

Information theory and neural coding

Alexander Borst¹ and Frédéric E. Theunissen²

¹ESPM-Division of Insect Biology and ²Dept. of Psychology, University of California, Berkeley, California 94720, USA

Correspondence should be addressed to A.B. (borst@nature.berkeley.edu)

Information theory quantifies how much information a neural response carries about the stimulus. This can be compared to the information transferred in particular models of the stimulus–response function and to maximum possible information transfer. Such comparisons are crucial because they validate assumptions present in any neurophysiological analysis. Here we review information-theory basics before demonstrating its use in neural coding. We show how to use information theory to validate simple stimulus–response models of neural coding of dynamic stimuli. Because these models require specification of spike timing precision, they can reveal which time scales contain information in neural coding. This approach shows that dynamic stimuli can be encoded efficiently by single neurons and that each spike contributes to information transmission. We argue, however, that the data obtained so far do not suggest a temporal code, in which the placement of spikes relative to each other yields additional information.

The brain processes sensory and motor information in multiple stages. At each stage, neural representations of stimulus features or motor commands are manipulated. Information is transmitted between neurons by trains of action potentials (spikes) or, less frequently, by graded membrane potential shifts. The ‘neural code’ refers to the neural representation of information, and its study can be divided into three interconnected questions. First, what is being encoded? Second, how is it being encoded? Third, with what precision? Neurophysiologists initially approached these questions by measuring stimulus–response curves, using mainly static stimuli. The stimulus (x -axis) indicates what is being encoded, the response (y -axis) and the curve’s shape determine how it is being encoded, and error bars indicate the code’s precision. By using different stimulus ensembles and different response measures, one can begin to answer questions one and two. The precision of the code is implicit in the variance but has also been addressed directly by quantifying how well stimuli can be discriminated based on neural responses.

Measuring neural reliability is important for many reasons related to how the three questions interconnect. The crucial first question cannot be answered directly but will always depend on the investigator’s intuition and experience in choosing relevant stimulus parameters. Moreover, how such parameters vary in the chosen stimulus ensemble can lead to different results. For example, an auditory physiologist interested in frequency tuning might obtain different results from pure tones versus white noise. One way to validate the choice of stimulus parameters and ensemble is to compare behavioral performance to the best performance possible by an ideal observer of the neural data. A match between behavioral and neural discrimination suggests that the chosen encoding description is relevant and perhaps directly involved in generating behavior^{1–3}.

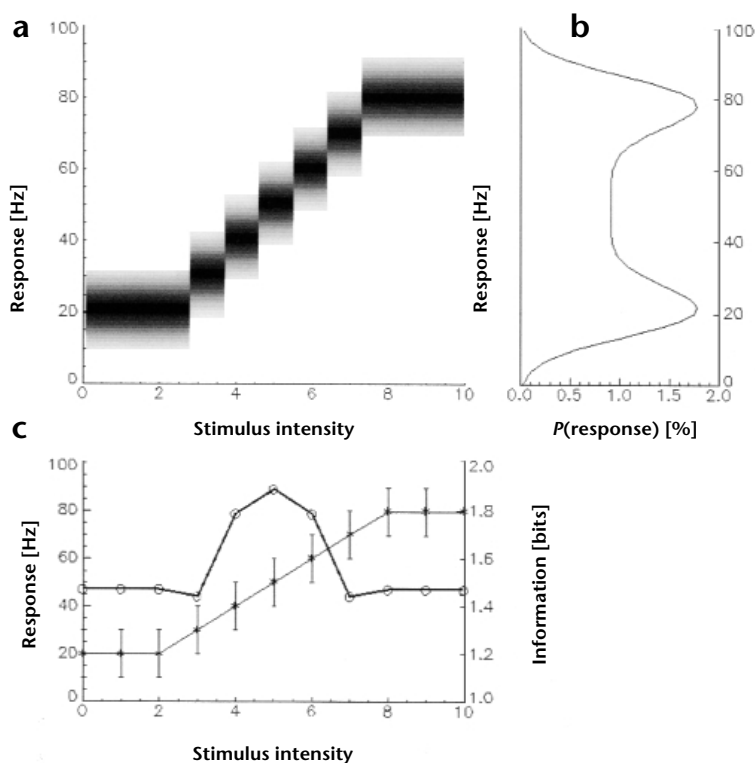
Information theory, the most rigorous way to quantify neural code reliability, is an aspect of probability theory that was developed in the 1940s as a mathematical framework for quantifying information transmission in communication systems⁴. The theory’s rigor comes from measuring information transfer precision by determining the exact probability distribution of outputs given any particular signal or input. Moreover, because of its mathe-

tical completeness, information theory has fundamental requirements on the maximum information transferrable in a particular communication channel. In engineering, information theory has been highly successful in estimating the maximal capacity of communication channels and in designing codes that take advantage of it. In neural coding, information theory can be used to precisely quantify the reliability of stimulus–response functions, and its usefulness in this context was recognized early^{5–8}.

We argue that this precise quantification is also crucial for determining what is being encoded and how. In this respect, researchers have recently taken greater advantage of information-theoretic tools in three ways. First, the maximum information that could be transmitted as a function of firing rate has been estimated and compared to actual information transfer as a measure of coding efficiency. Second, actual information transfer has been measured directly, without any assumptions about which stimulus parameters are encoded, and compared to the necessarily smaller estimate obtained by assuming a particular stimulus–response model. Such comparisons permit quantitative evaluation of a model’s quality. Third, researchers have determined the ‘limiting spike timing precision’ used in encoding, that is, the minimum time scale over which neural responses contain information. We review recent work using some or all of these calculations^{9–16}, focusing on the goodness of simple linear models commonly used to describe how sensory neurons encode dynamic stimuli. We conclude that these models often capture much of the transmitted information, and that each spike carries information.

Information-theoretic calculations also show that certain neurons use precise temporal (millisecond) spiking patterns in encoding. Precise spike timing had previously been identified in the auditory system, where it is important for sound localization¹⁷ and echolocation¹⁸, and also more recently elsewhere in the CNS¹⁹. The interesting question is whether spike timing precision is greater than necessary to encode the stimulus. New information-theoretic techniques address that question by quantifying spiking precision and comparing it to the minimal precision required for encoding in a variety of sensory systems^{2,9,13–16}.

Fig. 1. A mock neuron is tested with different stimulus intensities (from 0 to 10). For each stimulus intensity, it reveals a Gaussian distribution of spike responses around a mean value, ranging from 20 Hz for weak up to 80 Hz for strong stimuli. (a) Complete response distributions for each stimulus intensity; darker values indicate higher probabilities. (b) Summing these values along the horizontal lines leads to the overall response probability distribution (right), assuming that each stimulus is equally likely to occur. (c) Information theory allows one to replace the traditional stimulus–response curve (mean \pm s.d.) with an information curve (thick line) that indicates how well different values of the stimulus are encoded in the response. The information calculation is based not only the mean value of the response but also on its complete distribution at each stimulus condition. The distribution of responses obtained for this mock neuron at the middle of its operating range is more unique than the distribution of responses obtained for other stimulus values, leading to maximal values of information in that range.



We contrast the role of precise spiking in encoding dynamic stimuli to its potential role in situations where the stimuli do not vary rapidly in time, so that precise spike patterns could carry additional information not related to stimulus dynamics²⁰. Such temporal codes are suggested by data from single neurons and neuron ensembles^{21–24}.

General concepts

Information theory measures the statistical significance of how neural responses vary with different stimuli. That is, it determines how much information about stimulus parameter values is contained in neural responses. If stimulus A yields a mean response r_A and stimulus B yields r_B , information in the response could be measured as the difference between r_A and r_B . However, two neurons with the same differential response ($r_A - r_B$) may have different variability in their individual trial responses. Then the information obtained per trial is greater for the neuron with less variability. If response variability is described by the variance, then neuronal information can be described by the signal detection measure d' , which equals the differential response normalized by response variances²⁵. However, this is rigorously correct only if the distribution of response probabilities given particular stimulus conditions (conditional probability distribution) is completely specified by their mean and variance, as for Gaussian distributions.

The use of information as a statistical measure of significance is an extension of this process. Information theory allows one to consider not only response variance, but exact conditional probability distributions. In the example above, we can calculate conditional probabilities of various responses given stimulus condition A, $p(r|s_A)$, and again given stimulus condition B, $p(r|s_B)$, and then use information theory to calculate a distance between these two distributions. This analysis can be extended to a situation with many stimulus conditions $\{s_A, s_B, s_C, \dots\}$ to measure how the distribution of responses to any particular stimulus condition X is different from all other conditional distributions that can be obtained. This is done by comparing the conditional probability $p(r|s_X)$ to the unconditional probability $p(r)$ (the probability of the response under any stimulus condition) using the equation for $I(R, s_X)$ (Box 1). Plotting $I(R, s_X)$ as a function of stimulus condition X allows us to replace the traditional stimulus–response curve with a stimulus–information curve that shows how well an ideal observer could discriminate between the stimulus conditions based on a

single response trial (schematic example, Fig. 1; for actual examples, see refs. 8, 26, 27). The average information for all stimulus conditions $I(R, S)$ is then obtained by including the probability of occurrence of each condition (Box 1). In an experiment, stimulus condition probabilities are usually controlled and often equal. In such cases, $I(R, S)$ is obtained by summing all $I(R, s_X)$ for all possible stimulus conditions X and dividing by the total number of stimulus conditions. In natural situations, each stimulus condition has a different probability of occurrence, which might give very different mean information values. Information-theoretic values are strictly positive and traditionally measured in bits, representing the minimum length of a string of ‘zeros’ and ‘ones’ required to transmit the same information.

A second advantage is that information theory can be used to calculate maximal rates of information transfer. This measure, which is estimated from the set of all possible neuronal responses, is used to evaluate neuronal precision. For this purpose, we need to introduce entropy, which measures the information required to

Box 1. Information theory and significance of neuronal encoding.

$p(r_i)$	Probability that neural response takes the value r_i
$p(s_j)$	Probability that stimulus condition takes the value s_j
$p(r_i s_j)$	Probability that neural response takes the value r_i when stimulus condition s_j is presented (conditional probability)

Information about stimulus condition s_X :

$$I(R, s_X) = \sum_i p(r_i|s_X) \log_2 \frac{p(r_i|s_X)}{p(r_i)}$$

Average information obtained from all stimulus conditions:

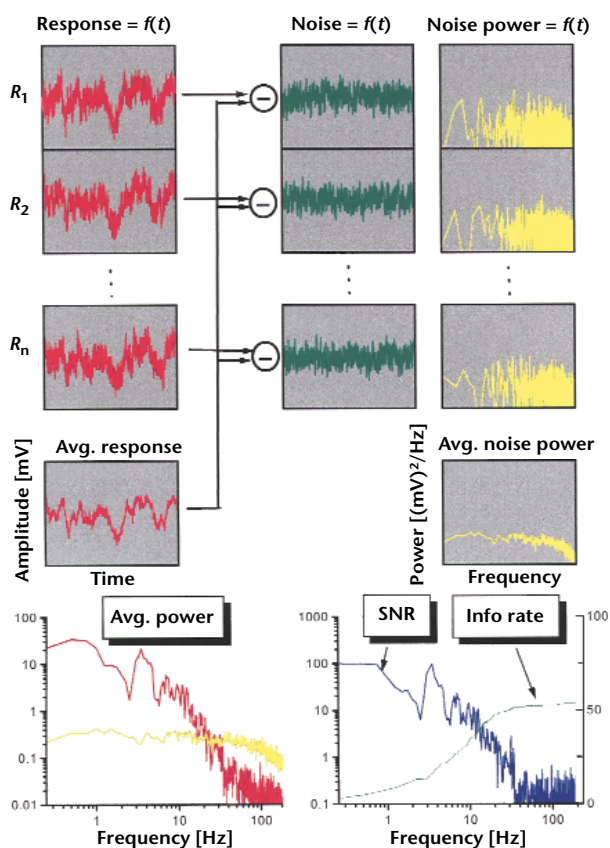
$$I(R, S) = \sum_i \sum_j p(s_j) p(r_i|s_j) \log_2 \frac{p(r_i|s_j)}{p(r_i)}$$

Fig. 2. Flow chart of how to measure the channel capacity of a neuron. The same stimulus is presented n times while the responses R_i are measured (left). These responses are averaged to obtain the average response R_{avg} . The difference between each R_i and R_{avg} become the noise traces N_i (middle). These are Fourier-transformed to the noise power spectra $N_i(f)$ (right), which can be averaged as well. Bottom left, power spectra of the mean response (red) together with the mean power spectra of the noise (yellow). Bottom right, ratio of these two functions, the so-called signal-to-noise ratio or SNR, together with the cumulative information rate. Response and noise data were created in a pseudo-random way from Gaussian distributions.

code a variable with a certain probability distribution by characterizing how many states it can assume and the probability of each. For example, a distribution with few conditions (such as light on and light off) contains less information (smaller entropy) than a distribution with many conditions (such as natural scenes). A distribution in which one condition is very probable and others very improbable has less entropy than a distribution in which all conditions are equally probable. Entropy, like information, is expressed in bits. The entropy of a distribution of stimulus conditions, $H(S)$ (Box 2), corresponds to the number of bits required to perfectly specify all stimulus conditions. Similarly $H(R)$, the entropy of the neural response, corresponds to the number of bits required to specify all possible responses under all possible stimulus conditions. Thus entropy is the information needed to encode all variability, or equivalently to eliminate all uncertainty about a variable.

Conditional probabilities are also used to calculate conditional entropies. In neural coding, $H(R|S)$ is the entropy in the neural response given the stimulus. This variable, called neuronal noise, measures the uncertainty remaining in the neural response when the stimulus conditions are known. Similarly, $H(S|R)$, called the stimulus equivocation, is the entropy remaining in the stimulus once the neural responses are known. Using Bayes' theorem, which relates joint probabilities (probability of a particular stimulus and response occurring together) to conditional probabilities, one can rewrite the information equation of Box 1 in terms of conditional entropies (Box 2). These new equations show that an information channel can be considered a channel for entropy transfer, in which some of the original entropy is lost and a different amount of new entropy is added (Box 2). The entropy of the stimulus $H(S)$ represents the maximum information that could be encoded, from which the stimulus equivocation $H(S|R)$ is lost. Therefore the information about the stimulus preserved in the neural response (termed 'mutual information') is $I(R, S) = H(S) - H(S|R)$. Adding the neuronal noise $H(R|S)$ to $I(R, S)$ gives the total neural response entropy, $H(R)$. Therefore $I(R, S)$ is also $H(R) - H(R|S)$. Note that entropy measures uncertainty and that information is defined as the difference of entropies—a reduction of uncertainty. In addition, information measures are symmetric in S and R , so that no causality is implied.

Because $H(R)$ represents the maximal information that could be carried by the neuron being studied, comparing $H(R|S)$ to $H(R)$ gives an estimate of the neural code's efficiency. However, $H(R)$ measured in an experiment still depends on the stimuli presented because they affect the range of neural responses observed. A more precise measure of efficiency is calculated by comparing the information transmitted by an actual neuron to the maximal possible response entropy. Similarly, the symmetry



of information measures in S and R can be used to measure how well the stimulus is being encoded. For example, $H(R|S)$ could be small in comparison to $H(R)$, but $H(S|R)$ could be large relative to $H(S)$. In that situation, even though neuronal efficiency is high, the possible stimulus conditions are not being encoded very well.

A final basic point of information theory is the 'data processing inequality' theorem. Its basis is the somewhat trivial statement that information cannot be recovered after being degraded. For example, consider a neural processing chain where S is encoded by a first neuron in a set of neuronal responses R_1 , and R_1 is then encoded by a second set of neuronal responses R_2 . The data processing inequality says that $I(S, R_1) \geq I(S, R_2)$. Note that this is true of all information channels, not just neurons. This theorem

Box 2. Entropy and information.

$$p(s, r) = p(s|r) \cdot p(r)$$

Bayes' theorem

$$H(S) = -\sum_i p(s_i) \log_2 p(s_i)$$

Entropy of S

$$H(R, S) = -\sum_i \sum_j p(s_i, r_j) \log_2 p(s_i, r_j)$$

Joint entropy of R and S

$$H(R|S) = -\sum_j p(s_j) \sum_i p(r_i|s_j) \log_2 p(r_i|s_j)$$

Conditional entropy of R given S or neuronal noise

$$H(R|S) = -\sum_j p(r_j) \sum_i p(s_i|r_j) \log_2 p(s_i|r_j)$$

Conditional entropy of S given R or stimulus equivocation

Equivalent forms for average information:

$$I(R, S) = H(R) - H(R|S)$$

$$I(R, S) = H(S) - H(S|R)$$

$$I(R, S) = H(R) + H(S) - H(R, S)$$

Fig. 3. Summary diagram for calculation of upper and lower bounds on information transfer. Top, situation where a stimulus S is corrupted by additive noise and subsequently fed through an unknown encoder to result in the response R . The lower bound is obtained with a linear reverse filter operation. The upper bound is obtained directly by comparing average and individual responses.

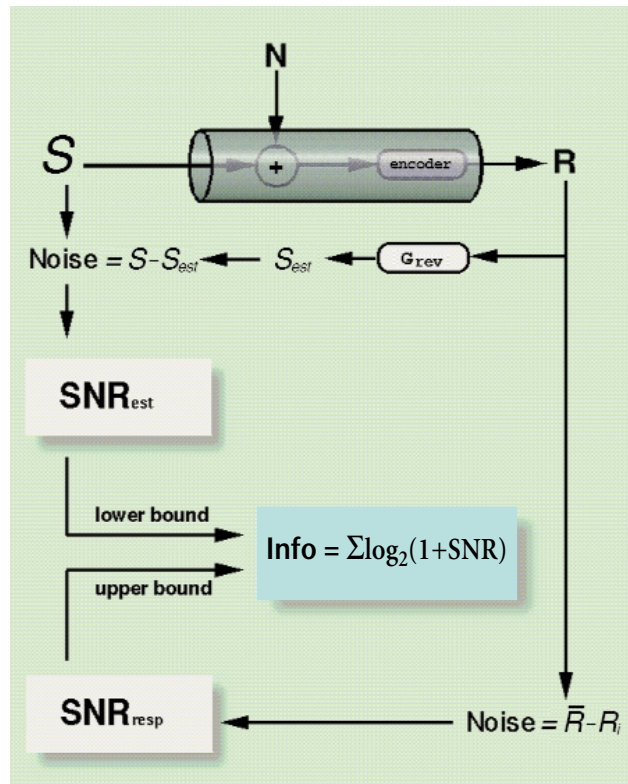
is a cornerstone of the method (below) used to find a lower bound on the amount of information about a dynamic stimulus transmitted in a neuronal channel.

By their choice of parameters to describe stimulus conditions or neural responses, and by the more fundamental choice of stimuli, neurophysiologists make assumptions that affect the information calculation. For that reason, information values are only applicable to a particular well-defined experimental context. However, because information-theoretic methodology allows one to quantify the accuracy of encoding and calculate maximal values of potential information transfer, it has become an essential tool to test the validity of these experimental assumptions.

In practice, it is easier to represent neural responses with minimal assumptions about the neural code than it is to find appropriate stimuli and the correct parameters to describe them. Neural responses can be represented with high temporal precision, and, with enough data, the relationship of any neural response measure to the stimulus conditions can be evaluated. Information measures can then be used to determine the limiting spike-timing precision involved in that particular encoding, for example by calculating the point at which information values stop increasing when analyzed over progressively shorter time windows. Similarly, information theory can guide the choice of parameters to represent the information being tested. Below we demonstrate how to estimate information transfer without making any assumptions about how the stimulus is encoded. This methodology can be used to test the validity of stimulus parameters and, more generally, of the stimulus ensemble, and thus to find the right model describing the neuron's stimulus-response function. We suggest that many parameters should be used initially (a 'rich' stimulus ensemble) to minimize assumptions. One can then search for the subset of parameters that most affect the information obtained in response to particular stimuli, compared to the average information obtained from all stimuli. We believe this process will lead to future experimental and theoretical breakthroughs.

Information theory and dynamic stimuli

Neuroscientists have recently used information theory to tackle the problem of characterizing information for continuously time-varying stimuli. This is difficult because the number of possible stimulus conditions quickly becomes enormous for any neural system with memory, as neural responses depend not only on the present stimulus but also on stimulus history. Therefore the stimulus must be specified as a vector of parameters, describing all preceding stimulus states relevant to the response. For example, if a certain stimulus parameter can have 8 different values, and the response depends on 7 previous states, suddenly 8^8 (that is, 16,777,216) different stimulus conditions must be represented. Because of this dimensional explosion, estimating the probabilities of stimulus and response is rarely practical. To avoid this problem, neurophysiologists have used three complementary methods. The first ('direct') method calculates information directly from the neural response by estimating its entropy, $H(R)$, and neural noise, $H(R|S)$. This method exactly determines the average information transmitted, but it does not reveal what aspects of the stimulus are being encoded.



ed. For the example in Fig. 1, the direct method would give the exact average value of the information curve without giving its shape. Because the direct method does not make any assumptions about response probability distributions, it also requires a lot of experimental data. The second method is similar to the first, with the added assumption that the neuronal response amplitudes, expressed in the frequency domain (see below), have Gaussian probability distributions. This method, which gives an upper bound for information transfer, requires significantly less data because Gaussian distributions are completely described by their mean and variance. The third method attempts to calculate information transfer for each possible stimulus condition to obtain the complete curve in Fig. 1. It therefore assumes a representation (choice of parameters)

Box 3. Entropy and information for Gaussian distribution and channel.

Gaussian distribution	$1 / \sqrt{2\pi\sigma_x^2} \cdot \exp(-x^2 / (2\sigma_x^2))$
mean = 0, variance = σ_x^2	
Gaussian entropy	$H(S) = \log_2(\sigma_s \sqrt{2\pi e})$
Gaussian channel	$R = S + N$, where S and N are Gaussian and independent.
Gaussian information	$I(S, R) = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_s^2}{\sigma_n^2} \right)$
Dynamic Gaussian channel	$I(S, R) = \int_0^k \log_2[1 + SNR(f)] df$

SNR(f) is the signal-to-noise power ratio at frequency f

Signal power at f is given by the variance of the Gaussian signal and is estimated by:

$\langle S(f)S^*(f) \rangle$

$S(f)$ is the Fourier transform of $s(t)$

$S^*(f)$ is the complex conjugate of $S(f)$

$\langle \rangle$ denotes the average over the experimental samples

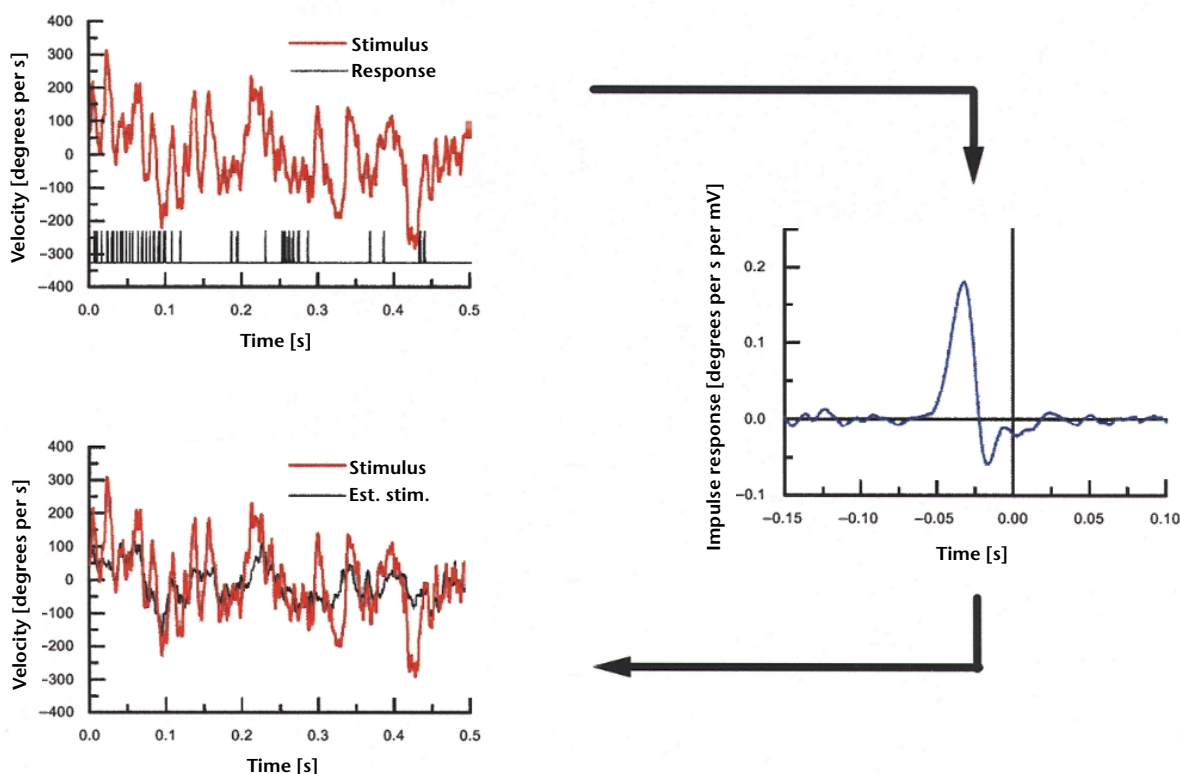


Fig. 4. An example of reverse reconstruction. A visual interneuron of the fly (HI cell) was stimulated by a grating moving in front of the animal using a pseudo-random waveform (upper diagram, red trace). The spike output of the cell (shown in black) follows the velocity signal only roughly. The impulse response of the reverse filter (right diagram) is negative in time and possesses band-pass characteristics. Applied to the neural response, the reconstructed or estimated stimulus function comes out as shown in black in the bottom diagram. Except for fast signal deflections, this signal is close to the stimulus (Haag and Borst, unpublished).

describing the stimulus conditions and a model relating these stimulus conditions to neural responses. This method always gives an average information rate lower than the actual information calculated by the direct method. Before describing these methods in detail, we briefly show how information-theory equations for time-varying signals (stimulus and response) are simplified when they are calculated in the frequency domain and the amplitude of each frequency component of the signal has a Gaussian distribution.

In a dynamic system with memory, probabilities are difficult to calculate because what happens now depends on what happened before. Thus all possible neural response and stimulus conditions must be considered simultaneously. However, this problem is simplified if the stimulus–response relationship depends on the relative time between their occurrences rather than the absolute time. Probability distributions describing such signals are called ‘stationary’ (although the signals are still dynamic). One can then use Fourier transformation to convert signals into a frequency-domain representation, that is, transforming probability distributions of signals at different times into probability distributions of signals at different frequencies. This is attractive because signals in the frequency domain, being the sums of many values collected at different times, are often statistically independent, unlike signals in the time domain. In this case, mutual information can be calculated independently for each frequency and summed to give the overall information.

However, to calculate information at a particular frequency, one still has to build probability distributions for stimulus and response amplitudes at that frequency. As in the time domain, this

requires large samples and is thus impractical. If the probability distributions can be approximated by Gaussian distributions, though, the situation changes completely. Then the entire probability distribution can be represented by its mean and variance, which can be estimated from the data. A case of particular simplicity and theoretical interest is when the response R can be obtained from a Gaussian stimulus S with zero mean simply by adding Gaussian noise with zero mean. This case is theoretically interesting because, for a given variance, the Gaussian distribution has the maximum possible entropy (see proof in ref. 14, section A.13). This property led to a famous formula on information capacity, defining the maximal information that can be transferred in an information channel given a particular signal variance, which Shannon proposed⁴. This property is also essential to the derivation of lower and upper information bounds discussed below. The equations for the entropy of a Gaussian distribution and for the information of a Gaussian channel (Box 3) depend only on the variance, σ^2 , of the distributions, as expected.

Calculating information transfer directly

The direct method is theoretically simple. Although the dimensionality explosion makes it practically impossible to calculate the joint probability of time-varying stimulus and response, the responses of single spiking neurons can be limited to strings of zeros and ones if the time windows used to divide the response are sufficiently small. Moreover, most possible strings occur rarely. Thus, one can directly estimate the total entropy of the spiking

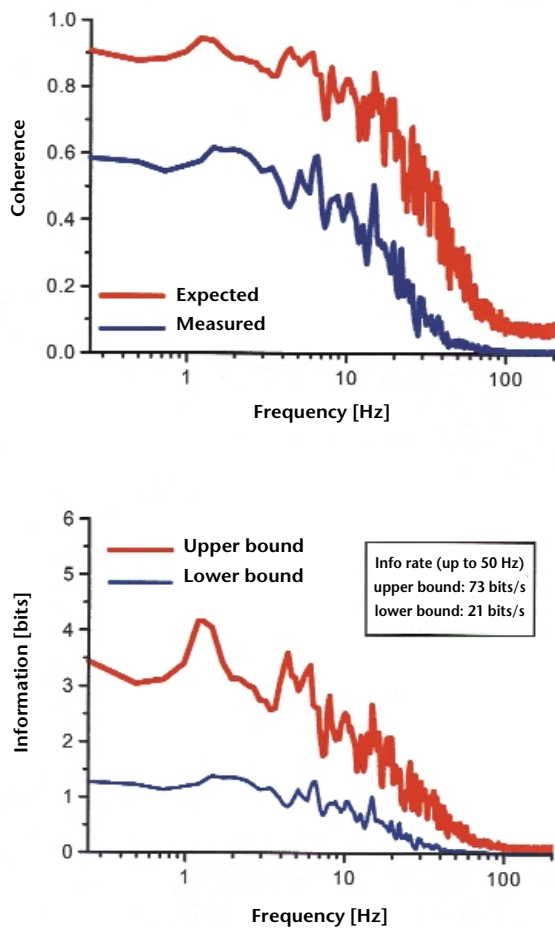


Fig. 5. Example of upper and lower bound of information as calculated from the spike train of a fly motion-sensitive nerve cell (H1 cell). The fly was stimulated by a moving grating while the spikes were recorded extracellularly. The lower bound was calculated from the coherence between the membrane potential and the stimulus velocity, the upper bound was calculated from the SNR. The upper integration limit was set to 50 Hz, because higher frequencies were not delivered by the stimulation device. The neural signal is carrying 21–73 bits per s about the stimulus velocity (Haag and Borst, unpublished).

responses (and their natural statistics) could be obtained. In practice, data collection is severely limiting, so we must use simple stimuli. Thus, the direct measure is most useful in gauging the goodness of the lower-bound estimates described below.

Calculating an upper bound on mutual information

The upper-bound calculation is a variant of the direct method that is used for dynamic stimuli. It assumes that the neuronal response and neuronal noise have Gaussian probability distributions in the frequency domain and that neuronal noise is additive. In this situation, we can define the stimulus S as the mean neuronal response obtained from many repetitions of identical stimulus conditions (Box 3). The actual response R is the response on individual trials, which then equals the mean signal plus a noise term. The noise is obtained from deviations of each individual response around the mean. This procedure (Fig. 3) is intended to separate deterministic aspects of encoding from those considered to be noise. As in the direct method, one then calculates the information from response entropy $H(R)$ and neuronal noise entropy $H(R|S)$, but in this case, both are obtained by simple averaging. This method for calculating neuronal noise is only valid when mean neuronal response and neuronal noise defined in this way are statistically independent (in other words, when the mean response reflects everything that can be learned about the stimulus). The direct method is more general in the sense that deviations from the mean response can carry information about the stimulus. For example, consider a temporal code in which the relative position of two spikes encodes the stimulus, but their absolute temporal position varies. In this case, the mean response carries no information about the stimulus, and deviations from the mean contain all the information.

response, $H(R)$, and the fraction of this entropy attributable to neuronal noise, $H(R|S)$. Spike train noise is determined by repeating a dynamic stimulus many times to get the response distribution under the same stimulus conditions. In this case, one does not worry about specifying parameters to describe the stimulus or calculating stimulus probabilities. Practically, however, this estimation is still difficult because one has to be careful when estimating the probabilities of occurrence of each spike response (for details, see ref. 28). The direct method has been used for dynamic systems when a lot of data could be obtained¹⁵ or with relatively simple stimuli¹⁶. This approach would seem to be the most satisfying because it gives a correct information measure, rather than an upper or lower bound, but it also has some limitations. First, for both experimental and computational reasons, researchers must limit the number of dimensions given by the size of the string of zeros and ones used to describe the response. This string size depends both on the size of the window used to parse the response into ones and zeros (the limiting temporal resolution) and on the length of time examined (memory of the system). Second, the direct method does not indicate which stimulus aspects are best represented. Finally, as with all estimates of information, one must remember that the information obtained depends on the stimulus ensemble. This is probably more obvious in the direct method than in the model-based method because it condenses a neuron's encoding properties to a single number. For this number to represent the average information transmitted by the neuron, we would have to sample many natural stimuli so that all possible

Box 4. Linear reconstruction formulas.

S_{est} is obtained by a linear filtering operation on R $S_{est}(f) = H(f) \cdot R(f)$

$H(f)$ is given by $H(f) = \frac{\langle R^*(f) \cdot S(f) \rangle}{\langle R^*(f) \cdot R(f) \rangle}$

The noise is given by $N = S - S_{est}$

The signal–noise ratio is $SNR = \frac{\langle S_{est} S_{est}^* \rangle}{\langle NN^* \rangle}$

The coherence between S and R is $\gamma^2 = \frac{\langle S^* R \rangle \langle R^* S \rangle}{\langle S^* S \rangle \langle R^* R \rangle}$

The signal to noise ratio is also $SNR = \gamma^2 / (1 - \gamma^2)$;

The information $Info_{LB} = -\int_0^\infty \log_2(1 - \gamma^2) df$;

The upper-bound method, however, estimates information transmission from significantly less data than the direct method. Because the upper-bound method assumes that mean response and noise have Gaussian probability distributions, it requires just enough data to correctly estimate the variance of the Gaussian probabilities used to model mean response and neuronal noise and to verify that the noise distribution is indeed Gaussian or almost Gaussian. The upper bound calculated by this method is the theoretical limit of information transmission obtainable from any input that causes similar power fluctuations in the neuronal response. This theoretical limit is called the channel capacity. The actual information may be lower because mean neuronal response statistics are not necessarily Gaussian (see also section 3.13, ref. 14). If this distribution is Gaussian with equal power at all frequencies (white noise), then the stimuli are optimally encoded (proof in appendix A15, ref. 14).

For example (Fig. 2), the mean response (signal) is estimated from many stimulus repetitions. The signal's power spectrum gives the variance at all frequencies. Noise is obtained by subtracting this mean response from each trial. The power spectrum of the noise is calculated to obtain the signal-to-noise ratio (SNR), which is then plugged into the equation for a dynamic Gaussian channel (Box 3). Information in the spike trains can still increase in the high-frequency range where SNR is smaller than one. For small SNR, this calculation must be done carefully. Estimation of SNR, and consequently of information, has a positive bias because power spectra can only be positive or zero. Thus, appropriate statistical tests must be used to decide whether the estimated SNR is significantly different from zero. Only in this case should cumulative information capacity increase. At frequencies where the SNR is no longer significantly different from zero, cumulative mutual information will flatten out. In Fig. 2, we used the jackknife resampling technique²⁹ to estimate the power spectra's significance.

Because the upper-bound method is based on strong assumptions, it can estimate information transmitted with much less data than the direct method. It can be applied both to spiking and non-spiking neuronal responses. Also, it estimates information transfer for different frequency components of the neuronal response. When linear models are used to estimate the lower bound, as described below, one can directly compare these estimates as a function of frequency to evaluate the model's quality.

Calculating the lower bound on information transfer

So far, we have calculated average information or, more practically, its upper bound without making any assumptions about what stimulus aspects are encoded. Here we describe how to investigate stimulus encoding by testing different encoding models. Because these models might not capture all the information, this gives a lower-bound estimate of information transmitted. One method of modeling stimulus encoding ('reverse reconstruction') describes how to calculate the best possible stimulus estimate from the neural responses. This estimate is then used to calculate the lower bound of information transmitted between stimulus and response. This method offers some advantages over the more traditional approach of estimating the response from the stimulus^{14,30}. In this procedure (Fig. 3), the stimulus signal S is encoded into response spike trains. A reconstruction algorithm of choice is then used to estimate S (S_{est}) from the response R . Mutual information between S and R is then estimated by calculating the information between S and S_{est} . From the data processing inequality, this information estimate is smaller than or equal to the information about S that is in R . Thus this procedure gives a lower bound on the information. If S is estimated well, the resulting lower bound is close to the real information transmitted about the stimulus. Otherwise, the lower bound is far from the real information. In this approach, the reconstruction algorithm models the neural encoding. The lower bound

lets us quantify the neuron's performance and (by comparison with the upper bound or direct estimation) the goodness of our model. This general methodology is applicable to any stimulus and any reconstruction model. Two progressively more restricted situations, Gaussian stimuli and use of a linear reconstruction algorithm, are of particular interest because they require significantly less experimental data. These additional assumptions can give good approximations to the actual information transmitted (calculated by the direct method).

First, to take advantage of the properties of a Gaussian channel, we use a stimulus with Gaussian distributions. We define noise as the difference between S and S_{est} . Assuming the noise is Gaussian and independent of S_{est} , we then estimate a lower bound on information transmitted by calculating the information in a channel with input S_{est} and output $S_{\text{est}} + N$ (in Box 3, $S = S_{\text{est}}$ and $N = S - S_{\text{est}}$). We

Table 1. Methods and assumptions of four ways to calculate neural information.

Method of estimation	Driving principle	Simplifying assumption	Further assumption
Lower bound	Find 'best' S_{est} from R $I(S, R) \rightarrow I(S, S_{\text{est}})$	Gaussian S Calculate $N = S - S_{\text{est}}$ \rightarrow Equations in Box 3	Linear decoder \rightarrow Box 4 and actual coherence
Absolute lower	Find 'best' S_{est} from R Find smallest $I(S', S_{\text{est}}')$ that would give the same error as $(S - S_{\text{est}})^2$	1) Use Gaussian S \rightarrow Equation 3.1 in ref. 31	
Upper bound (when all assumptions are true)	Separate R into a deterministic and a random component by repeating S many times $I(S, R) \rightarrow I(R, R_{\text{det}})$	Additive noise: $R_{\text{det}} = R_{\text{avg}}$ $N = R - R_{\text{avg}}$ $I(R, R_{\text{det}}) \rightarrow I(R, R_{\text{avg}})$	If N is Gaussian, \rightarrow Box 3 and expected coherence
Direct	Separate R into a deterministic and a random component by repeating S many times $I(S, R) \rightarrow I(R, R_{\text{det}})$	None except temporal resolution	

S = stimulus, S_{est} = estimated stimulus, R = response, R_{avg} = average response, R_{det} = deterministic part of the response, N = noise. All four methods reformulate the problem of finding the mutual information between S and R , $I(S, R)$, as an equivalent information calculation that is easier to perform.

can further relax the assumption of Gaussian noise because we know that the entropy of any non-Gaussian noise with the same variance is smaller than the entropy of Gaussian noise. In other words, this lower-bound estimate is most accurate if the noise is Gaussian, and the estimate is lower if the noise probability distribution deviates from this assumption.

Within this general framework, the key problem is finding the best way to estimate the stimulus from the response. Although this can be difficult, there are many signal-processing and systems-analysis methods for calculating transformations between two time-varying signals. Transformations with memory (history dependence) are called filters. In linear filters, the simplest type of memory transformations, the effect of the stimulus on the response at one time is added to the effect at all previous times. This linear operation in the time domain is called convolution. In the frequency domain, convolution corresponds to simple multiplication between corresponding frequency components of stimulus and response. Signal processing methods can be used to calculate optimal linear filters that transform R into S_{est} to minimize the difference between S and S_{est} .³¹ The optimal linear filter is obtained with a relatively simple formula, almost identical to the one used to calculate regression coefficients. This formula is the product of the frequency components of stimulus and response (also known as their cross-correlation) divided by the response power spectra. Reverse reconstruction (Box 4) is similar to the process termed 'reverse correlation'³² (used, for example, to calculate dynamic receptive fields for visual³³ or auditory³⁴ interneurons), except that the reverse filter is normalized by response power.

Note that we define noise relative to the estimated stimulus, unlike the 'estimated noise' in ref. 14, which is defined relative to the stimulus. In the linear case, effective noise is $N = S - S_{\text{est}}/\gamma^2$. Using the effective noise and replacing the stimulus with the estimated stimulus as the signal leads to the same equations (see refs. 30 and 35 for a more detailed derivation of both formulations).

As mentioned above, all information calculations depend on the choice of stimuli. Therefore, the information transmitted by a neuron may be smaller in a natural situation than in the laboratory. To calculate a lower bound that is more independent of the stimuli, one can use an information measure called the rate-distortion

Table 2. Neural information and spike precision in response to dynamic stimuli.

Animal system (Neuron) Stimulus	Method	Bits per second	Bits per spike (efficiency)	High-freq. cutoff or limiting spike timing
Fly visual ¹⁰ (H1) Motion	Lower	64	~1	~2 ms
Fly visual ¹⁵ (H1) Motion	Direct	81	—	0.7 ms
Fly visual ³⁷ (HS, graded potential) Motion	Lower and upper	36	—	—
Monkey visual ¹⁶ (area MT) Motion	Lower and direct	5.5	0.6	~100 ms
Frog auditory ³⁸ (Auditory nerve) Noise and call	Lower	Noise 46 Call 133	Noise 1.4 (~20%) Call 7.8 (~90%)	~750 Hz
Salamander visual ⁵⁰ (Ganglion cells) Random spots	Lower	3.2	1.6 (22%)	10 Hz
Cricket cercal ⁴⁰ (Sensory afferent) Mechanical motion	Lower	294	3.2 (~50%)	> 500 Hz
Cricket cercal ⁵¹ (Sensory afferent) Wind noise	Lower	75–220	0.6–3.1	500–1000 Hz
Cricket cercal ^{11,38} (10-2 and 10-3) Wind noise	Lower	8–80	Avg = 1	100–400 Hz
Electric fish ¹² (P-afferent) Amplitude modulation	Absolute lower	0–200	0–1.2 (~50%)	~200 Hz

function³⁵, which is the absolute lower bound of information obtainable for a particular error. This error can be calculated with the reverse reconstruction as explained above. Thus, if a similar error is obtained in natural situations with different stimuli, then this method provides a more accurate estimate of the lower bound.

Linear reconstruction algorithms have been used almost exclusively for reverse reconstruction (Table 2). For example (Fig. 4), a spiking neuron called H1 was recorded in the visual system of the fly. The stimulus was a grating moving randomly back and forth in front of the fly's eyes (Fig. 4a). Using the formulas in Box 4, an optimal linear reverse filter (green trace, Fig. 4b) was calculated based on the cross-correlation between velocity signal and spike output, divided by spike train power. Convolution of the spike train with this filter gives a stimulus estimate

(Fig. 4c) that matches the original stimulus fairly well, particularly in the slow components of its time course.

Combining the methods

We can now compare these various estimates of information transfer. When we have enough data to estimate an upper bound, but not enough for the direct method, we know the true information is between the two bounds but we can still attempt to improve the lower bound to find a better decoding scheme. In particular, because most experimenters use a linear filter to approximate decoding, using these comparisons to validate the decoding model allows us to measure deviations from linearity. If encoding were truly linear, the direct estimate and lower bound would be equal. Moreover, if stimuli have Gaussian distributions, then the upper bound equals the exact information, because the response signal is then Gaussian and noise is independent of the signal. We can define an 'expected' coherence (see Box 4) as given by the SNR measured directly from the response train in the estimation of the upper-bound information rate. Deviations of expected coherence from one are due to the system's intrinsic noise. Moreover, using the linear decoding scheme, actual coherence calculated from the lower bound equals expected coherence. In most cases, however, expected and actual coherence are different. This difference can be used to estimate the system's degree of linearity. In particular, if noise is truly independent and the signal is Gaussian, then differences between expected and actual coherence accurately measure the system's non-linearity^{13,36}.

In summary, lower and upper bounds of the information estimate can be derived from a stimulus-response set (see Fig. 3 and Table 1). From repeated presentation of identical stimuli, one calculates the signal-to-noise ratio (see Fig. 1), converts it to an expected coherence and uses the formula in Box 4 to determine the upper bound. The lower bound in the linear decoding case is obtained using the same formula, but using the actual coherence (Box 4). The assumptions in each of these steps are summarized in Table 1.

Figure 5 shows an example of actual and expected coherence as calculated from a visual interneuron of the fly, a so-called HS cell. In the low-frequency range, up to about 10 Hz, the expected coherence is about 0.9 on average. Thus about 10% of the missing coherence (with respect to a perfect representation) is due to response noise. In this frequency range, actual coherence is about 60%. Thus, another 30% can be attributed to response nonlinearities. For higher frequencies, both measured and expected coherence drop off to asymptote at zero level. In the low-frequency range, the information estimate is between 1 and 3 bits. Assuming that signals are independent for each frequency, the total information rate estimate therefore is between 21 and 73 bits per second³⁷.

What have we learned from information theory?

We have briefly reviewed the use of methods of systems analysis and information theory to estimate the precision of the neural code and the goodness of our models of encoding for dynamical stimuli. The absolute measure of information quantifies the code's precision, whereas comparison of lower or upper versus direct estimates tests the goodness of our neuronal decoding scheme and therefore of our understanding of neuronal processing. However, both absolute and relative measures depend on the choice of stimulus^{12,14,16,35,38} (in an obvious example, its bandwidth). The use of information to measure absolute precision is therefore subject to the same constraint as any other method of estimating the neural code's precision: the choice of stimulus ensemble. However, because information values are in absolute units, they can be used to evaluate the effectiveness of

different stimuli. In particular, one can search for stimuli that give large information values, and we suggest that these ensembles might be most valuable for determining neural encoding properties. An interesting hypothesis is that stimulus ensembles with naturalistic properties yield the highest information values, suggesting that neural processing is optimized to represent natural stimuli, as found in the auditory system of the frog³⁹.

Not many researchers have yet used information-theory and systems analysis techniques to characterize neural encoding of dynamic stimuli (Table 2). Most of them used only a linear decoding filter to model the stimulus-response function, and obtained only the corresponding lower-bound estimate of information. The linear filters' forms implied that most neurons could be thought of as low-pass or band-pass filters. For peripheral sensory neurons, this result is not particularly striking. Moreover, the high-frequency cutoff for most neurons studied was relatively low, and much higher frequencies are encoded by auditory neurons, as discussed below. However, these papers go beyond simply describing neuronal tuning properties. The absolute values of information revealed the importance of every spike and the relatively low neural noise. Information measures of around one bit per spike were found. This suggests that every spike allows an ideal observer to reduce uncertainty about the stimulus identity by half. By comparing overall information transfer to maximum spike train entropy achievable with identical spike rates, researchers have further quantified this high level of encoding^{39,40}. The ratio of these numbers can be used to define a coding efficiency (see also ref. 14, pp 166–175). In the cricket cercal system, this coding efficiency is about 50% (Table 2). This efficiency measure is sensitive to the stimulus. In the frog auditory system, efficiencies of 90% were measured in response to a natural stimulus ensemble (Table 2)³⁹. Using an upper-bound measure based on the finding that electric fish P receptors fire only one spike per cycle of carrier frequency¹², another group also found coding efficiencies of about 50% (Table 2). These measures of coding efficiency could not be obtained from classical stimulus-response characterizations. We and others suggest that 50% is a very high number, considering the high entropy obtainable from considering any possible spike pattern given a fixed number of spikes. These results verify that single sensory neurons at the periphery have high fidelity, as sensory neurophysiologists have long known⁴¹. Note, however, that high fidelity of single neurons does not necessarily imply high fidelity of stimulus encoding. When information is compared to source entropy, or when information is plotted as a function of stimulus condition, one finds that only a limited bandwidth of the stimulus is represented and that when the bandwidth is large, the relative information encoded compared to the stimulus entropy can be low. In those cases, joint consideration of neural responses would lead to high stimulus encoding, but this consideration should not consist of simply averaging responses because this causes information loss.

The other very promising strength of these information-theoretic measures is the possibility of calculating the absolute amount of information transmitted (upper and direct estimation of information) to test the goodness of encoding models. In all papers with this comparison so far^{10,15,16,37}, linear decoding only captures a fraction of overall information transmitted, albeit a large fraction. The information-theoretic methodology allows one to identify system non-linearities and can validate any nonlinear model investigated in the future. This will help to bridge the gap between the very quantitative analysis used to describe linear neurons found at the sensory periphery and the more qualitative description of nonlinear neurons such as combination-

sensitive auditory neurons in bats⁴² and songbirds⁴³ or face-selective neurons in primates⁴⁴. This discrepancy results from the difficulty of systematically deriving nonlinear models from neural data. On the other hand, nonlinear encoding is arguably more interesting because it occurs in higher-order neural processing involved in complex feature extraction.

Spike timing precision and temporal codes

Here we analyze the nature of the neural code. A universal finding in information calculations for dynamic stimuli is the relatively high importance of single spikes, in the sense that the information per spike is high. Because the temporal placement of spikes is also well preserved, this suggests that temporal spike patterns are an important aspect of the code. This general statement, however, does not imply that the neural code is temporal rather than based on spike number.

For dynamic stimuli, both 'what' and 'when' aspects of the stimulus could be encoded in spike train patterns. Humans and other animals are sensitive to 'when' aspects of dynamic stimuli. Psychophysical measurements of stimulus occurrence detection reveal microsecond precision for a multitude of sensory modalities. In the auditory system, echolocation¹⁸ and sound localization¹⁷ require particularly fine temporal resolution. Such behaviors must be mediated by precise representation of time in the CNS. In certain situations, spike patterns show a finer spike precision than necessary. When such spike patterns encode 'what' aspects of the stimulus that are not encoded in the firing rate, then encoding truly can be labeled 'temporal'. This definition distinguishes spike timing required by stimulus dynamics from spike timing used to encode non-dynamic aspects^{20,45}. In contrast, precise spike timing is often contrasted with a rate code. Because a rate code can be estimated with an arbitrarily small time window, high spike timing precision and rate coding are not mutually exclusive, whereas the difference between temporal encoding and rate coding can be rigorously defined.

The methods discussed here can be used to measure stimulus encoding accuracy and the corresponding spike timing precision. In general, because no assumptions are made about the encoding, spike timing precision can be used both for temporal encoding ('what') and for time coding ('when'). This calculation can also be done for both single neurons and neuron ensembles, although we focused on examples from single neurons, reflecting current progress. We elaborate on spike timing issues when the lower bound is obtained by linear decoding, the most common case so far.

To estimate the lower bound of information about dynamic stimuli, we did all our calculations in the frequency domain. A universal result in such analyses is the existence of an upper frequency cutoff at which dynamic stimulus aspects stop being encoded in the spike train response. This upper frequency limit is the frequency at which information goes to zero. To encode dynamic stimulus changes up to that cutoff frequency, limiting spiking precision must be at least roughly the time resolution given by half the inverse of the cutoff frequency (called the Nyquist limit). When one assumes linear encoding, the energy at a particular frequency in the spike trains encodes the same frequency in the stimulus. In those cases, spike placement with time resolution smaller than the window given by the Nyquist limit has no effect in representing the stimulus. The number of spikes within that time window can be used to encode 'what' aspects of the stimulus (for example, amplitude). For a linear encoder, we effectively assume a rate code where the rate is estimated for time windows given by the cutoff frequency of the stimulus encoding, and the linear filter can be thought of as one of the

most appropriate transformations to obtain a mean firing rate^{20,46}. Any other window used to estimate firing rate may decrease information in the spike train (by low-pass filtering).

Therefore with the linear model and the knowledge that high-frequency stimulus components are being encoded, the corresponding limiting spike timing is not surprising, nor is it indicative of a temporal code. On the contrary, it is necessary to represent the time-varying stimulus. What is surprising, however, is the combination of high information transmission with relatively low spiking rates. High total information can be obtained by encoding a large bandwidth or by encoding a smaller bandwidth very precisely. Absolute measures of total information reach values of ~300 bits per second for the lower bound (Table 2). When the spike rate is taken into account, most cases yielded on the order of one bit per spike. This result supports the statement that 'every spike counts'. Therefore spike timing in these examples is essential, even though it can still be called a rate code. Such a firing rate is obtained not by averaging over many neurons (or stimulus repetitions) but by convolving the spike train with an appropriate filter.

Information measures in bits per spike do not translate directly into spike timing precision in milliseconds, but in the linear case, the high-frequency cutoff of stimulus encoding corresponds to the limiting accuracy of spike timing. To determine this cutoff, one needs to find the point at which information becomes statistically indistinguishable from zero. To do so correctly requires obtaining error bars on information estimates (or more precisely, their exact distribution). This is particularly important because information estimates have positive bias⁴⁷. To obtain correct estimates of the bias and standard errors of the estimates, different resampling techniques can be used^{29,48}. Estimation of the upper frequency limit of information transmitted is similar to estimation of the upper frequency limit of phase locking calculated for the owl auditory system, where very high frequency limits, and therefore spike precision, occur⁴⁹.

In general, the limiting temporal accuracy of stimulus encoding might not equal the limiting spiking resolution. This might occur when nonlinear decoding²⁰ or the direct entropy method¹⁵ is used to estimate the lower information bound. In such cases, one can test for spiking precision by repeating information calculations for a range of time windows. The information should increase as the window size is made smaller until it plateaus. If particular care is taken to correct for bias, this particular time window represents the spike timing resolution. This approach revealed spike timing resolutions of roughly one millisecond¹⁵. That result, however, does not determine whether this fine spiking resolution is used for temporal encoding, that is, whether it carries additional information beyond that required to characterize the dynamics of the stimulus. One way to answer this question is by looking at encoding in the frequency domain and estimating whether higher-frequency components used in a lower-bound estimation (with nonlinear decoding filters) could carry additional information that is not present in the lower-frequency components²⁰.

The studies reviewed here demonstrate that both fine spike-timing resolution and high reliability are found in peripheral neurons that encode dynamic stimuli. These results highlight the importance of each spike to the neural code. However, an example of temporal encoding for dynamic stimuli has not yet been found. On the other hand, temporal encoding for stimuli with very slow dynamics (usually presented as static stimuli) has been shown both in single neurons²¹ and in neuronal ensembles^{23,24,52-54}. In ensembles, synchronized activity encodes 'what' aspects of the stimulus that were completely absent in joint consideration of the firing-rate estimate at time scales corresponding to those of the stimulus presentation. It remains to

be seen how including the dynamics of natural stimuli would affect these results in single neurons, and whether precise spike timing could be used to simultaneously encode not only 'when', but also 'what'.

ACKNOWLEDGEMENTS

We are grateful to B. Bialek, A. Doupe, F. Gabbiani and J. Haag for comments on the manuscript.

RECEIVED 23 DECEMBER 1998; ACCEPTED 23 SEPTEMBER 1999

- Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4746 (1992).
- Gabbiani, F., Metzner, W., Wessel, R. & Koch, C. From stimulus encoding to feature extraction in weakly electric fish. *Nature* **384**, 564–567 (1996).
- Parker, A. J. & Newsome, W. T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
- Shannon, C. E. The mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
- MacKay, D. N. & McCulloch, W. S. The limiting information capacity of a neuronal link. *Bull. Math. Biophys.* **14**, 127–135 (1952).
- Stein, R. B., French, A. S. & Holden, A. V. The frequency response, coherence, and information capacity of two neuronal models. *Biophys. J.* **12**, 295–322 (1972).
- Eckhorn, R. & Popel, B. Rigorous and extended application of information theory to the afferent visual system of the cat. I. Basic concepts. *Biol. Cybern.* **16**, 191–200 (1974).
- Eckhorn, R. & Popel, B. Rigorous and extended application of information theory to the afferent visual system of the cat. II. Experimental results. *Biol. Cybern.* **17**, 7–17 (1975).
- de Ruyter van Steveninck, R. & Bialek, W. Real-time performance of movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc. R. Soc. Lond. B Biol. Sci.* **234**, 379–414 (1988).
- Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R. & Warland, D. Reading a neural code. *Science* **252**, 1854–1857 (1991).
- Theunissen, F., Roddey, J. C., Stufflebeam, S., Clague, H. & Miller, J. P. Information theoretic analysis of dynamical encoding by four identified interneurons in the cricket cercal system. *J. Neurophysiol.* **75**, 1345–1364 (1996).
- Wessel, R., Koch, C. & Gabbiani, F. Coding of time-varying electric field amplitude modulation in a wave-type electric fish. *J. Neurophysiol.* **75**, 2280–2293 (1996).
- Haag, J. & Borst, A. Encoding of visual motion information and reliability in spiking and graded potential neurons. *J. Neurosci.* **17**, 4809–4819 (1997).
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, Massachusetts, 1997).
- de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R. & Bialek, W. Reproducibility and variability in neural spike trains. *Science* **275**, 1805–1808 (1997).
- Buracas, G. T., Zador, A. M., DeWeese, M. R. & Albright, T. D. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* **20**, 959–969 (1998).
- Knudsen, E. I. & Konishi, M. Mechanisms of sound localization in the barn owl (*Tyto alba*). *J. Comp. Physiol.* **133**, 13–21 (1979).
- Simmons, J. A. Perception of echo phase information in bat sonar. *Science* **204**, 1336–1338 (1979).
- Mainen, Z. F. & Sejnowski, T. J. Reliability of spike timing in neocortical neurons. *Science* **268**, 1503–1506 (1995).
- Theunissen, F. E. & Miller, J. P. Temporal encoding in nervous systems: a rigorous definition. *J. Comput. Neurosci.* **2**, 149–162 (1995).
- McClurkin, J. W., Optican, L. M., Richmond, B. J. & Gawne, T. J. Concurrent processing and complexity of temporally encoded neuronal messages in visual perception. *Science* **253**, 675–677 (1991).
- Vaadia, E. I. *et al.* Dynamics of neuronal interactions in monkey cortex in relation to behavioral events. *Nature* **373**, 515–518 (1995).
- Laurent, G. & Davidowitz, W. Encoding of olfactory information with oscillating neural assemblies. *Science* **265**, 1872–1875 (1994).
- Laurent, G. Odor images and tunes. *Neuron* **16**, 473–476 (1996).
- Theunissen, F. E. & Doupe, A. J. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. *J. Neurosci.* **18**, 3786–3802 (1998).
- Theunissen, F. E. & Miller, J. P. Representation of sensory information in the cricket cercal sensory system. II Information theoretic calculation of system accuracy and optimal tuning curve width of four primary interneurons. *J. Neurophysiol.* **66**, 1690–1703 (1991).
- Kjaer, T. W., Hertz, J. A. & Richmond, B. J. Decoding cortical neuronal signals: network models, information estimation and spatial tuning. *J. Comput. Neurosci.* **1**, 109–139 (1994).
- Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R. & Bialek, W. Entropy and information in neural spike trains. *Phys. Rev. Lett.* **80**, 197–200 (1998).
- Thomson, D. J. & Chave, A. D. in *Advances in Spectrum Analysis and Array Processing* Vol 1. (ed. Haykin, S.) 58–113 (Prentice Hall, Englewood Cliffs, New Jersey, 1991).
- Theunissen, F. E. *An Investigation of Sensory Coding Principles Using Advanced Statistical Techniques*. Thesis, Univ. California, Berkeley (1993).
- Marmarelis, P. & Marmarelis, V. *The White Noise Approach* (Plenum, New York, 1978).
- Boer, E. & Kuyper, P. Triggered correlation. *IEEE Trans. Biomed. Eng.* **15**, 169–179 (1968).
- DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. Receptive-field dynamics in the central visual pathways. *Trends Neurosci.* **18**, 451–458 (1995).
- Eggermont, J. J., Johannesma, P. I. M. & Aertsen, A. M. H. J. Reverse correlation methods in auditory research. *Q. Rev. Biophys.* **16**, 341–414 (1983).
- Gabbiani, F. Coding of time-varying signals in spike trains of linear and half-wave rectifying neurons. *Network Comput. Neural Sys.* **7**, 61–65 (1996).
- Roddey, J. C., Girish, B. & Miller, J. P. Assessing the performance of neural encoding models in the presence of noise. *J. Comput. Neurosci.* (in press).
- Haag, J. & Borst, A. Active membrane properties and signal encoding in graded potential neurons. *J. Neurosci.* **18**, 7972–7986 (1998).
- Clague, H., Theunissen, F. & Miller, J. P. Effects of adaptation on neural coding by primary sensory interneurons in the cricket cercal system. *J. Neurophysiol.* **77**, 207–220 (1997).
- Rieke, F., Bodnar, D. & Bialek, W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory neurons. *Proc. R. Soc. Lond. B Biol. Sci.* **262**, 259–265 (1995).
- Warland, D., Landolfa, M. A., Miller, J. P. & Bialek, W. in *Analysis and Modeling of Neural Systems* (ed. Eckman, F.) 327–333 (Kluwer, Norwell, Massachusetts, 1991).
- Perkel, D. H. & Bullock, T. H. Neural coding. *Neurosci. Res. Prog. Bull.* **6**, 223–344 (1967).
- Suga, N. Cortical computational maps for auditory imaging. *Neural Net.* **3**, 3–21 (1990).
- Margoliash, D. & Fortune, E. S. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J. Neurosci.* **12**, 4309–4326 (1992).
- O'Scalaidhe, S. P., Wilson, F. A. & Goldman-Rakic, P. S. Areal segregation of face-processing neurons in prefrontal cortex. *Science* **278**, 1135–1138 (1997).
- Ferster, D. & Spruston, N. Cracking the neuronal code. *Science* **270**, 756–757 (1995).
- Abbott, L. F. Decoding neuronal firing and modeling neural networks. *Q. Rev. Biophys.* **27**, 191–331 (1994).
- Miller, G. A. in *Information theory in Psychology: Problems and Methods* 95–100 (Free Press, Glencoe, Illinois, 1955).
- Golomb, D., Hertz, J., Panzeri, S., Richmond, B. & Treves, A. How well can we estimate the information carried in neuronal responses from limited samples? *Neural Comput.* **9**, 649–665 (1997).
- Koppl, C. Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnoocellularis of the barn owl, *Tyto alba*. *J. Neurosci.* **17**, 3312–3321 (1997).
- Warland, D. K., Reinagel, P. & Meister, M. Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.* **78**, 2336–2350 (1997).
- Roddey, J. C. & Jacobs, G. A. Information theoretic analysis of dynamical encoding by filiform mechanoreceptors in the cricket cercal system. *J. Neurophysiol.* **75**, 1365–1376 (1996).
- Gray, C. M., König, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
- Abeles, M., Bergman, H., Margalit, E. & Vaadia, E. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.* **70**, 1629–1638 (1993).
- DeCharms, R. C. & Merzenich, M. M. Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* **381**, 610–613 (1996).