independence for each possible stimulus conditional response is greater than 1, and is greater than 5 for at least 80% of stimulus conditional responses. In situations where this was not the case, the statistical

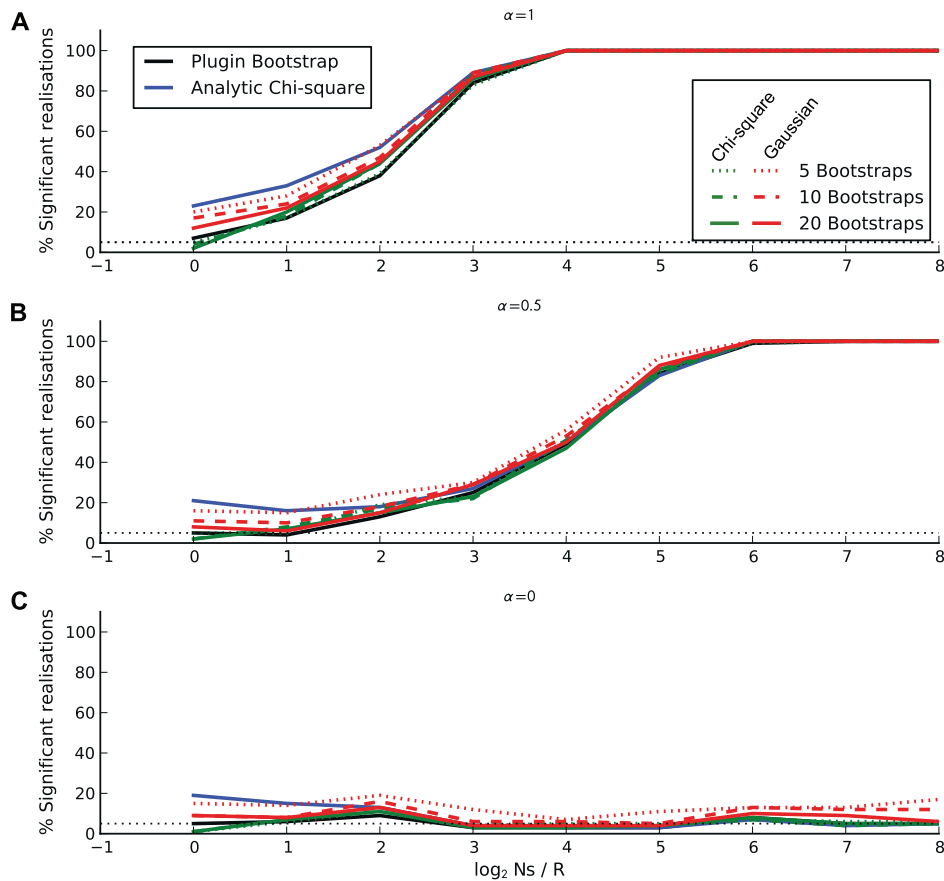


Fig. 2. Parametric models of null hypothesis bootstrap distribution. Model LFP gamma BLP responses to eight different values of Michelson contrast were simulated (see text for details). The number of trials generated per stimulus was varied; for each number of trials 100 realizations of the simulation were generated and the calculations were performed on each realization. The x -axes show the N_s/R ratio, the critical parameter for information estimates (see text, here $R=4$). Results are shown for the full information model ($\alpha=1$), the $\alpha=0.5$ model and the no information ($\alpha=0$) model. For each realization a chi-square (green lines) or Gaussian (red lines) distribution was fit to either 5 (dotted lines), 10 (dashed lines) or 20 (solid lines) shuffled bootstrap values, and this distribution was used to estimate the significance of the measured information value.

test based on the chi-square distribution performed poorly, with a marked increased false positive rate.

In many situations it is not possible to compute a large number of bootstraps and so it is difficult to compute significance for small p -values. Besides using the analytic chi-square approximation, another possible way to compute small p -value significance under such circumstances is to fit parametric models to null hypothesis distributions based on small numbers of bootstrap samples. These empirically determined parametric models can then be used to evaluate the significance of the observed information value. In Fig. 2 we tested the effectiveness of such procedures, by fitting either chi-square or Gaussian distributions to distributions of information values obtained from small numbers (5–20) of independent bootstrap samples. Even in under-sampled situations where, as discussed above, the analytic chi-square approximation fails for $N_s/R < 2^3$ (shown by increased false positive rate in Fig. 2C and small KS test p -values in Fig. 1B), good results were still obtained with these parametric fits. The chi-square fit performance was very close to the full non-parametric bootstrap (95th percentile of 100 bootstrap samples), even when only 5 bootstrap samples were used. This provides a large computational saving. The Gaussian model also performed well, but 20 bootstrap trials were required to maintain a close fit to the non-parametric bootstrap in the under-sampled regime, and control of the false positive rate throughout the sampling ranges considered. Although the degrees of freedom are high, suggesting the chi-square is close to a Gaussian, when fitting the Gaussian model two parameters must be estimated from the data rather than just one for the chi-square distribution. This

could explain the increased performance of the simpler model for low numbers of samples.

In summary, our analysis of the tests of independence gave a number of practically useful results: (i) while accurate estimates of information values require bias correction, when testing for significance of information, best results are obtained with uncorrected plug-in estimates. (ii) The analytic chi-square null hypothesis distribution allows a very rapid and highly precise test even with small p -values provided sampling is sufficient ($N_s/R > 8$ in our simulation). The regime in which the analytic chi-square is accurate can be determined in real data by using the KS test between empirical bootstrap and analytic chi-square distributions for sub-samples of different sizes. (iii) Even in sub-sampled situations where the analytic chi-square approximation is not accurate and it is not possible to compute many bootstrap samples, it is still possible to obtain good results by empirically fitting a chi-square distribution to a small number of bootstrap samples.

2.4. The effect of number of bins used in the discretization of a continuous signal upon the evaluation of significance of information

An important step in the information analysis of discretized analog signals, such as the LFP BLP response, is the binning of data into a number of categorical bins. Two issues regarding the binning procedure require further discussion. The first issue is the binning algorithm to be chosen. Here, we set the boundaries of the bins so that they are approximately equally occupied over all

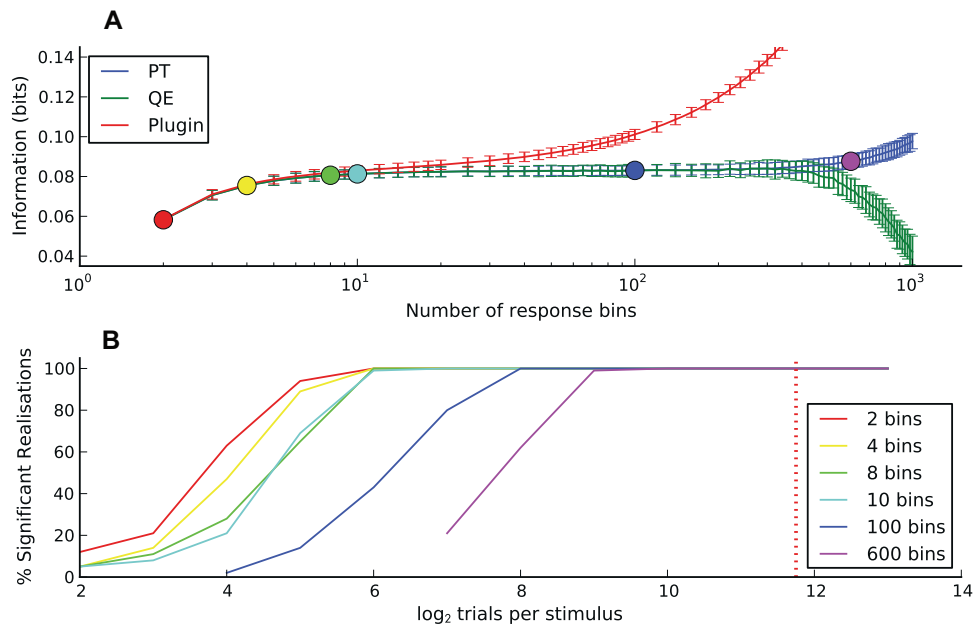


Fig. 3. Effect of number of bins on information estimation and statistical significance. Model LFP gamma BLP responses to eight different values of Michelson contrast were simulated (see text for details). Panel (A) shows average information estimates over 100 realizations (errorbars show standard deviation) of the simulation with the same number of trials as the experimental data (3460 trials per stimulus) as a function of number of bins used to discretize the responses. Information is estimated without bias correction ('Plugin', red line), with Panzeri–Treves bias correction ('PT', blue line) and with quadratic extrapolation ('QE', green line). Panel (B) shows the percentage of realizations found to be statistically significant (100 realizations, $p=0.05$, bootstrap method), as a function of the number of trials per stimulus, for different numbers of response bins. The corresponding information values for each curve are indicated by corresponding colored marks in panel (A). The vertical red dotted line indicates the number of trials available in the experimental data set and used for the simulations in panel (A).

the available trials (the unconditional or marginal distributions are approximately uniform). This avoids sampling problems caused by bins having very low probability of occupancy, for example if equal-width bins are used to represent a Gaussian signal, those sampling the tails of the distribution have a very low probability of occurrence. This procedure makes the information essentially a reduced rank statistic, since the relative position (rank) of a particular trial determines which bin it is in. An advantage of this procedure is that, like other rank statistics, it is non-parametric and robust to outliers. The second issue to be considered is that, whatever the binning method used, the number of bins used for discretization is an important, yet partly arbitrary, parameter that may greatly affect the information calculation.

The choice of the number of bins is a compromise between the need to use a fine enough discretization in order to capture all possible informative features in the data, and then need to use a discretization coarse enough that the probability of each discretized response can be sampled accurately with the available data. For any given dataset, selecting more bins means restricting the average number of trials per bin. In the following we use computer simulation to illustrate how to empirically evaluate what is the optimal trade-off between these two conflicting constraints in order to evaluate the significance of information values.

Fig. 3A shows how estimates of the mutual information depend on the number of response bins chosen for the simulated data set generated with the model described in Section 2.3. In Fig. 3A the simulated dataset had exactly the same number of trials as the real LFP data analyzed in Section 5 (3460 trials in each of the 8 stimulus categories). We studied the dependence of information values upon the number of bins using bias-corrected information estimates, because the dependence of bias-uncorrected estimates on the number of bins is masked by the linear increase of the bias with this parameter (see Eq. (2) and the bias-uncorrected information curve in Fig. 3A). For very low numbers of bins, increasing the number of bins increases the information, because the more finely

discretized responses better capture the underlying continuous distributions. However, bias-corrected information values plateau over a large range of bin numbers, in which the information values are stable and largely independent of the bias correction procedure used. This means that, within the range of number of bins at which information plateaus the sampling is sufficient to allow an accurate estimation of the information, but increasing the number of bins does not have an effect; there are already sufficient bins to capture the relevant features of the distribution. For an accurate estimate of the information value, any parameter within the plateau region can be chosen. When estimating the significance of the information, it is intuitive that using more bins reduces the available samples per bin, and hence the statistical power of the test, but there must be sufficient bins for information to be detectable. Fig. 3B shows the percentage of realizations in which significance was detected ($p=0.05$, bootstrap method) for the simulated system described above, as a function of the amount of generated data. In this case, using just 2 bins provides the greatest statistical power, since the information with two bins is already large enough and is a large fraction of the plateau value of information (Fig. 3A).

These results suggest some practical considerations that may guide the choice of the number of bins for significance testing. Whenever possible, bias-corrected information estimates of how information scales with the number of bins, or empirical observation about the shape and width of the distributions, should be used to evaluate the range of number of bins that gives sufficiently high information values. The smallest number of bins sufficient to preserve a large fraction of the information should be used for the information significance testing in order to maximize the statistical power of the significance tests. Whenever possible, it is also important to verify the stability of the results over a range of bin numbers because in some cases using too few bins may lead to very large losses of information with respect to their plateau values.

3. The multiple feature tuning problem in the context of information theory

If a set of stimuli are defined by multiple features, F_i , we can consider the mutual information between the neural response and each individual feature:

$$I(R; F_i) = \sum_{f_i} P(f_i) \sum_r P(r|f_i) \log_2 \frac{P(r|f_i)}{P(r)} \quad (5)$$

where f_i represents values of the feature F_i and the probabilities are as in Eq. (1). However, if these features are not independently distributed across different presentations of the stimuli, a positive value of information $I(R; F_i)$ about feature F_i does not necessarily imply that the neural response specifically encodes that feature. The neural response may not be selective to F_i , but instead be selective to another feature (say F_j), which is correlated with F_i . When using correlated stimuli, an important problem is how to distinguish the true selectivity to a feature from that acquired only from its relations to other features.

3.1. Multiple feature tuning and information complementarity

Here we address the issue of determining genuine feature selectivity in the presence of multiple correlated features by considering, for simplicity, the case in which the stimulus is defined by two different correlated features, F_1 and F_2 . The amount of dependency between the two features is measured, in information theoretic terms, by the information $I(F_1; F_2)$ between them:

$$I(F_1; F_2) = \sum_{f_1, f_2} P(f_1, f_2) \log_2 \frac{P(f_1, f_2)}{P(f_1)P(f_2)} \quad (6)$$

where $P(f_1, f_2)$ is the joint probability of feature values f_1 and f_2 being presented in the same stimulus. The two features are dependent if $I(F_1; F_2) > 0$. If the neural response is selective to both features (i.e. $I(R; F_1) > 0$ and $I(R; F_2) > 0$, with $I(R; F_1) > I(R; F_2)$), how can we make sure that the given neural response truly encodes both features, and does not encode F_2 only because it correlates with the truly encoded feature F_1 ? One way to address this problem is to look at the information the response conveys about F_2 at fixed values of F_1 . Since the value of F_1 is fixed, any dependence between the response and F_2 represents *genuine* complementary information that cannot be explained by the dependency between the features. This quantity is called the conditional mutual information (CMI) and is defined as:

$$I(R; F_2|F_1) = \sum_{f_1} P(f_1) \sum_{f_2, r} P(r, f_2|f_1) \log_2 \frac{P(r, f_2|f_1)}{P(r|f_1)P(f_2|f_1)} \quad (7)$$

where $P(r, f_2|f_1)$ is the joint probability of observing response r and feature f_2 at fixed f_1 . In a situation where

$$I(R; F_2|F_1) = 0 \quad (8)$$

it can be proved (Cover and Thomas, 1991) that the following equation holds for all values of f_2, f_1 and r

$$P(f_1, f_2, r) = P(f_2)P(f_1|f_2)P(r|f_1) \quad (9)$$

This illustrates the fact that the variable F_2 affects the neural response only through its correlation with F_1 . If instead $I(R; F_2|F_1) > 0$ then F_2 still has an effect on the variable r at fixed F_1 and so $P(f_1, f_2, r)$ cannot be expressed by the product in the right-hand side of Eq. (9). In summary, the feature F_2 has a genuine effect upon the neural response, beyond the indirect effect of correlations with F_1 , if and only if $I(R; F_2|F_1) > 0$.

It is important to note that, by using the chain rule (Cover and Thomas, 1991), the CMI can be rewritten as:

$$I(R; F_2|F_1) = I(R; F_2, F_1) - I(R; F_1) \quad (10)$$

where $I(R; F_2, F_1)$ is the information that the neural response carries about the joint presence of F_1 and F_2 , defined as follows:

$$I(R; F_1, F_2) = \sum_{f_1, f_2} P(f_1, f_2) \sum_r P(r|f_1, f_2) \log_2 \frac{P(r|f_1, f_2)}{P(r)} \quad (11)$$

Therefore the condition $I(R; F_2|F_1) > 0$ is equivalent to the condition

$$I(R; F_2, F_1) - I(R; F_1) > 0 \quad (12)$$

The requirement that the joint information be higher than the information carried about the variable F_1 implies that the neural response carries “true” information about F_2 above and beyond the information that it carries about F_1 , or in other words, that the information that the neural response carries about F_2 is complementary to that it carries about F_1 (Panzeri et al., 2010). Alternatively, due to the symmetry of information, the above condition means a better prediction of the response could be made based on observation of both features than could be made based on the observation of F_1 alone. It is intuitive that genuine coding of F_2 requires that the response carry information about F_2 complementary to that which it carries about F_1 . The above equations formalize this intuition.

This framework can in principle be extended to the case in which the stimulus is defined by an arbitrary number n of features (F_1, \dots, F_n). The stimulus feature F_n is genuinely encoded by the neural response if the information $I(R; F_n|F_1 \dots F_{n-1})$ between response R and feature F_n , conditional to all other variables is significantly positive: $I(R; F_n|F_1 \dots F_{n-1}) > 0$. However, while the extension to multiple features is straightforward in principle, the formalism may be difficult to apply to many features because of the difficulties in evaluating multivariate response probabilities from finite amounts of data (Panzeri et al., 2007).

3.2. Effect of correlation of features on establishing information complementarity

We mentioned above that the problem of correlations between stimulus features complicates the problem of determining which stimulus features are genuinely encoded by the neural response. In order to develop a method suited to tackle this problem on experimental data, we investigate theoretically how statistical dependencies between features influence the amount of joint information carried about the features. For simplicity we will limit our discussion to a stimulus defined by only two features, F_1 and F_2 .

The relationship between the CMI, $I(R; F_2|F_1)$, and the correlation between variables $I(F_1; F_2)$ can be defined as follows (Adelman et al., 2003; Lüdtke et al., 2008; Schneidman et al., 2003):

$$I(R; F_2|F_1) = I(R; F_2) + I(F_1; F_2|R) - I(F_1; F_2) \quad (13)$$

where $I(F_1; F_2)$ is the amount of correlation between variables without regard to the neural response (see Eq. (6)), and $I(F_1; F_2|R)$ is the amount of correlation between the features at fixed neural response, defined as follows:

$$I(F_1; F_2|R) = \sum_r P(r) \sum_{f_1, f_2} P(f_1, f_2|r) \log_2 \frac{P(f_1, f_2|r)}{P(f_1|r)P(f_2|r)} \quad (14)$$

Eq. (13) has several important implications. First, it shows that if the features are independently distributed (i.e. $I(F_1; F_2) = 0$) then the condition $I(R; F_2) > 0$ is sufficient to ensure that F_2 carries information complementary to that of F_1 . This is because the term $-I(F_1; F_2)$ is the only one in the right-hand side of Eq. (13), which can be negative. Moreover, and for the same reason, if the value of $I(F_1; F_2)$ is

quantitatively much smaller than that of the information carried by the neural response about each of the considered features, it is safe to conclude that the information about each feature is genuine. This gives a useful rule of thumb for interpretation of the results. Second, Eq. (13) shows that the presence of correlation between features does not always tend to suppress the complementarity of information between features. In fact it can be shown that the term $I(F_1; F_2|R) - I(F_1; F_2)$, sometimes called the interaction information, can be either positive or negative depending on whether correlations between features across all scenes are weaker or stronger than correlations between features at fixed neural response (Pola et al., 2003; Schneidman et al., 2003). This interaction information is equal to the synergy (Brenner et al., 2000b) usually defined as

$$\text{Syn}(F_1, F_2; R) = I(F_1, F_2; R) - I(F_1; R) - I(F_2; R) \quad (15)$$

Together, using Eqs. (13) and (15) we see that

$$I(R; F_2|F_1) = I(F_2; R) + \text{Syn}(F_1, F_2; R) \quad (16)$$

This shows that the absence of complementary information ($I(R; F_2|F_1) = 0$) implies the presence of redundancy (negative synergy) between F_1 and F_2 . Similarly, if the features are coded synergistically, the complementary conditional mutual information can be greater than the unconditional information conveyed about a single feature. In such cases, the coding of groups of features would be difficult to fully characterize with standard uncorrelated stimulus designs. These theoretical considerations reinforce the importance of such methods to enable quantitative investigation of the coding of correlated features found in natural stimulation conditions. Indeed, it has been proposed that responses to natural stimuli are optimized to cope with correlations in the environment (Barlow, 1989; Dan et al., 1996; Geisler, 2008; Olshausen and Field, 1996); the methods discussed here allow quantitative investigations of these ideas.

4. Testing the significance of complementary tuning to multiple features using conditional mutual information

We now consider the problem of how to extend the above tests for significance of mutual information to evaluating the significance of complementarity of tuning to multiple features using CMI. This problem is difficult because the inclusion of multiple features exponentially increases the size of the stimulus space, which correspondingly reduces the number of trials available for each stimulus combination and compounds the sampling difficulties. The correlations between the features can also cause difficulties because they tend to concentrate the joint distributions of the presented combinations of features, and the potentially uneven sampling introduced by correlations among features must be correctly taken into account when designing the appropriate bootstrap shuffling and algorithms that evaluate significance of complementary information values.

4.1. The statistical significance of conditional mutual information

The conditional mutual information can be written as

$$I(R; F_2|F_1) = \sum_{f_1} P(f_1) I(R; F_2|f_1) \quad (17)$$

where $I(R; F_2|f_1)$ is the information between R and F_2 , conditioned on a specific fixed value f_1 of feature F_1 . By substituting in the above equation $P(f_1) = N_{f_1}/N$, where N_{f_1} is the number of trials in which feature F_1 takes the value f_1 , one obtains

$$I(R; F_2|F_1) = \frac{1}{2N \ln(2)} \sum_{f_1} 2N_{f_1} \ln(2) I(R; F_2|f_1) \quad (18)$$

As discussed in Section 2, when using plugin uncorrected information estimates, each element of the sum on the right-hand side of Eq. (18) follows a chi-square distribution with $(\hat{R} - 1)(\hat{F}_2 - 1)$ degrees of freedom, under the null hypothesis that R and F_2 are independent at fixed values of F_1 . From the additivity property of the chi-square distribution we find (assuming the number of observed responses \hat{R} and feature values \hat{F}_2 are the same at each fixed f_1 , see below) that $2N \ln(2) I(R; F_2|F_1)$ is chi-square distributed with $\hat{F}_1(\hat{R} - 1)(\hat{F}_2 - 1)$ degrees of freedom.

It is also possible to evaluate the statistical significance of the conditional mutual information using the bootstrap approach as described in Section 2. However, care must be taken when implementing the shuffling procedure to ensure it samples from the required null hypothesis. Since the null hypothesis is that feature F_2 and response R are independent, at fixed values of F_1 , samples should be drawn by shuffling the combination of F_2 and R at fixed F_1 . The simpler procedure of shuffling F_2 values without considering the corresponding F_1 value would indeed be incorrect (and we verified that it is also less accurate, results not shown) because it removes all the information in F_2 , and not only that additional to F_1 .

When the response features are highly correlated or when the neural response is selective only to a specific combination of features, then the number of relevant response bins for both features may be smaller than the product of the number of relevant bins for each feature. The random permutation of F_2 without considering the value of F_1 would then make the joint feature space at fixed response larger than in the real data. As a consequence the null hypothesis distribution will have a larger bias than the true one (Panzeri et al., 2007), and also a larger variance. This would make the evaluation of the required null hypothesis problematic. Indeed, in such cases the above expression for the degrees of freedom of the chi-square approximation should be modified. Instead of taking $\hat{F}_1(\hat{R} - 1)(\hat{F}_2 - 1)$, the values of $(\hat{R} - 1)(\hat{F}_2 - 1)$ observed for each value f_1 should be summed to obtain the total degrees of freedom.

4.2. Numerical investigation of the performance of statistical tests for information complementary

To investigate the statistical power of these tests, we performed simulations with a system similar to that described in Section 2, but extended to include two correlated features. Again, the model is based as closely as possible on the statistics of the experimental data considered in Section 5, and it matches the first and second order statistics of V1 gamma BLP LFP responses to variations in spatial contrast and temporal contrast during presentation of Hollywood color movies. The correlations in the simulated features also matched those observed in the movies used in the experiments. As in Section 2, the simulated gamma BLP response was binned in $R=4$ equi-populated levels, and each simulated feature was binned into $\hat{F}_i = 8$ equi-populated classes (see Appendix A for full details).

It is intuitive that the sensitivity of any test for complementarity of information must depend on the actual amount of additional genuine complementary information of the second feature (the larger it is, the easier it is to detect). Thus in our simulation the response dependence on F_1 (Michelson contrast) was fixed in all simulations, but we considered two different cases with regard to the second correlated feature F_2 (temporal contrast): in the first simulation it carried an amount of additional information to F_1 , and in a second simulation it did not carry any additional information, although it was informative per se due to its correlation to F_1 (see Appendix A).

We tested this bootstrap procedure in Fig. 4. We applied this procedure to uncorrected plugin estimates of information because, for the reasons explained in Section 2, these uncorrected information estimates are better at determining significance. Fig. 4A shows how the percentage of simulation realizations where

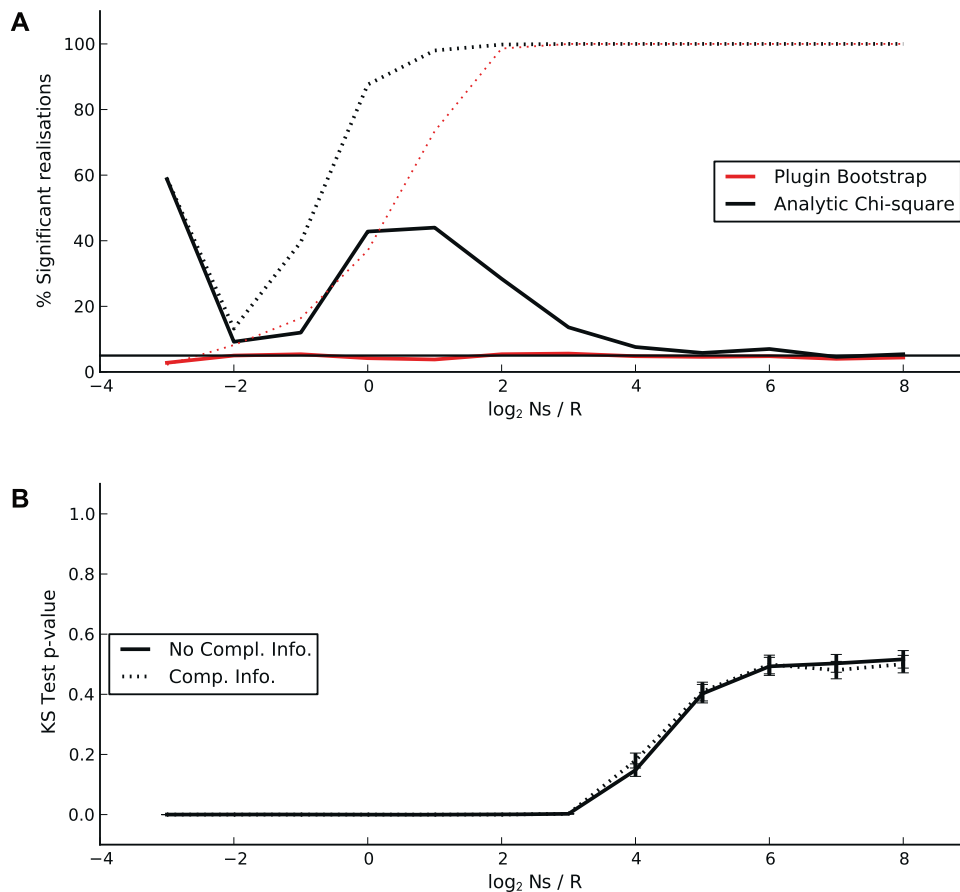


Fig. 4. Determining the significance of conditional mutual information. Model gamma responses to eight different values of Michelson contrast and eight different values of temporal contrast were simulated (see text for details). The number of trials generated per stimulus was varied; for each number of trials 500 realizations of the simulation were generated and the calculations performed on each realization. The x-axes show the N_s/R ratio where N_s is the average number of trials for each possible stimulus feature pair ($R = 4$). Panel (A) shows the percentage of realizations where the CMI between response and temporal contrast given Michelson contrast was found to be significant using the uncorrected bootstrap (red lines) and analytic chi-square (black lines) methods. Dotted lines show the model with independent representation of temporal contrast; solid lines show the model with no independent representation of temporal contrast. Panel (B) shows the mean (errorbars show standard deviation) over 100 realizations of the p -value from the Kolmogorov–Smirnov test of the goodness of fit of the 100 bootstrap samples to the analytic chi-square distribution.

the CMI was correctly found to be significant (i.e. percentage of true positives) depended upon the N_s/R ratio. Note that in this case N_s is the average number of trials for each possible stimulus feature pair, and R is the number of responses. The bootstrap test was able to reliably detect the presence of significant complementary information for systems with $N_s/R > 4$ (16 trials for each pair of stimulus feature values, 128 trials per F_1 stimulus feature value).

Again, the chi-square approximation was accurate provided the sampling was sufficient. For the single feature system shown in Fig. 1, the chi-square approximation started to fail at $N_s/R = 2^3$. For this conditional system $N_s/R = 2^5$ is required. This is shown by the increased false positive rate for the analytic chi-square approximation when no genuine complementary information is present (Fig. 4A), and by the low p -values for the KS-test between the chi-square approximation and the bootstrap distribution for both simulated systems (Fig. 4B). The fact that more trials are required here for the analytic chi-square to hold is likely due to the strong correlations between features changing the statistics of the F_1 -conditional distributions; the f_2 values at fixed f_1 are not necessarily uniformly distributed as they were in the single feature model. We verified this fact by demonstrating that in simulations in which the correlation between F_1 and F_2 was set to zero, the chi-square approximation was effective even with $N_s/R = 2^3$, as in the individual feature case (results not shown).

In summary, our analysis of the effectiveness of tests of significance of complementarity of information revealed that: (i) a

bootstrap test based upon computing uncorrected plugin information estimates after shuffling a features conditional to the value of other(s) is extremely effective at detecting significance with small amounts of both false positives and negatives (ii) The bootstrap test can be replaced by an analytical chi-square test provided sampling is sufficient ($N_s/R > 32$ in our simulation). The regime in which the analytic chi-square is accurate can be determined in real data by using the KS test between empirical bootstrap and analytic chi-square distributions for sub-samples of different sizes.

5. Application of the formalism to the encoding of visual features by gamma BLP in V1

To illustrate the feasibility of the aforementioned formalism on neurophysiological data, we used it to analyze LFPs recorded in V1 in response to repeated binocular presentation of a color movie. This is a useful application of our tests for at least two reasons.

The first reason is that the precise origin, meaning and stimulus tuning of local field potentials is not yet fully known and is the subject of continuing studies (Belitski et al., 2008; Berens et al., 2008; Jia et al., 2011; Katzner et al., 2009; Kayser and Konig, 2004; Mazzoni et al., 2008). Because of this, there is a pressing need for tools, which are able to correctly identify and quantify the tuning of these signals to complex stimuli.

The second reason is that these data collect responses to complex naturalistic stimuli, which contain many visual features that co-vary over time (Simoncelli, 2003); for instance moving objects may enter or leave the RF or be occluded by other fast moving objects, or local illumination may change dynamically, both causing a large number of features to change in a correlated fashion (Bartels et al., 2008). Evidently such correlated changes may pose problems to studies examining the selectivity of neuronal populations to individual features of complex visual images. During the presentation of simplified, artificial visual stimuli, the problem can be readily solved by varying different visual attributes in an uncorrelated manner (Brenner et al., 2000a; de Ruyter van Steveninck and Bialek, 1988; Petersen et al., 2008; Rust et al., 2005; Touryan et al., 2005; Yamada and Lewis, 1999). In such designs, demonstration of neuronal feature-selectivity simply follows the selective stimulus parameterization. Yet, such an uncorrelated stimulus design cannot be applied when using naturalistic stimuli such as movies or sounds capturing the complexity and characteristics of a natural environment. Analysing neural responses to naturalistic stimuli is important for at least two reasons. First, naturalistic stimuli are likely to engage complex patterns of activity in cortical microcircuits that may not be elicited in the presence of simple stimuli optimized for the study of basic properties of single neurons (Felsen and Dan, 2005; Reinagel, 2001). Second, naturalistic visual stimuli have been shown to evoke more reliable responses (Hasson et al., 2009), and more specific inter-regional correlations (Bartels and Zeki, 2005), suggesting that coding in the visual system may be optimized for processing naturalistic stimuli (Dan et al., 1996). Here we probe the ability of our methods to quantify the statistical significance of information carried by LFPs about a given visual feature above and beyond that carried by other, potentially co-varying visual features.

5.1. Experimental procedures

Before proceeding with the analysis, we briefly summarize the experimental procedures used to record neural responses to naturalistic color movies in V1. A detailed description can be found in previous studies (Belitski et al., 2008, 2010; Montemurro et al., 2008). All procedures were approved by the local authorities (Regierungspräsidium) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals. We recorded neural activity with an array of extracellular electrodes from the opercular (foveal and para-foveal field-representations) portion of V1, of anesthetized macaques presented with commercially available color movie clips consisting of scenes from *The Last Samurai* (2003, Warner Bros. Pictures) and *Star Wars: Episode I – The Phantom Menace* (1999, Lucasfilm). We analyzed a total of 55 recording sites from 3 different monkeys. Movie clips lasted 150–210 s and were presented binocularly at a resolution of 640×480 pixels (field of view: $30^\circ \times 23^\circ$, 24 bit true color, 60 Hz refresh) using a fiber-optic system (Avotec, Silent Vision, Florida). Each movie was presented 35–50 times per session in order to adequately sample the probability distribution of the neural responses to each part of the movie.

The extracellular field potential was filtered to extract Multi Unit Activity (MUA) and local field potentials using standard signal conditioning techniques as described in detail elsewhere (Belitski et al., 2008). The MUA was used only to determine the receptive fields, as described below. The gamma band (30–100 Hz) of the LFPs was extracted by means of zero-phase shift, bidirectional filters (Kaiser filter with a transition band of 1 Hz, stop-band attenuation of 60 dB, and pass-band ripple of 0.01 dB). The instantaneous power of this band-limited signal, shortened in the following as gamma band limited power, was computed as the squared modulus of the discrete-time analytic signal obtained via the Hilbert transform.

We focus here on the gamma BLP as a neural response since it has been previously shown to be informative about naturalistic movie stimulation (Belitski et al., 2008) and is known to be modulated by visual features such as spatial contrast (Henrie and Shapley, 2005) and orientation (Kayser and Konig, 2004).

5.2. Extraction of visual features

We first estimated the aggregate receptive field (RF) of each recording site by using the reverse correlation technique (Ringach and Shapley, 2004), which measures the sensitivity of the MUA response to the time contrast of each pixel of the screen. When using reverse correlation with spatially correlated stimuli (such as the natural movies used here) the obtained RF is likely biased by the stimulus correlation and is therefore likely to be larger than the true RF (Theunissen et al., 2001; Touryan et al., 2005). For this reason, the RF chosen for the feature extraction analysis was fixed as a $1^\circ \times 1^\circ$ region around the peak of the neural sensitivity, even when the result of reverse correlation resulted in a larger map. This size of RF was chosen because it is typical of multiple unit V1 receptive fields (Sceniak et al., 1999). An example of such RF estimation is reported in Fig. 5A. Reverse correlation was used here to obtain an estimate of the location of the RF area and not to estimate its size. To verify that the precise size set for the RF was not critical for the information results, we repeated the analysis selecting RFs sizes of $0.5^\circ \times 0.5^\circ$ and $1.5^\circ \times 1.5^\circ$. We found that the features extracted from RFs of these larger or smaller sizes were essentially identical to those obtained with the $1^\circ \times 1^\circ$ RF (correlation between the time courses of features extracted with different RF sizes was >0.95 for all electrodes and recording sessions, and information values were not significantly different, $p > 0.5$). As an additional check that the location of the RF was correctly determined, we compared the RF positions obtained by reverse correlation with those manually mapped by the experimenter with small polar stimuli during the experimental session, and we found excellent agreement (not shown) for all sites.

We subsequently computed three different types of image features within each RF: spatial contrast (Michelson contrast, calculated as the difference between the maximal and minimal luminance divided by the average between maximal and minimal luminance), orientation (computed as the direction of gradient of contrast (Kayser et al., 2003)), and a measure of time contrast, defined as the total amount of pixel-wise luminance change from frame to frame, that is, the absolute value of the luminance difference between two successive frames, calculated for every pixel and then averaged over the RF (Bartels et al., 2008).

5.3. Calculation of information conveyed about different features

For each recording site, we computed the information carried by the gamma BLP about the current value of the movie features in the corresponding RF. For this calculation, the stimulus set was created as follows (see also Reich et al. (2000)). We first divided the movie presentation into non-overlapping adjacent windows each 10 frames long. This window length was chosen because it matched the timescale of the feature autocorrelations. We averaged the value of each feature over the 10 frames in each stimulus window. Unless otherwise stated, we discretized the resulting feature values across all windows into 8 equally populated classes, and each class was considered as a different “stimulus” value for the information calculation. Gamma BLP values were averaged over the same intervals and subsequently discretized into $R=4$ equally populated bins. We chose these parameters because they allow accurate sampling given the amount of data available. We chose $R=4$ based on the considerations in Section 2.4, which suggest cautious choices of the number of bins, especially considering that the

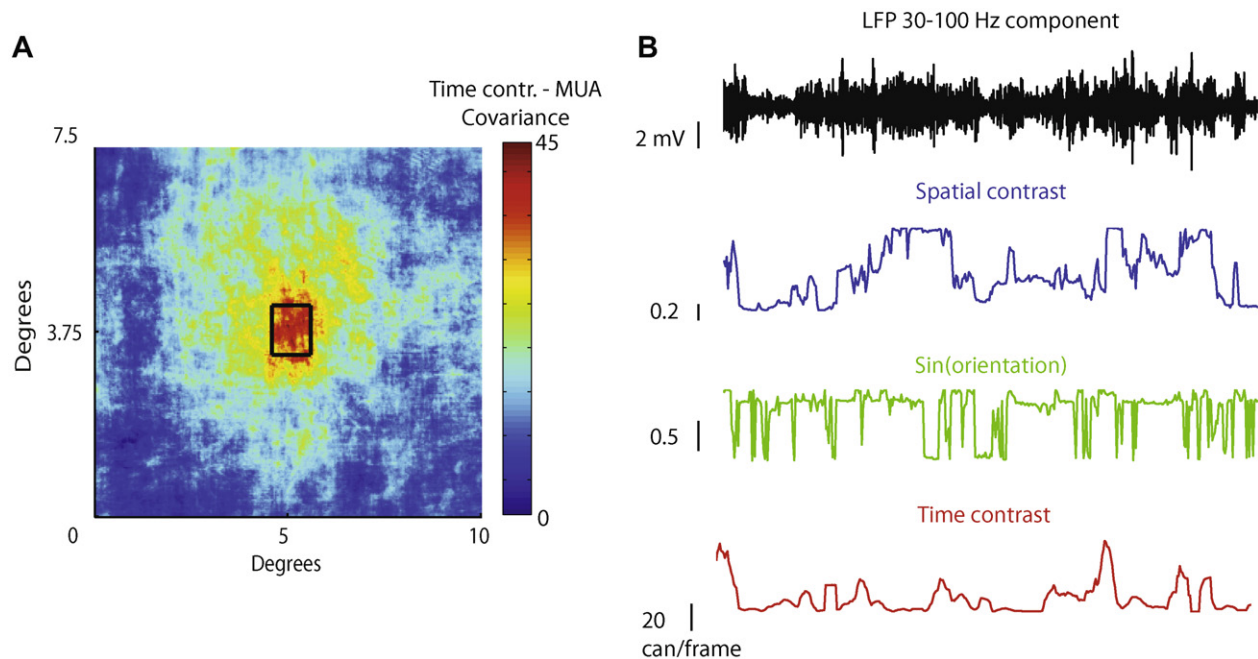


Fig. 5. Example of movie features and neural activity. (A) For each recording site, the receptive field was determined as the 1° square area in the screen where the covariance between the multi-unit activity recorded in the site and the time contrast was highest. The local field potential (LFP) recorded in the site was then studied as a function of the visual features of the movie computed within the associated receptive field. (B) Comparison of the gamma activity recorded during a single trial in a representative recording site (from session D04nm1) with the movie features computed in the associated receptive field. From top to bottom: LFP gamma band (30–100 Hz), spatial contrast, sine of the orientation, time contrast.

sampling demands of conditional mutual information are greater than for the single feature information considered there. We chose a larger number of bins for the stimulus features, since our aim was to characterize the role of correlations between them in evaluating tuning of neurons to multiple features.

We computed different sets of information measurements. For each information value, we first computed the significance of the uncorrected plug-in information estimate with the analytic chi-square approximation described in Section 2.2 for the absolute information and in Section 4.1 for the conditional information. For this analysis we had 20,000–30,000 trials available ($25,500 \pm 3100$, mean \pm std). This corresponds to $N_s/R \sim 800$ for the estimation of information about an individual feature, and $N_s/R \sim 100$ for the estimation of CMI. These numbers are well inside the regimes where the chi-square approximations perform well (Sections 2 and 4). We then computed an estimate of the information, using Eq. (5) for the information between gamma BLP and features, Eq. (6) for the information between features, and Eq. (7) for the conditional information, correcting for the bias with the PT correction method (Panzeri and Treves, 1996) described by Eq. (2). All information measures on experimental data were computed using the Information Breakdown Toolbox² (Magri et al., 2009).

5.4. Results: complementarity coding of orientation, spatial contrast and time contrast by the LFP gamma power

The time course of three extracted movie features (spatial contrast, time contrast and orientation) in the RF (Fig. 5A) on an example recording site is shown in Fig. 5B, together with the associated gamma band of the LFP recorded from that site. From this example, it is apparent that the gamma BLP is tuned, to different extents, to these three features. This is consistent with previous

reports about gamma-band selectivity obtained with simple stimuli (Frien et al., 2000; Henrie and Shapley, 2005).

To quantify these dependencies, we computed the mutual information between the gamma BLP and the different features and we tested the significance of tuning of each of the 55 LFP recording sites. To this end, we used the analytical chi-square information test on the bias uncorrected plugin estimates (because this is the most sensitive test under the sampling conditions analyzed). We found that all recording sites carried significant information at $p < 0.001$ confidence level for all three features (Fig. 6A). While responses were significantly modulated by all stimulus features considered, we found that the population average (computed with PT bias correction) of the information about spatial or temporal contrast was much higher than the information about orientation (Fig. 6B), showing that orientation caused weaker (although still highly significant) response modulation than the other features.

We evaluated the robustness of the significance test by using only a subset of the movie presentations (and henceforth of the trials), similarly to what was done in Fig. 1A. We found (Fig. 6C) that for the most informative features (spatial and temporal contrast) significant response modulation was detected in almost all recording sites (>95%) even when only 6 movie presentations were considered (of the 35–50 available), and using only two movie presentations was sufficient to detect significant information about these two contrast features in more than 50% of recording sites. For the orientation feature, more data was required to find dependence in a sizable fraction of recording sites because of the smaller information. However, even with only two movie presentations, more than 30% of recording sites were found to be significantly modulated by all three features.

As described in Section 3, to understand whether the gamma BLP was truly tuned to all three stimulus features and was not carrying information about some of them only because of correlation to another truly encoded feature, we had to characterize whether or not the different features in the RF vary independently during the movie time course. We found that the visual features did not

² <http://www.ibtb.org>.

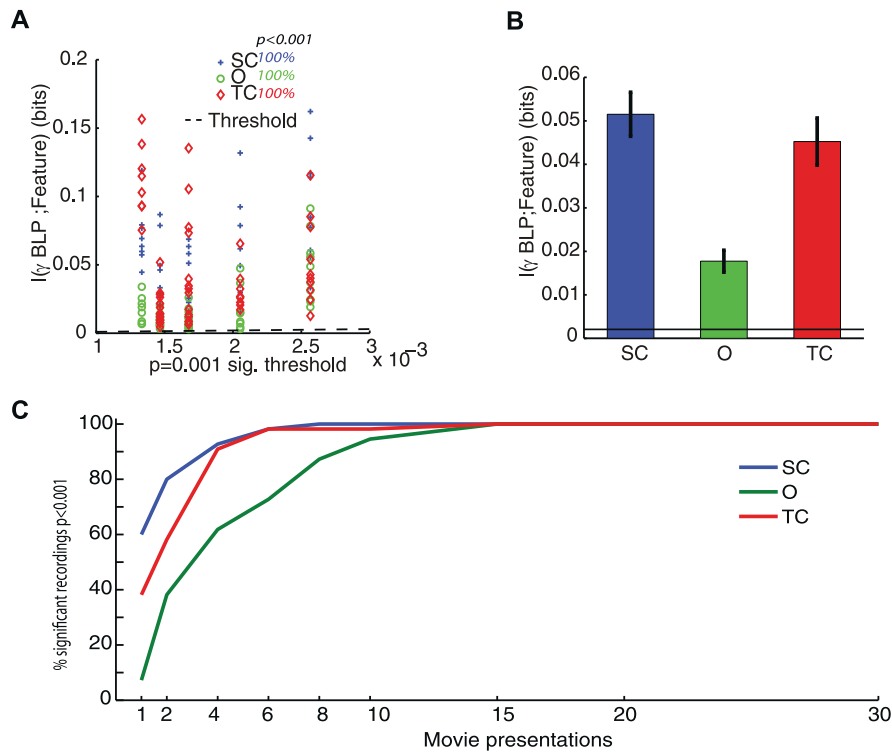


Fig. 6. Information carried by gamma band limited power (BLP) about spatial contrast, orientation and time contrast. (A) Information about the value of the different features carried by gamma BLP for each recording site, computed with 4 classes for gamma BLP, 8 classes for features and no bias correction. Significance was computed with the chi-square approximation, $p < 0.001$. Since the number of classes for gamma BLP and features are fixed, the analytic chi-square significance threshold (in bits) depends only on N , the number of trials available (Section 2.2). Data points are organized into vertical lines since different numbers of trials were produced by different experimental protocols (depending on the duration and number of presentations of the movie); experiments with the same N form each vertical line. Threshold $p = 0.001$ is indicated by dashed line. Numbers in legends indicate the fraction of recordings for which the gamma BLP conveyed significant information. (B) Information about the value of the different features carried by gamma BLP (mean \pm SEM, $n = 55$ recording sites). Black line indicates mean value of significance threshold ($p = 0.95$ of average bootstrap over 400 permutations). (C) Fraction of recordings for which the gamma BLP conveyed significant information (chi-square information test; $p < 0.001$) as a function of the number of movie presentations used for the analysis.

vary independently from one another in the movie section used here: the percentage of recording site RFs with significant information between features at $p < 0.001$ confidence level was 71%, 62% and 73% for spatial contrast with orientation, temporal with spatial

contrast and orientation with temporal contrast respectively (Fig. 7A). Furthermore, the bias corrected population average values of information (Fig. 7B) were higher than those between features and neural response. This, as explained in previous

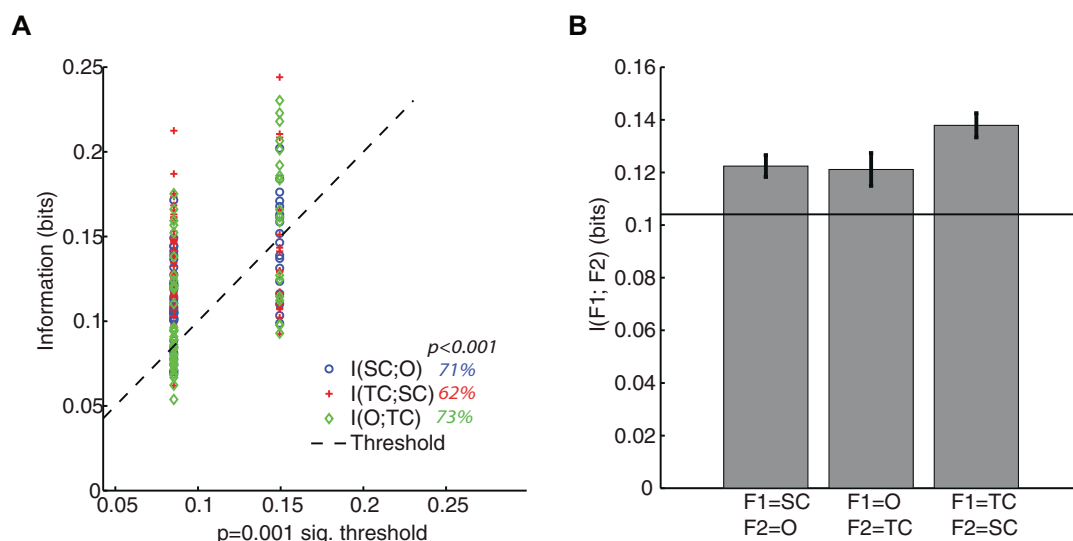


Fig. 7. Mutual information between visual features. (A) Information between features computed for each RF and values averaged over 10 frame windows using 8 feature classes. Significance of information is computed with the chi-square approximation, $p < 0.001$. Threshold $p = 0.001$ is indicated by dashed line. Numbers in italics indicate the fraction of recordings for which the mutual information between the features was significant. (B) Mutual information between features (mean \pm SEM, $n = 55$ RFs) computed with same parameters of (A), and PT bias correction. Black line indicates mean of significance threshold ($p = 0.95$ of population average of bootstrap over 400 permutations). Results show that the considered visual features of the movie section used here are significantly correlated.

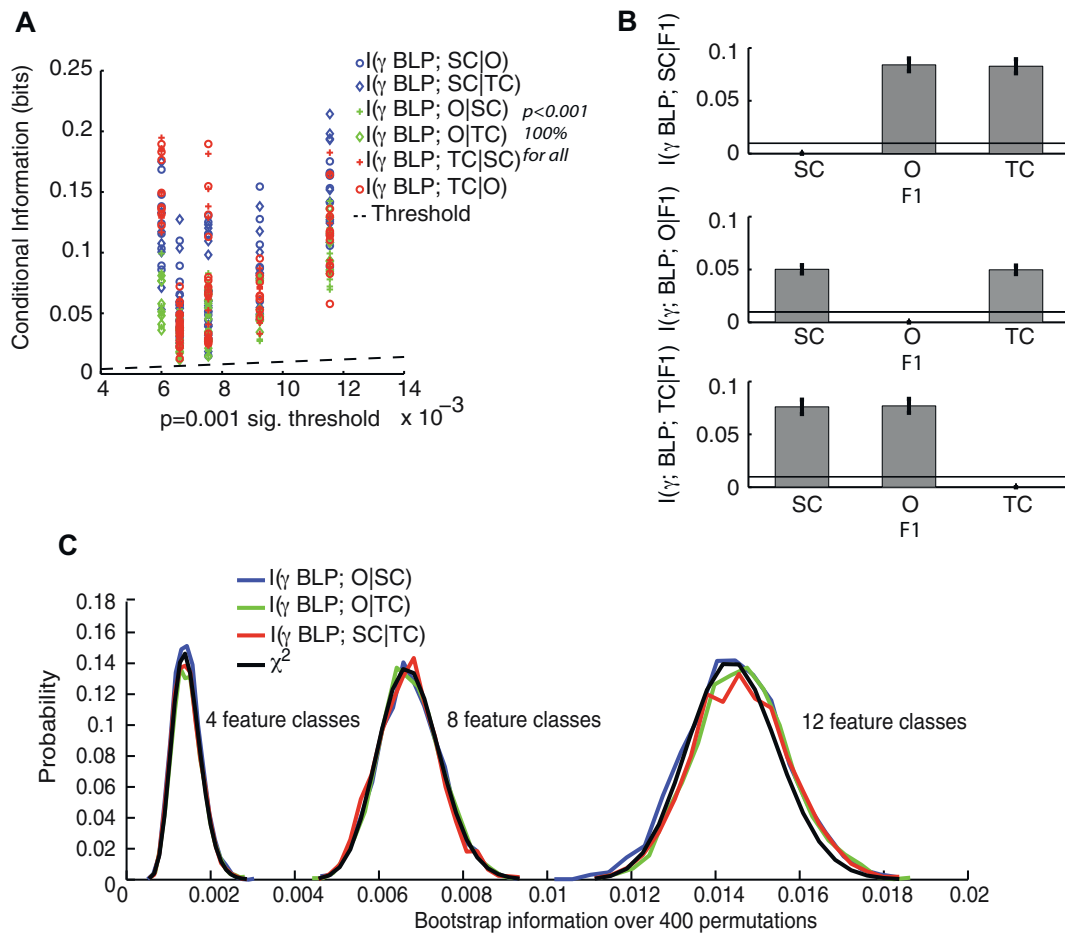


Fig. 8. Conditional information carried by gamma BLP about visual features. (A) Information carried at each recording site by gamma BLP about the value of each feature conditioned on each of the other features, computed with no bias correction. Significance ($p < 0.001$) is assessed with the chi-square approximation, with threshold value of $p = 0.001$ indicated by black dashed line. Numbers in legend report the percentage of recordings for which the gamma BLP conveyed significant CMI. (B) Information carried by gamma BLP about the value of each feature conditioned on each of the other features (mean \pm SEM, $n = 55$). Information is computed with PT bias correction. Black line indicates mean of significance threshold ($p = 0.95$ of average bootstrap, 400 permutations). (C) Distribution of bootstrap values of uncorrected CMI for different pairs of features (colored lines, 400 random permutations) compared with the corresponding chi-square distribution, for different numbers of feature classes. Results are averaged over all recordings from session H06nm6.

sections, implied that the single-feature information analysis was not enough to establish if all three considered features were truly encoded by the gamma BLP.

To shed light on this issue, we computed, for each pair of features, the conditional information $I(\gamma \text{ BLP}; F_2|F_1)$ carried by the gamma BLP about feature F_2 given the value of feature F_1 . We estimated the significance of this value using the chi-square approximation for the CMI described in Section 4. We found that for every pair of features F_1 and F_2 the CMI was significant at the $p < 0.001$ level for all recording sites (Fig. 8A). We then computed population averages of PT bias corrected CMI values for each pair of features (Fig. 8B). Again, orientation (Fig. 8B, middle row) had weaker conditional modulation than the other features, although it remained statistically significant when conditioned on either remaining feature. Interestingly, the values of conditional information were higher in all cases than the corresponding unconditioned single-feature information values, indicating that correlations between features in the natural movie actually increased (rather than decreased) the joint information that the gamma BLP carried about these multiple features.

The origin of this effect was further investigated by computing $I(F_1; F_2|\gamma \text{ BLP} = 1, 2, 3, 4)$, i.e. by analysing how the mutual information between the feature pairs was related to the different levels of LFP gamma band power. Fig. 9 shows that the mutual information

between feature pairs at fixed response was significant for all levels of gamma BLP (chi-square approximation test on uncorrected information; $p < 0.001$), and was also significantly varying among different levels of gamma BLP (ANOVA test; $p < 0.001$). The fact that the degree of dependence between features differs across sections of stimuli that evoke different responses explains why correlations between features increase information. In this case, the degree of dependence between features can be used to predict the neural response and so it increases the overall information between features and responses (see Panzeri et al. (1999) or Pola et al. (2003) for a proof). Moreover, the average over all four levels of gamma BLP of $I(F_1; F_2|\gamma \text{ BLP} = 1, 2, 3, 4)$ was higher overall than $I(F_1; F_2)$ for all feature pairs, the average gain being +26%, +26% and +22% for $I(\text{SC}; \text{O})$, $I(\text{O}; \text{TC})$ and $I(\text{SC}; \text{TC})$ respectively. In this case, according to Eq. (13) statistical dependencies among features must increase information.

We then evaluated if a given feature carried information that was complementary to that carried by both other features, by calculating $I(\gamma \text{ BLP}; F_2|F_1, F_3)$ using 4 classes for the features. We found that this was also significant (chi-square approximation, $p < 0.001$) for all permutations of features. The results of the above computations show that the gamma BLP conveyed genuine information about each single feature, even though they were strongly correlated.

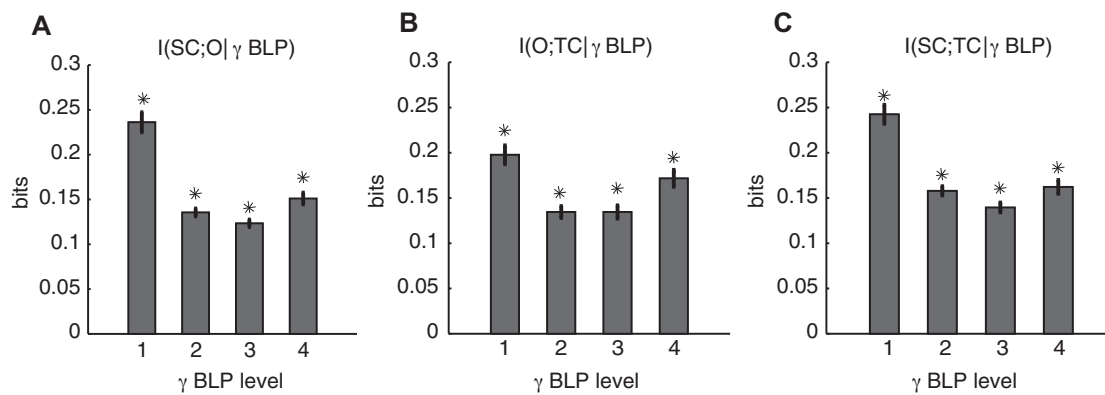


Fig. 9. Information between visual features at fixed response. Mutual information between features conditioned to four different levels of gamma BLP (mean \pm SEM, $n = 55$). (A) Spatial contrast and orientation, (B) orientation and time contrast, (C) spatial contrast and time contrast. Information is computed for values averaged over 10 frame windows with PT bias correction and 8 feature classes. Asterisks indicate that the mutual information is significant for each level of response (chi-square approximation test on uncorrected information, $p < 0.001$). ANOVA test ($p < 0.001$) shows that the information between feature pairs changes significantly for different levels of the response.

We performed all the significance tests using the analytic chi-square approximation on the uncorrected plugin information values because they allow fast computation of highly significant p -values, and because the considerations of Section 3 predicted this test to be highly effective under the sampling conditions described here. To illustrate the accuracy of the chi-square approximation, we compared the distribution of the bootstrap values (over 400 permutations) of the uncorrected plugin information estimates with the analytical chi-square distribution with the degrees of freedom predicted by the equations in Section 4.1. As shown in Fig. 8C for a representative session, with this dataset there was an excellent agreement between the bootstrap and the chi-square distribution. In particular, when changing the number of bins in to which the features were discretized, the bootstrap distribution followed the change in the number of degrees of freedom of the chi-square distribution predicted in Section 3: when the features were binned with 4, 8 and 12 classes the average p -values of the KS test were 0.49 ± 0.025 , 0.36 ± 0.025 , 0.34 ± 0.015 respectively (mean \pm SEM over all recording sites). These results both highlight the value of the analytical approximations reported in our study and demonstrate that the chi-square could be correctly used in this dataset to accurately compute significance values down to small p -values.

6. Discussion

We have considered the application of mutual information as a statistical test to establish dependency between stimulus features and neuronal response variables. While accurate bias free information values are crucial to enable meaningful quantitative comparisons, in many applications it is sufficient to establish the presence of a statistically significant dependence. We have shown that for determining significance with the bootstrap method, uncorrected plugin information estimates should be used, and that if sufficient data are available, the bootstrap procedure can be bypassed completely and significance determined through an analytical form for the distribution of values under the null hypothesis. This allows accurate calculation of p -values and significance with greatly reduced computational requirements.

This finding, particularly the accuracy of the analytic chi-square approximation, has implications for analysis of other neurological signals, for example functional magnetic resonance imaging (fMRI) data. An approach often taken with such data is a mass-univariate analysis to determine dependencies between various stimulation conditions and the time course of the blood oxygen level dependent (BOLD) signal at each voxel. The ability to directly calculate a significance p -value for any dependence cheaply using the chi-square approximation suggests that the methods discussed here

could be easily applied to such data, even when they consist of tens or even hundreds of thousands of voxels. The advantages of such an approach is that they are completely non-parametric and do not require any convolution of the stimulus time course with a hemodynamic response function (HRF). This means they can be applied to situations where the HRF is not well characterized or is not guaranteed to be universal across conditions and areas. This may for instance be the case when investigating the relationship between the fMRI BOLD response and the power of EEG or LFP responses in different frequency bands, or when collecting fMRI responses to complicated dynamic time varying stimuli with multiple correlated features such as the movie stimuli considered here. Additionally, the general and non-linear nature of mutual information allows it to capture any type of dependency, in contrast to conventional linear analysis. Similar considerations can be applied to the problem of estimating significance of activation in other types of neuroimaging signals, such as the sources reconstructed from EEG or MEG data (Gross et al., 2001).

While we have shown that the use of bias corrections has no beneficial effect on establishing the presence of a statistically significant stimulus response modulation, once that significance has been determined, bias corrected values are of course crucial to obtain an accurate quantification of the strength of the modulation. The bias corrected values allow direct quantitative comparisons of the amounts of the information carried by different systems or experiments with different numbers of trials, and of the amounts of information carried by neural codes characterized by a different response dimensionality (such as for example, spike times and spike counts). Our findings suggest that it may be worth implementing a hybrid approach, where uncorrected information values are used (according to the procedures described here) to determine the statistical significance of any dependence effect, while bias corrected values are computed to produce an accurate estimate of the effect size. In practice, since most bias correction techniques involve calculation of the uncorrected values this approach carries little computational overhead. However, care must be taken to avoid statistical errors due to 'non-independent' analysis (Vul et al., 2009), which arise when reporting mean values where the mean is calculated only from samples exceeding some threshold. For this reason, in the analysis of real data presented here we reported bias corrected mean information values over the whole population rather than the information values averaged over the subpopulation with significant information. The accurate determination of the statistical significance can of course be used as a selection tool for other types of independent analysis, for example calculation of receptive fields of significantly tuned neurons.

We considered the problem of characterizing the neural encoding of features of naturalistic stimuli given their unavoidable (and natural) covariance. We tackled this problem in the context of information theory, and we proposed a set of quantities and statistical tests to address this issue. We illustrated the feasibility of the method with simulations and analysis of real neural responses to natural stimuli, and found that our method was successful in establishing (despite the presence of strong correlations between features) that gamma power is genuinely tuned to all features inspected, i.e. space and time contrast as well as orientation. A number of previous studies have shown that the gamma BLP in V1 is sensitive to both contrast and orientation separately (Friedman-Hill et al., 2000; Frien et al., 2000; Henrie and Shapley, 2005; Kayser and Konig, 2004), but to our knowledge this is the first study investigating the effect of correlated features on the gamma BLP response. As discussed above, since visual features can be strongly correlated in naturalistic stimuli as shown in Fig. 7 and discussed in Simoncelli (2003) it was necessary to verify that the gamma power was conveying complementary information about each feature in a naturalistic environment. In particular in our movie clips the strength of the orientation inside each receptive field was much lower than in experiments utilizing gratings or bars, and the orientation was highly correlated to the other features. However, the gamma BLP was found to be genuinely modulated by variations in the image orientation, even when the correlated spatial and temporal contrast features were fixed. In fact, we have shown that the CMI is in all cases higher than the individual feature information, suggesting that the features are coded in a highly synergistic fashion. More detailed investigation of how this synergistic coding takes place, and the effect of different levels of stimulus feature correlations is an important topic for future research.

We applied our method to the determination of the genuine tuning of neural responses to multiple stimulus features. However, it is important to note that, due to the symmetry of mutual information, the same approach developed here can be readily applied to the problem (currently intensively investigated, see Panzeri et al. (2010) and Schyns et al. (2011)) of determining if a given neural code (e.g. the neural response in a given frequency band) carries information about stimuli which is truly complementary to that carried by another type of simultaneously observed neural response (e.g. the power in another band).

The major limitation of the present approach arises from the fact that information theory is substantially more data intensive than other analysis methods, because it requires the estimation of the full stimulus-response probability distribution (Brown et al., 2004). It is thus difficult to apply this approach to very large number of features. In such cases, the information theoretic approach is feasible only if stimulus-response relationships can be well approximated by a simple parametric model (e.g. by assuming Gaussian stimulus conditional response distributions, which have been shown to be a useful model for LFP power responses, see Magri et al. (2009)), so that many features could be considered with the amount of data typically recorded from a neurophysiological experiment. If no model of stimulus-response probabilities is available, then other approaches should be considered, such as those based on linear or nonlinear spatio-temporal receptive field (STRF) estimation and boosting algorithms seeking to reduce the impact of correlation by sparsification of stimulus response characteristics (Theunissen et al., 2001, 2000; Willmore et al., 2010)

An advantage of the method considered here stems from the fact that mutual information, unlike other forms of single trial analysis such as stimulus decoding techniques, captures the effect of all possible relationships between stimuli and neural responses and quantifies the overall knowledge about the stimulus gained with the considered single-trial neuronal responses. For example, evaluating single-trial stimulus selectivity by simply decoding the most

likely stimulus that elicited the neural response in each trial may neglect some type of information carried by the neural response, such as the information that the neuron provides about the identity of other stimuli that are very unlikely given the observed neural response. Alternatively, evaluating stimulus selectivity with a linear decoder may fail to capture the knowledge gained by nonlinear stimulus-response interactions. Since the mutual information quantifies the effect of all possible relationships between feature and response, demonstrating that a neuron carries significant extra mutual information about a second feature means that the neuronal response has selectivity for the second feature that cannot possibly be explained from selectivity to the first feature. A similarly strong conclusion could not be reached, for example, by showing that a linear regression based on two stimulus features explains the neural response better than either feature alone. In such case it is still possible that the predictability of the neural response gained by adding the second feature reflects some response predictability explainable by a nonlinear relationship between the response and the first feature, which could not be captured by the linear model. For these reasons, the information theoretical approach discussed here does allow strong conclusions when it is applicable.

Acknowledgements

We thank the anonymous referees for thoughtful and detailed critiques, which significantly improved the revised version of the manuscript. We are grateful to C. Magri for useful discussions, to C. Kayser for sharing his code for orientation computation, and to members of the EPSRC-funded COLAMN network for useful interactions. This work was supported by the BMI project of IIT, by the Max Planck Society, by the Compagnia di San Paolo, and was part of the research program of the Bernstein Center for Computational Neuroscience, Tübingen, funded by the German Federal Ministry of Education and Research (BMBF; FKZ: 01GQ1002).

Appendix A.

This Appendix A describes the simulated processes used in Figs. 1–4. All information measures on simulated data were computed in the Python programming language using the pyEntropy library³ (Ince et al., 2009).

A.1. Simulation 1: single feature gamma BLP LFP responses

The gamma BLP LFP response of a single recording site (session D04nm1, channel 3) in response to a natural movie was chosen from the data described in Section 5. The Michelson contrast of the movie was quantized into 8 equally occupied levels, and the neural response to each stimulus was quantified by the mean and standard deviation of the cube root of the gamma BLP (see main text). These means and standard deviations were manipulated according to Eq. (4) in order to modulate the information conveyed by the system resulting in three different model systems, with parameters $\alpha = 1$ (full information), $\alpha = 0.5$ and $\alpha = 0$ (no information). The conditional distributions for each system are visualized in Fig. A1. This shows the structure of the system as determined by the experimental data; the response is modulated most strongly for stimulus categories 0 and 6. Category 0 consists of the lowest values of Michelson contrast and modulates the response most strongly; when contrasts in this group are presented a lower response is more likely to be observed. Category 6, the second highest category of contrast values also modulates the response strongly but

³ <http://code.google.com/p/pyentropy/>.

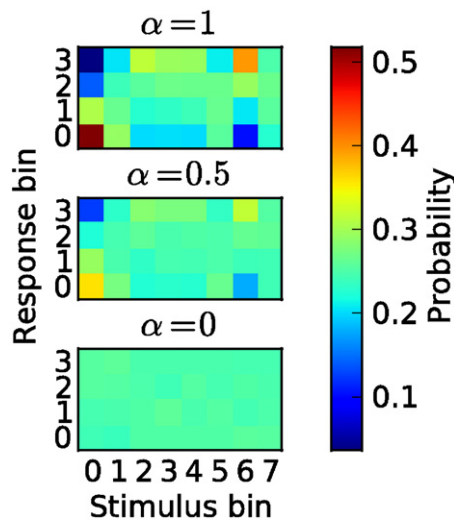


Fig. A1. Conditional distributions of simulated data. The conditional probabilities $P(r|s)$ are shown for each possible response bin r and stimulus bin s for the simulated Gaussian systems, generated as described in the text (Eq. (4) and Appendix A). Probabilities were estimated from a generated data set with 8192 trials per stimulus value (65,536 trials in total).

in the opposite direction; when these contrasts are presented and higher response is more likely to be observed. For each model system 100 realizations were generated with a fixed number of trials per stimulus. For each of these realizations the information was calculated, and stimulus and response were shuffled 100 times and the calculation repeated to obtain the bootstrap distribution of information under the null hypothesis of independence. The true value of each realization was compared to the 95th percentile of the bootstrap distribution, and to the corresponding $p=0.05$ value of the analytical chi-square distribution to determine significance.

A.2. Simulation 2: multiple feature gamma BLP LFP responses

The gamma BLP response of a single recording site (session D04nm1, channel 3) in response to a natural movie was chosen from the data described in Section 5. The Michelson spatial contrast (SC) and temporal contrast (TC) of the movie (see main text) were quantized into 8 equally occupied levels, and the neural response to each stimulus was quantified by the mean and standard deviation of the cube root of the gamma BLP (see main text). In addition the conditional distribution, $P(\text{TC}|\text{SC})$, of the quantized feature values was determined for each of the 8 SC values using the maximum likelihood (histogram) method. To generate the stimulus values for the simulation, uniform integers were generated to represent the quantized SC feature (due to the equi-populated binning the feature symbols were uniformly distributed over the course of the movie). Then TC features were generated at each time point from the appropriate conditional feature distribution. In the simulation with no genuine complementary information (Fig. 4A and B, solid lines), the response was drawn from a Gaussian distribution with the parameters found from the data for the appropriate TC contrast value (the TC single feature conditional Gaussian model). In the simulation with genuine complementary information about SC (Fig. 4A and B, dotted lines), a value was drawn from the appropriate SC single feature conditional Gaussian model and a separate value was drawn from the appropriate TC single feature conditional Gaussian model. The response was then taken as the sum of these two values. In both simulations the continuous responses were quantized into 4 equally populated bins. For each model system 500 realizations were generated with a fixed number of trials for each SC value. For each of these realizations the CMI was calculated, and

then the TC values were shuffled at fixed SC 100 times and the calculation repeated to obtain the bootstrap distribution of CMI under the null hypothesis of independence of response and TC at fixed SC. The true value of each realization was compared to the 95th percentile of the bootstrap distribution, and to the corresponding $p=0.05$ value of the analytical chi-square distribution to determine significance. The degrees of freedom for the chi-square were calculated as the sum of the degrees of freedom of each specific SC value conditional information calculation based on the number of observed response and TC values for the specific SC value.

References

Adelman TL, Bialek W, Olberg RM. The information content of receptive fields. *Neuron* 2003;40:823–33.

Barlow HB. Unsupervised learning. *Neural Comput* 1989;1:295–311.

Bartels A, Zeki S. Brain dynamics during natural viewing conditions—a new guide for mapping connectivity in vivo. *Neuroimage* 2005;24:339–49.

Bartels A, Zeki S, Logothetis NK. Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cereb Cortex* 2008;18:705–17.

Belitski A, Grettton A, Magri C, Murayama Y, Montemurro MA, Logothetis NK, et al. Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. *J Neurosci* 2008;28:5696–709.

Belitski A, Panzeri S, Magri C, Logothetis NK, Kayser C. Sensory information in local field potentials and spikes from visual and auditory cortices: time scales and frequency bands. *J Comput Neurosci* 2010;29:533–45.

Berens P, Keliris GA, Ecker AS, Logothetis NK, Tolias AS. Comparing the feature selectivity of the gamma-band of the local field potential and the underlying spiking activity in primate visual cortex. *Front Syst Neurosci* 2008;2:2.

Borst A, Theunissen FE. Information theory and neural coding. *Nat Neurosci* 1999;2:947–57.

Brenner N, Bialek W, de Ruyter van Steveninck R. Adaptive rescaling maximizes information transmission. *Neuron* 2000a;26:695–702.

Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR. Synergy in a neural code. *Neural Comput* 2000b;12:1531–52.

Brown EN, Kass RE, Mitra PP. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat Neurosci* 2004;7:456–61.

Cochran WG. Some methods for strengthening the common chi-squared tests. *Biometrics* 1954;10:417–51.

Cover TM, Thomas JA. Elements of information theory. New York: Wiley; 1991.

Dan Y, Atick JJ, Reid RC. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J Neurosci* 1996;16:3351–62.

de Ruyter van Steveninck R, Bialek W. Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc R Soc Lond B: Biol Sci* 1988;234:379–414.

de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W. Reproducibility and variability in neural spike trains. *Science* 1997;275:1805–8.

Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR. Efficiency and ambiguity in an adaptive neural code. *Nature* 2001;412:787–92.

Fan J, Hung HN, Wong WH. Geometric understanding of likelihood ratio statistics. *J Am Stat Assoc* 2000;95:836–41.

Felsen G, Dan Y. A natural approach to studying vision. *Nat Neurosci* 2005;8:1643–6.

Friedman-Hill S, Maldonado PE, Gray CM. Dynamics of striate cortical activity in the alert macaque: I. Incidence and stimulus-dependence of gamma-band neuronal oscillations. *Cereb Cortex* 2000;10:1105–16.

Frien A, Eckhorn R, Bauer R, Woelbern T, Gabriel A. Fast oscillations display sharper orientation tuning than slower components of the same recordings in striate cortex of the awake monkey. *Eur J Neurosci* 2000;12:1453–65.

Geisler WS. Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol* 2008;59:167–92.

Gross J, Kujala J, Hämäläinen M, Timmermann L, Schnitzler A, Salmelin R. Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proc Natl Acad Sci USA* 2001;98:694–9.

Hasson U, Malach R, Heeger DJ. Reliability of cortical activity during natural stimulation. *Trends Cogn Sci* 2010;14:40–8.

Henrie JA, Shapley R. LFP power spectra in V1 cortex: the graded effect of stimulus contrast. *J Neurophysiol* 2005;94:479–90.

Ince RAA, Mazzoni A, Petersen RS, Panzeri S. Open source tools for the information theoretic analysis of neural data. *Front Neurosci* 2010;4.

Ince RAA, Petersen RS, Swan DC, Panzeri S. Python for information theoretic analysis of neural data. *Front Neuroinformatics* 2009;3:4.

Jia X, Smith MA, Kohn A. Stimulus selectivity and spatial coherence of gamma components of the local field potential. *J Neurosci* 2011;31:9390–403.

Katzner S, Nauhaus I, Benucci A, Bonin V, Ringach DL, Carandini M. Local origin of field potentials in visual cortex. *Neuron* 2009;61:35–41.

Kayser C, Einhauser W, König P. Temporal correlations of orientations in natural scenes. *Neurocomputing* 2003;52–54:117–23.

Kayser C, König P. Stimulus locking and feature selectivity prevail in complementary frequency ranges of V1 local field potentials. *Eur J Neurosci* 2004;19:485–9.

Kullback S. Information theory and statistics. New York: Dover Publications; 1968.

- Larntz K. Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *J Am Stat Assoc* 1978;73:253–63.
- Lüdtke N, Panzeri S, Brown M, Broomhead DS, Knowles J, Montemurro MA, et al. Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks. *J Roy Soc Interface* 2008;5:223–35.
- Magri C, Whittingstall K, Singh V, Logothetis NK, Panzeri S. A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. *BMC Neurosci* 2009;10:81.
- Mazzoni A, Panzeri S, Logothetis NK, Brunel N. Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. *PLoS Comput Biol* 2008;4:e1000239.
- Miller GA. Note on the bias of information estimates. *Information theory in psychology: problems and methods*, vol. 2; 1955. p. 95–100.
- Montani F, Kohn A, Smith MA, Schultz SR. The role of correlations in direction and contrast coding in the primary visual cortex. *J Neurosci* 2007;27:2338–48.
- Montemurro MA, Rasch MJ, Murayama Y, Logothetis NK, Panzeri S. Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Curr Biol* 2008;18:375–80.
- Nemenman I, Bialek W, de Ruyter van Steveninck R. Entropy and information in neural spike trains: progress on the sampling problem. *Phys Rev E* 2004;69:56111.
- Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 1996;381:607–9.
- Optican LM, Gawne TJ, Richmond BJ, Joseph PJ. Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. *Biol Cybern* 1991;65:305–10.
- Optican LM, Richmond BJ. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* 1987;57:162–78.
- Paninski L. Estimation of entropy and mutual information. *Neural Comput* 2003;15:1191–253.
- Panzeri S, Brunel N, Logothetis NK, Kayser C. Sensory neural codes using multiplexed temporal scales. *Trends Neurosci* 2010;33:111–20.
- Panzeri S, Petroni F, Petersen RS, Diamond ME. Decoding neuronal population activity in rat somatosensory cortex: role of columnar organization. *Cereb Cortex* 2003;13:45–52.
- Panzeri S, Schultz SR, Treves A, Rolls ET. Correlations and the encoding of information in the nervous system. *Proc R Soc B: Biol Sci* 1999;266:1001–12.
- Panzeri S, Senatore R, Montemurro MA, Petersen RS. Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* 2007;98:1064–72.
- Panzeri S, Treves A. Analytical estimates of limited sampling biases in different information measures. *Network* 1996;7:87–107.
- Percival DB, Walden AT. *Spectral analysis for physical applications: multitaper and conventional univariate techniques*. Cambridge, UK: Cambridge University Press; 1993.
- Petersen RS, Brambilla M, Bale MR, Alenda A, Panzeri S, Montemurro MA, et al. Diverse and temporally precise kinetic feature selectivity in the VPM thalamic nucleus. *Neuron* 2008;60:890–903.
- Pola G, Thiele A, Hoffmann KP, Panzeri S. An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network* 2003;14:35–60.
- Reich DS, Mechler F, Purpura KP, Victor JD. Interspike intervals, receptive fields, and information encoding in primary visual cortex. *J Neurosci* 2000;20:1964–74.
- Reinagel P. How do visual neurons respond in the real world? *Curr Opin Neurobiol* 2001;11:437–42.
- Ringach D, Shapley R. Reverse correlation in neurophysiology. *Cogn Sci* 2004;28:147–66.
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP. Spatiotemporal elements of macaque v1 receptive fields. *Neuron* 2005;46:945–56.
- Sceniak MP, Ringach DL, Hawken MJ, Shapley R. Contrast's effect on spatial summation by macaque V1 neurons. *Nat Neurosci* 1999;2:733–9.
- Schneidman E, Bialek W, Berry 2nd MJ. Synergy, redundancy, and independence in population codes. *J Neurosci* 2003;23:11539–53.
- Schyns PG, Thut G, Gross J. Cracking the code of oscillatory activity. *PLoS Biol* 2011;9:e1001064.
- Simoncelli EP. Vision and the statistics of the visual environment. *Curr Opin Neurobiol* 2003;13:144–9.
- Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. *Annu Rev Neurosci* 2001;24:1193–216.
- Strong SP, Koberle R, de Ruyter van Steveninck R, Bialek W. Entropy and information in neural spike trains. *Phys Rev Lett* 1998;80:197–200.
- Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 2001;12:289–316.
- Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 2000;20:2315–31.
- Touryan J, Felsen G, Dan Y. Spatial structure of complex cell receptive fields measured with natural images. *Neuron* 2005;45:781–91.
- Victor JD. Approaches to information-theoretic analysis of neural activity. *Biol Theory* 2006;1:302–16.
- Vul E, Harris C, Winkielman P, Pashler H. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect Psychol Sci* 2009;4:274–90.
- Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Stat* 1938;9:60–2.
- Willmore BDB, Prenger RJ, Gallant JL. Neural representation of natural images in visual area V2. *J Neurosci* 2010;30:2102–14.
- Wilson EB, Hilferty MM. The distribution of chi-square. *Proc Natl Acad Sci USA* 1931;17:684–8.
- Wolf D. Mutual information as a Bayesian measure of independence; 1995 arXiv:comp-gas/9511002v1.
- Yamada WM, Lewis ER. Predicting the temporal responses of non-phase-locking bullfrog auditory units to complex acoustic waveforms. *Hear Res* 1999;130:155–70.