

Unbiased measures of transmitted information and channel capacity from multivariate neuronal data[★]

L. M. Optican¹, T. J. Gawne², B. J. Richmond², and P. J. Joseph¹

¹ Laboratory of Sensorimotor Research, National Eye Institute, Bethesda, Md, USA

² Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, Md, USA

Received February 28, 1991/Accepted in revised form May 22, 1991

Abstract. Two measures from information theory, transmitted information and channel capacity, can quantify the ability of neurons to convey stimulus-dependent information. These measures are calculated using probability functions estimated from stimulus-response data. However, these estimates are biased by response quantization, noise, and small sample sizes. Improved estimators are developed in this paper that depend on both an estimate of the sample-size bias and the noise in the data.

Introduction

Brain functions depend upon mechanisms that encode, process and transmit information. Despite decades of study, many basic neuronal mechanisms, such as the intrinsic neuronal code used by the brain to transmit information, remain unknown. Thus, in quantifying a neuron's response some arbitrary measure of its activity must be chosen. It is generally accepted that the timing of neuronal action potentials carries information. However, how the signal is encoded by that timing is not known. One approach to solving this coding mystery is to assume different codes, and then to compare the amount of stimulus-dependent information conveyed by each code.

Information theory can be helpful in quantifying this comparison (Eckhorn and Pöpel 1974; Eckhorn et al. 1976; Fuller and Looft 1984; Optican and Richmond 1987). However, accurate estimation of transmitted information from experimental data faces three obstacles. First, the data must be quantized, because the discrete form of information theory is most easily applied to biological problems. Second, biological systems often have highly variable or noisy responses. Third, experi-

mental considerations often limit the number of replications per condition, resulting in small sample sizes. In this paper we present a method for removing the biases from information measures caused by these three problems.

Information is usually calculated by converting experimental data into discrete input/output histograms called contingency tables. However, information calculated from contingency tables is known to be biased upward. This bias is additive, and if known it may be subtracted off to improve the estimation of information (Carlton 1969; Fagen 1978). There is an exact equation for this bias, but its application requires complete knowledge of all the stimulus-response probabilities (Carlton 1969). Unfortunately, these probabilities are usually not known. Several approximations have been developed, but these either overstate the bias for small sample sizes (Miller 1955; Carlton 1969; Macrae 1971), or require assumptions about the distribution of probabilities underlying the process (Macrae 1971).

Fagen (1978) corrected the sample-size bias in one-dimensional tables by two methods. In one method, an estimate of the sample bias was obtained with the jack-knife statistical technique. In the second method, that estimate was obtained from first-order approximation formulae. Fagen's methods do not deal with the quantization problem nor do the closed-form approximations seem applicable to neurophysiological data, where the underlying distributions are usually not known.

To avoid the quantization artifact, Sakitt proposed a method of forming contingency tables by quantizing responses according to their rank order, rather than their value (Sakitt 1980; Sakitt et al. 1983). However, this technique can not be applied to multidimensional data (Optican and Richmond 1987), and it biases the information upward at either end of the data range (Crowe et al. 1988).

Optican and Richmond (1987) used a non-parametric kernel estimation method to minimize the quantization bias in the contingency tables, but did not correct for sample-size bias.

[★] This work was supported in part by Air Force Office of Scientific Research Grant AFOSR-ISSA-88-0073

Thus, none of the available methods for calculating information measures accounts for all three bias sources presented above. This paper develops such a method for estimating transmitted information and channel capacity from multidimensional data. The new method is applied to simulated data sets with known distributions, and to data from individual neurons recorded in primate visual cortex.

Methods

1 Response quantification

The mean firing rate is often used as a univariate measure of neuronal response. However, the responses of neurons are clearly multivariate (Eckhorn et al. 1976; Cataneo et al. 1981; Fuller and Looft 1984; Richmond and Optican 1987; Richmond and Optican 1990). In making unfounded assumptions about how stimulus-dependent neuronal messages are encoded some aspects of neuronal function may be neglected. Thus, a complete, multivariate representation must be used to quantify a neuron's responses. The principal component (or Karhunen-Loève) decomposition of a continuous representation of neuronal activity may be used as a complete, multidimensional quantification of a neuron's response (Richmond and Optican 1987).

2 Transmitted information

The amount of information transmitted by a coded message is a measure of the reduction in the receiver's uncertainty about which message was sent. The amount of information transmitted about the particular stimulus s_j , averaged over all the responses in the set R , is the *conditional transmitted information*, $T(s_j; R)$:

$$T(s_j; R) = \sum_k \frac{p(s_j, r_k)}{p(s_j)} \log \frac{p(s_j, r_k)}{p(s_j)p(r_k)}. \quad (1)$$

The summation is over all the members of the set R . The *average transmitted information* can be calculated as the weighted sum of all the conditional transmitted informations:

$$T(S; R) = \sum_j p(s_j) T(s_j; R). \quad (2)$$

3 Channel capacity

Use of a communication channel is controlled by setting the a priori probabilities of stimulus occurrence, $p(s_j)$. Altering these probabilities can change the reduction in uncertainty caused by receipt of a message. Thus, transmitted information should be regarded as a function of both the channel itself and the way the channel was used in the experiment (Blahut 1987):

$$T(S; R) = T(p(s_j); p(r_k | s_j)). \quad (3)$$

The *channel capacity* is a measure of the maximum amount of information that the channel is capable of transmitting with any a priori input distribution:

$$C = \max_{p(s_j)} T(p(s_j); p(r_k | s_j)). \quad (4)$$

Blahut gives an iterative procedure for finding the set of stimulus probabilities that maximizes the information transmitted by a given channel (Blahut 1972, 1987).

4 Joint stimulus-response probability function

An experiment provides the data needed to calculate the joint probability of occurrence of any stimulus-response pair, $p(s_j, r_k)$, since that is just the probability that a given stimulus and a given response occurred together in the experiment. This probability function is usually estimated by quantizing the data pairs into bins to form a histogram. This quantization moves values near the edges of a bin to its center, which reduces the data's variability and thus decreases its entropy. Since the transmitted information is the difference between the prior and posterior entropies, decreasing the posterior entropy will bias the transmitted information upward. This quantization artifact can be reduced by using the kernel approach to obtain an estimate of the joint *pdf*, $p(s_j, r_k)$ (Fukunaga 1972; Silverman 1986; Optican and Richmond 1987). In kernel estimation, each data point is replaced with a continuous density function, centered on that point. The kernel estimate is the average of all these density functions. Fukunaga has suggested that a good kernel for such an estimator is a Gaussian pulse with the same variance as the distribution of the data points themselves. Such a kernel has the advantage that the statistical properties of the data, up to the second moment, are taken into account (Fukunaga 1972).

The kernel estimate of the multivariate pdf (for a single stimulus) based on n_j data points is:

$$p(s_j, \mathbf{r}) = \frac{1}{n_j} \sum_{k=0}^{n_j-1} g(\mathbf{r}, \mathbf{r}_k, h) \quad (5)$$

where g is the kernel function, \mathbf{r} is the v -dimensional response vector, and \mathbf{r}_k is the k -th v -dimensional data vector.

The multivariate Gaussian kernel function is given by (Fukunaga 1972):

$$g(\mathbf{r}, \mathbf{r}_k, h) = (2\pi)^{-v/2} h^{-v/2} |\Sigma|^{-1/2} \times \exp \left[-\frac{1}{2h^2} (\mathbf{r} - \mathbf{r}_k)' \Sigma^{-1} (\mathbf{r} - \mathbf{r}_k) \right] \quad (6)$$

where Σ is the covariance matrix of the kernel, $|\Sigma|$ is its determinant, h is given in (8), and \mathbf{z}' denotes the vector transpose of \mathbf{z} . A different covariance matrix, Σ_j , is used in the kernel for each subpopulation:

$$\Sigma_j = \frac{1}{n_j} \sum_{k=0}^{n_j-1} (\mathbf{r}_k - \bar{\mathbf{r}}_j)(\mathbf{r}_k - \bar{\mathbf{r}}_j)' \quad (7)$$

where $\bar{\mathbf{r}}_j$ is the average of all n_j responses to the given stimulus.

The function h gives the kernel estimator its dependence on the sample size:

$$h(n_j) = n_j^{-1/(3\delta_j)} \quad (8)$$

where δ_j is a measure of the intrinsic dimensionality of the response space (9), and the exponent value of 1/3 gives the estimator uniform consistency (Fukunaga 1972).

The value of δ_j is usually the number of dimensions in the response variable. However, in our method the joint probability function is estimated separately for the subpopulation formed by the responses to each individual stimulus. Thus, δ_j should measure the number of degrees of freedom of each subpopulation, rather than the number of dimensions in the response measure. For example, suppose three-dimensional responses to stimulus s_1 fall along one line, and to s_2 fall along another line. Here, each set of responses (considered separately) is only unidimensional. One way to determine this intrinsic dimensionality is from the proportion of total variance in the response subpopulation distributed along each response dimension. However, for both simulated noisy data and real neuronal data, an estimate based on the proportion of the total standard deviation distributed along each dimension gave better convergence with small sample sizes. Suppose, for the j -th stimulus-response subset, the standard deviation along the i -th dimension is λ_{ji} . Suppose the maximum standard deviation is along the m -th dimension. An estimate of the number of degrees of freedom in the j -th subpopulation is then given by the normalized sum of the standard deviations:

$$\delta_j = \frac{1}{\lambda_{jm}} \sum_{i=1}^v \lambda_{ji} \quad (9)$$

where v is the number of dimensions in the response space.

5 Multidimensional density estimation with a non-separable kernel

In a multivariate code the Σ_j matrix obtained for the j -th stimulus-response subset of the data is not always diagonal. Thus, the multivariate kernel is not generally separable and can not be calculated by multiplying a set of one-dimensional kernels together. However, these non-separable kernels can be generated easily by combining a linear transformation of the data with a Monte Carlo simulation technique for estimating probability density functions. Basically, the procedure obtains the desired probability function by 1) transforming to a domain where the distribution is separable (Guttman et al. 1971), 2) generating an appropriately distributed set of points, 3) transforming the set of points back to the original data domain, and 4) building a histogram of the points from the transformed set.

6 The unbiased estimators

The upward bias of transmitted information (T) and channel capacity (C) calculated using the kernel estimate of $p(s_j, r_k)$ is shown in panels a and b of Fig. 1 for three-dimensional simulated data.

To understand where this upward bias comes from, suppose that the stimulus and response are completely

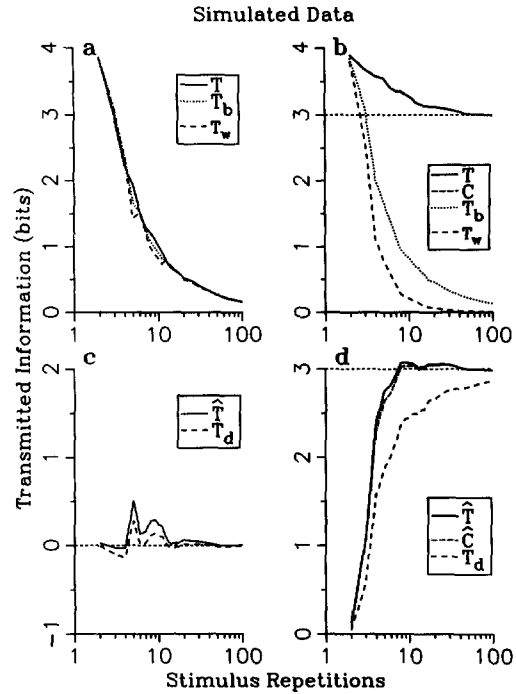


Fig. 1a–d. Dependence of transmitted information on sample size for simulated data. The data for the analysis in a and c were generated by a three-dimensional pseudo-random process (two components were independent Gaussian noise, and the third component was independent uniform noise). No information should have been transmitted by these data. Data for b and d were generated by adding a deterministic signal (step size five times the standard deviation of the noise) to the stochastic process used in a. The horizontal dotted line is the amount of information carried by the deterministic component of the signal. In a and b, T is the raw transmitted information, T_b is the bootstrap estimate of the bias, T_w is the noise-factor-weighted bias estimate, and C is the raw channel capacity (nearly identical to T in this example). In c and d, T_d is an estimator based on the difference $(T - T_b)$. \hat{T} is the improved estimator based on the difference $(T - T_w)$. \hat{C} is the improved estimator of channel capacity. Note that noise-factor weighting accelerates convergence to the correct value

unrelated, i.e., the response contains zero bits of information about the stimulus. The only distribution of points in the $p(s_j, r_k)$ matrix that yields zero bits of information is the one that has exactly the same distribution in every row (i.e., for every stimulus). If the stimulus-response relation is random, then as N increases the distribution actually found in each row will converge to the same distribution. However, for small values of N , the bins in each row will not be filled identically. This unequal filling will result in the apparent transmission of information. This non-zero value is a sample-size bias in the estimate of transmitted information. Note, however, that in the case of noise-free data, the bins in each row would be filled correctly, irrespective of N . Thus, the bias is dependent upon both sample size and noise.

Our technique for eliminating bias depends on building many artificial stimulus-response data sets, and thus is similar to Fagen's jack-knife technique (Fagen 1978). Each data set is made by randomly shuffling the real data set, which creates a new data set of the same

size and with the same statistical distribution of the stimuli and the responses. However, in each shuffled data set the stimulus and the response are independent of each other. Suppose that the number of these shuffled data sets is N_b . Let T_i be the transmitted information calculated from the i -th shuffled data set. Then a bootstrapped estimate of the transmitted information when the stimulus and response are independent is:

$$T_b = \frac{1}{N_b} \sum_{i=0}^{N_b-1} T_i. \quad (10)$$

The variance of this estimator is (Efron 1982):

$$\sigma_{T_b}^2 = \frac{1}{N_b^2} \sum_{i=0}^{N_b-1} (T_i - T_b)^2. \quad (11)$$

Subtracting the bootstrap estimate of the sample-size bias from the raw estimate produces an improved estimate of transmitted information (T_d in Fig. 1c and d), but it converges slowly as N increases. This difference estimate can be improved upon by incorporating an estimate of the noise in the data. Various measures of noise could be derived from the data set based on parametric estimates (e.g., the ratio of within-group variance to between-group variance). However, these will depend upon assumptions about the noise process (e.g., its distribution). Such parametric estimates are undesirable, since one advantage of using information theory is that its measures are model-free. Fortunately, an estimate of the noise can be obtained directly from the information measures themselves.

The ratio of the estimated bias (T_b) to the raw transmitted information (T) is a measure of the noise in the data. If the data were pure noise, then T and T_b would be virtually the same. If there were no noise, then T would be much greater than T_b . So, the ratio T_b/T can serve as a factor that is sensitive to the noise in the data set.

An improved estimator is formed from the raw estimate, T , and the product of the noise and bias estimates:

$$\hat{T} = T - \frac{T_b}{T} T_b. \quad (12)$$

The correction term can be applied to either the average or conditional transmitted information per stimulus. The performance of the improved estimator is shown in panels c and d of Fig. 1 (\hat{T}).

Subtracting the weighted bias term is equivalent to multiplying by a correction factor:

$$\hat{T} = \left[1 - \left(\frac{T_b}{T} \right)^\gamma \right] T \quad (13)$$

with $\gamma = 2$. This suggests another interpretation of \hat{T} : it is the raw T scaled by a quadratic measure of the noise-to-signal ratio. Varying γ gives a family of new estimators. The smaller the value of γ , the more conservative the estimator is for small sample sizes. With our data, the choice of $\gamma = 2$ gave well behaved results for sample sizes as small as ten.

The new estimator has three desirable properties. First, \hat{T} is an asymptotically unbiased estimator of the true transmitted information. To see this, note that T itself is asymptotically unbiased, since it is based on an asymptotically unbiased kernel estimate of $p(s_j, r_k)$ (Fukunaga 1972). Also, T_b is asymptotically unbiased, since as N increases the probability of finding the shuffled responses evenly distributed across the response space approaches one. The information transmitted by such an even distribution is zero. Thus T_b , the average value of T_i , also tends to zero as N increases. Hence \hat{T} , which is just the weighted difference of T and T_b , must also be an unbiased estimator of the transmitted information.

Second, \hat{T} is nearly zero for random data for all values of N , since if the stimulus-response relation is random, then shuffling should have no effect on the distribution of the responses. Hence, T and T_b will be nearly the same, and their difference will be nearly zero. Theoretically, \hat{T} can never be less than zero (Abramson 1963; Blahut 1987). However, if the transmitted information is low enough, or the data are noisy enough, it is possible for the average mutual information, $\hat{T}(S; R)$, to be slightly negative. This occurs whenever the true data points are more dispersed in the response space than their shuffled counterparts.

Third, \hat{T} will underestimate information for small values of N , because T_b is overestimated more than T . This conservative behavior is more desirable than the large overestimates given by T for small N .

An improved estimator of the capacity can be formed following an exactly parallel argument:

$$\hat{C} = \left[1 - \left(\frac{C_b}{C} \right)^2 \right] C \quad (14)$$

where C_b is the Monte Carlo estimate of C after shuffling the data. The reduction of the sample-size bias effect can be seen in Fig. 1d (\hat{C}).

7 Variance estimates

The uncertainty about \hat{T} depends upon the uncertainties of T and T_b . The variance of the estimate of T in closed form is (Optican and Richmond 1987):

$$\text{var}[T(s_j; R)] = \frac{1}{n_j} \left[\sum_k \frac{p(s_j, r_k)}{p(s_j)} \left(\log \frac{p(s_j, r_k)}{p(s_j)p(r_k)} \right)^2 - T^2(s_j; R) \right] \quad (15)$$

where n_j is the number of repeated data points for stimulus s_j . This closed form estimate is based on implicit assumptions about the underlying distribution (e.g., normality), and is thus only approximate (Fagen 1978). We compute $T(S; R)$ as the weighted sum of conditional transmitted informations, $T(s_j; R)$ (2). Thus, applying the rule for propagation of errors (Bevington 1969), the formula for the closed form variance estimate of $T(S; R)$ becomes:

$$\text{var}[T(S; R)] = \sum_j p(s_j)^2 \text{var}[T(s_j; R)]. \quad (16)$$

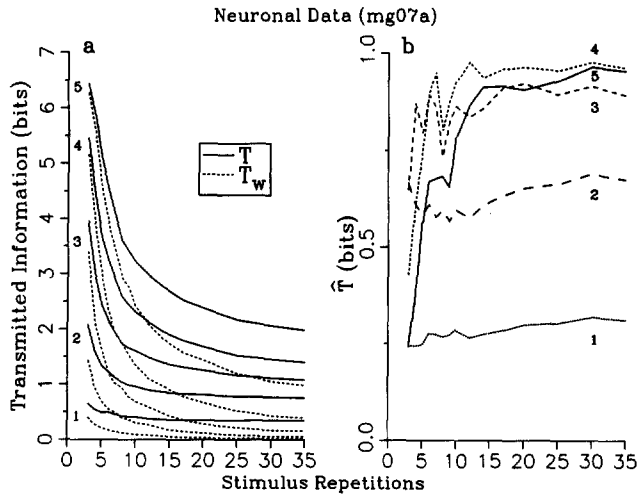


Fig. 2a, b. Dependence of transmitted information on sample size for data from a complex cell (mg07a) in monkey striate cortex. T was based on temporal codes with 1–5 principal components. **a** shows the raw estimate of transmitted information (T) and the noise-weighted bias estimate (T_w). **b** shows the improved estimate of transmitted information (\hat{T}) in 1–5 dimensions

The variances for T (16) and T_b (11) can be combined to give the variance of \hat{T} (Bevington 1969):

$$\sigma_{\hat{T}}^2 \approx \sigma_T^2 \left(1 + \frac{T_b^2}{T^2} \right)^2 + \sigma_{T_b}^2 \left(-2 \frac{T_b}{T} \right)^2. \quad (17)$$

A similar equation can be derived for the improved estimator of capacity, \hat{C} .

Results

The improved estimators developed above were applied to recordings from individual neurons (Richmond et al. 1990). The data used here came from two complex cells in the primary visual cortex of a monkey performing a fixation task. The stimulus set consisted of 128 pictures, and the response codes were based on from one through five principal components of the response (Optican and Richmond 1987). Figure 2 shows the dependence of \hat{T} for neuronal data as different numbers of trials were included in the analysis. The performance of \hat{T} is very good for sample sizes as small as ten to fifteen (Fig. 2b), especially when compared with the performance of the data of the raw estimator in more than two dimensions (Fig. 2a).

Using biased estimators of transmitted information we have shown that the temporal modulation of a neuron's response carries more information about the stimulus than the strength of the response alone (Optican and Richmond 1987; Richmond and Optican 1990). Figure 3 shows that this increase in information transmitted by a temporal code of increasing dimensionality persists with the unbiased estimators.

The curve labeled *pc* in Fig. 3 shows the information transmitted by each data component alone. The information per component decreases rapidly because

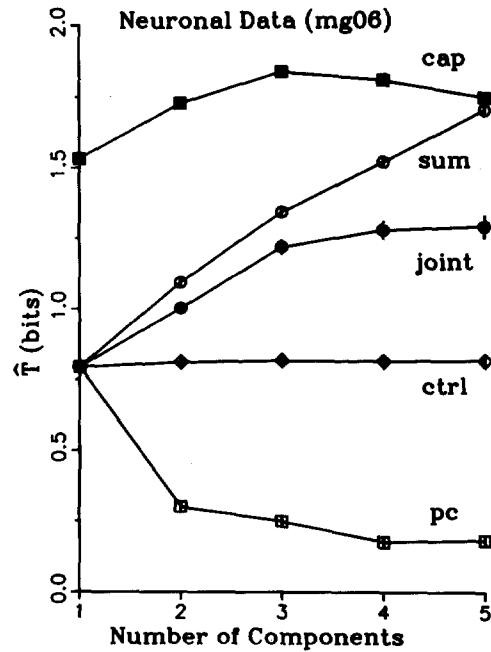


Fig. 3. Transmitted information carried by temporal modulation in a complex cell (mg06) in monkey striate cortex. \hat{T} for individual principal components (*pc*). The joint \hat{T} calculated using codes of increasing dimensionality (*joint*). The sum of \hat{T} for each component (*sum*) is more than the joint \hat{T} , implying some redundancy in the temporal code. The joint channel capacity \hat{C} is also shown (*cap*). As a control on the correct multidimensional estimation of $p(s_j, r_k)$, joint information was calculated using the same principal component along 1–5 dimensions (*ctrl*). Vertical bars are \pm one SEM

successive principal components account for less and less of the variability in the data set. The curve labeled *sum* is just the cumulative sum of \hat{T} for the individual components. If the information transmitted by each principal component were independent, then the amount of information transmitted jointly by all the components would be their sum. The curve labeled *joint* is \hat{T} calculated from the components jointly. The *joint* curve is below the *sum* curve, which suggests that there is some redundancy in the information transmitted by the individual components. Redundancy is useful in encoding information to be transmitted over noisy channels (Shannon 1948), and thus some of this redundancy may be part of the brain's encoding strategy.

As a control on the estimation of $p(s_j, r_k)$ in more than one dimension, the line labeled *ctrl* in Fig. 3 shows that increasing the dimensionality of the $p(s_j, r_k)$ matrix while keeping the data one-dimensional (i.e., placing the data along a diagonal in the response hyperspace) does not change \hat{T} .

The curve labeled *cap* shows the improved estimator of channel capacity. The choice, number, and presentation frequency of stimuli in these experiments was arbitrary. The average transmitted information per stimulus was small, because many of the stimuli could not be differentiated by the responses they elicited. The channel capacity, however, reflects the most information that could have been transmitted by that neuron about those stimuli if a better a priori distribution of

presentation frequencies had been chosen. For example, stimuli that were clearly differentiated by their elicited responses would be shown more often, reducing the chance of errors.

Figure 3 shows that the channel capacity of the neuron is greater than the average transmitted information, but this difference decreases as the dimensionality of the code increases. This suggests that neurons could utilize a larger fraction of their available channel capacity using multivariate temporal codes than they could using a univariate code (e.g., mean rate).

Discussion

Crowe et al. (1988) have suggested that for information theory to be useful in biological studies, a new method must be developed that eliminates the problems of quantization and sample-size bias in estimating the transmitted information. Our new estimators of transmitted information and channel capacity, based on a quantized kernel estimate of the joint probability function, a Monte Carlo estimate of the small-sample bias, and a ratio estimate of the noise in the data set, satisfy those criteria. Results from simulated data and neuronal data show that the new estimators overcome the upward bias effects on information measures from three sources: response quantization, noise, and small sample sizes. The improved estimators thus allow results from different experiments, or from different analyses on data from the same experiments, to be compared.

One disadvantage of using the average transmitted information alone as a measure of coding efficiency is that it tends to decline as the number of stimuli increases. Thus, the channel capacity provides another important measure of the code's efficiency, since it is largely independent of the number of stimuli used. Indeed, the limited channel capacity of neurons, coupled with their adaptive properties, raises the intriguing possibility that neurons might adapt (over long or short time courses) to optimize their use of a limited channel capacity. Such mechanisms need to be tested by formulating specific models of neuronal function.

Using the new tools presented here it is possible to avoid assumptions about which parameters of a neuron's response are important. Now, candidate neuronal codes (such as mean firing rate or temporal modulation) can be compared quantitatively. Furthermore, these improved estimates can be used as an independent test of the predictive ability of parametric models of neuronal function (because information theory is model-free). Thus, these new estimators make it possible to use information theory as a tool to aid in understanding how neurons encode and transmit information.

Acknowledgements. We would like to thank Dr. Richard E. Blahut of IBM Corp. for discussions on calculating information measures. We are grateful to Dr. Karen D. Pettigrew of the Theoretical Statistics and Mathematics Branch, National Institute of Mental Health, for discussions on estimator bias.

References

- Abramson N (1963) Information theory and coding. McGraw-Hill, New York
- Bevington PR (1969) Data reduction and error analysis for the physical sciences. McGraw-Hill, New York
- Blahut RE (1972) Computation of channel capacity and rate-distortion functions. *IEEE Trans Inf Theory* IT-18:460–473
- Blahut RE (1987) Principles and practice of information theory. Addison-Wesley, Reading, Mass
- Carlton AG (1969) On the bias of information estimates. *Psychol Bull* 71:108–109
- Cattaneo A, Maffei L, Morrone C (1981) Patterns in the discharge of simple and complex visual cortical cells. *Proc R Soc London B* 212:279–297
- Crowe A, de Ruiter T, Blaauw M, Oosthoek B (1988) Information transmission in non-visual fingertip matching along a horizontal track in the median plane. *Biol Cybern* 58:141–148
- Eckhorn R, Grüsser O-J, Kröller J, Pellnitz K, Pöpel B (1976) Efficiency of different neural codes: information transfer calculations for three different neuronal systems. *Biol Cybern* 22:49–60
- Eckhorn R, Pöpel B (1974) Rigorous and extended application of information theory to the afferent visual system of the cat. I. basic concepts. *Kybernetik* 16:191–200
- Efron B (1982) The jackknife, the bootstrap and other resampling plans. *SIAM Monograph*, vol 38. CBMS-NSF. SIAM Philadelphia
- Fagen RM (1978) Information measures: statistical confidence limits and inference. *J Theor Biol* 73:61–79
- Fukunaga K (1972) Introduction to statistical pattern recognition. Academic Press, New York
- Fuller MS, Looft SJ (1984) Information-theoretic analysis of cutaneous receptor responses. *IEEE Trans Biomed Eng BME-31*:377–383
- Guttman I, Wilks SS, Hunter JS (1971) Introductory engineering statistics. Wiley, New York
- Macrae AW (1971) On calculating unbiased information measures. *Psychol Bull* 75:270–277
- Miller GA (1955) Note on the bias of information estimates. *Inf Theory Psychol Probl Methods II-B*:95–100
- Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* 57:162–178
- Richmond BJ, Optican LM (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex: II. Quantification of response waveform. *J Neurophysiol* 57:147–161
- Richmond BJ, Optican LM (1990) Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. II. Information transmission. *J Neurophysiol* 64:370–380
- Richmond BJ, Optican LM, Spitzer H (1990) Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. *J Neurophysiol* 64:351–369
- Sakitt B (1980) Visual-motor efficiency (VME) and the information transmitted in visual-motor tasks. *Bull Psychonom Soc* 16:329–332
- Sakitt B, Francis L, Zeffiro TA (1983) The information transmitted at final position in visually triggered forearm movements. *Biol Cybern* 46:111–118
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
- Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
- Dr. Lance M. Optican
Chief, Section on Neural Modeling
Laboratory of Sensorimotor Research
Building 10, Room 10-C-101
National Eye Institute, NIH
Bethesda, MD 20892
USA