

Signal transcoding by nonlinear sensory neurons: Information-entropy maximization, optimal transfer function, and anti-Hebbian adaptation

FRANÇOIS CHAPEAU-BLONDEAU AND FRÉDÉRIC RAGUIN

*Faculté des Sciences, Université d'Angers,
2 Boulevard Lavoisier, 49000 Angers, France*

[Received 10 September 1996 and in revised form 28 October 1996]

A principle of information-entropy maximization is introduced in order to characterize the optimal representation of an arbitrarily varying quantity by a neural output confined to a finite interval. We then study the conditions under which a neuron can effectively fulfil the requirements imposed by this information-theoretic optimal principle. We show that this can be achieved with the natural properties available to the neuron. Specifically, we first deduce that neural (monotonically increasing and saturating) nonlinearities are potentially efficient for achieving the entropy maximization, for any given input signal. Secondly, we derive simple laws which adaptively adjust modifiable parameters of a neuron toward maximum entropy. Remarkably, the adaptation laws that realize entropy maximization are found to belong to the class of anti-Hebbian laws (a class having experimental groundings), with a special, yet simple, nonlinear form. The present results highlight the usefulness of general information-theoretic principles in contributing to the understanding of neural systems and their remarkable performances for information processing.

Keywords: neuron; sensory processing; signal transcoding; information entropy; entropy maximization; anti-Hebbian adaptation.

1. Introduction

The information-processing abilities of neurons in the nervous system are currently under intense investigation. Theoretical and mathematical modelling is one approach that can contribute to this endeavour. The most common methodology for the modelling of neural systems is to start with a collection of experimental observations and then try to elaborate a model which will account for these observations. The basic assumptions of such models usually have their origin in the biophysical and biochemical properties of the neurons. An example of this approach is the celebrated Hodgkin–Huxley neuron model (Hodgkin & Huxley, 1952).

We choose here to adopt another, much less common, methodology for the modelling of neural systems and ultimately of their information-processing abilities. We start with an abstract principle, which we express in information-theoretic terms, and which guarantees an optimal representation of information by neurons. We then draw the consequences of this principle, and show that properties known to belong to neurons come out as deductions from the principle.

We consider the case of sensory neurons, in charge of the transduction or transcoding of signals from the external world, at the entrance of a neural system. Sensory neurons can

convey information about external stimuli in the form of an output firing frequency at which output spikes (action potentials) are emitted in response to a stimulus. Such an output firing frequency is bound to vary in a finite interval, between no firing and a finite maximum firing rate accessible to the neuron. For an arbitrarily varying stimulus, the neuron has to adapt its response in order to realize an efficient transcoding that makes optimal use of the finite interval of representation on the output. We introduce here a principle of information-entropy maximization to characterize the condition for optimality of this transcoding. We then examine under which conditions a neuron can fulfil the requirements imposed by this information-theoretic principle, and show that this can be achieved with the natural properties of the neuron.

In particular, we shall derive adaptation laws by which a neuron can modify some of its parameters to realize optimal signal transcoding in the sense of our information-theoretic principle, and we shall point out the connections of these adaptation laws with known properties of neural adaptation. An account of the many aspects of adaptation in biological neurons has been given by Kandel (1991). Briefly, neural adaptation consists of changes that take place in the neuron, under the influence of the signals exchanged during activity. Adaptation mainly concerns the efficacy of the synapses, but it can also affect other parameters like the threshold of activity or the responsiveness of the neuron (Zucker, 1989). At the level of a neuron, adaptation occurs in a local manner, under the control of the signals received or emitted by the neuron under consideration. A common form is Hebbian adaptation (Brown *et al.*, 1990), in which the changes are driven by the joint action of a presynaptic and a postsynaptic activity, or of an input and an output activity, related to a given neuron, and with a notion of accumulation or building-up of the effect over time. The time scales of the changes involved in this type of adaptation can vary from a few tens of milliseconds up to several seconds or minutes, and sometimes longer (Brown *et al.*, 1990; Zucker, 1989; von der Malsburg, 1994). This general category of Hebbian adaptation encompasses numerous specific forms, that may differ in details, especially when one tries to cast them into mathematical formulations (Brown *et al.*, 1990). Such mathematical forms are usually constructed in an a priori manner, based on empirical arguments or experimental observations. In contrast, we shall see that our information-theoretic optimal principle applied to neural-signal transcoding, allows us to *deduce*, as consequences, adaptation laws of a Hebbian character.

2. Entropy of the neural signals

A model is considered for a peripheral sensory neuron excited by a pool of sources or sensors with sufficient homogeneity to justify the definition of a single scalar input signal X . Typically, X gives an image of a physical quantity of the environment, uniform on the sensors feeding the neuron. We consider X as a continuous random variable with a probability density function (p.d.f.) $p_X(x)$. The neuron produces a nonlinear continuous output response Y of p.d.f. $p_Y(y)$. This response Y is interpreted here as a firing rate, since we assume that a firing rate is an appropriate and natural variable to convey and conceive the information dealt with by sensory neurons. The information entropy $H(X)$ for X is defined as (Shannon, 1948)

$$H(X) = - \int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx , \quad (1)$$

and a similar definition applies for the entropy $H(Y)$ of Y .

The neuron output Y , representing a firing rate, is naturally limited to a finite range $[y_{\min}, y_{\max}]$. As a consequence, the output entropy $H(Y)$ reaches its overall maximum for the uniform p.d.f.:

$$p_Y^{\text{opt}}(y) = \begin{cases} \frac{1}{y_{\max} - y_{\min}} & \text{for } y \in [y_{\min}, y_{\max}], \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

And this absolute maximum of $H(Y)$ is:

$$H_{\max} = \log(y_{\max} - y_{\min}) . \quad (3)$$

Maximizing the entropy of the output Y is, in some sense, an optimal use of the finite interval of representation $[y_{\min}, y_{\max}]$. We now propose that the sensory neuron, when operating in an environment which imposes an input X with given $p_X(x)$, seeks to implement a transcoding from X to Y that aims at maximizing $H(Y)$, thus optimizing the neural representation of the signal on the output.

The neuron input-output transformation from X to Y is expected to be a nonlinear function, monotonically increasing from y_{\min} to y_{\max} , that we write as $Y = \varphi(X)$. The neuron receives an input with a given p.d.f. $p_X(x)$, and we ask whether it is possible to find a monotonically increasing neural function φ that allows the output Y to reach the absolute maximum of its entropy in (3).

The conservation of probabilities in the transformation $Y = \varphi(X)$ reads: $p_Y(y)dy = p_X(x)dx$, which yields

$$p_Y(y)\varphi'(x) = p_X(x) . \quad (4)$$

For $H(Y)$ to reach the absolute maximum in (3) requires that $p_Y(y)$ be $p_Y^{\text{opt}}(y)$ of (2). Equation (4) thus gives

$$\varphi'(x) = (y_{\max} - y_{\min})p_X(x) , \quad (5)$$

which integrates into

$$\varphi(x) = y_{\min} + (y_{\max} - y_{\min}) \int_{-\infty}^x p_X(u)du . \quad (6)$$

The expression in (6) which involves the distribution function $\int_{-\infty}^x p_X(u)du$ of X , represents a monotonically increasing and bounded function φ that realizes the absolute maximum of the output entropy $H(Y)$. This result demonstrates that for any input p.d.f. $p_X(x)$ there exists an input-output transformation of a *neural-type* (a monotonically increasing and saturating nonlinearity) that is able to reach the absolute maximum of the output entropy. Such a remarkable outcome suggests that sensory neurons possess input-output nonlinearities which can be made especially efficient for the optimal transcoding of signals.

This property and its information-theoretic interpretation have already been reported by Laughlin (1981, 1987). Laughlin observed that the input-output transfer function of the large monopolar cells in the blowfly compound eye, is very close to the cumulative probability distribution of its input, leading to a uniform, and hence an entropy, output

p.d.f. that is maximum. Laughlin did not give an adaptive or constructive mechanism for the neurons to reach this property; we shall do this below.

In actuality, a neuron nonlinearity, while expected to remain monotonically increasing and bounded, is only adaptable to a limited extent, through specific parameters such as threshold, slope, or gain. As a consequence, the optimal transfer function of (6), although representing a monotonically increasing bounded function, may not be exactly realizable, for any $p_X(x)$, through adaptation of the nonlinearity of a given neuron.

In the following we show that, for realistic input p.d.f.s $p_X(x)$, common models of neuron nonlinearities can be adapted to come very close to the absolute maximum entropy of (3), and, furthermore, this is with adaptation laws which display an anti-Hebbian character.

We note that (4) leads to the following expression for $H(X)$:

$$-\int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx = -\int_{-\infty}^{+\infty} p_Y(y) \log[p_Y(y)\varphi'(x)] dy, \quad (7)$$

from which we deduce that the transformation of the entropy in the input-output transcoding can be expressed as

$$H(Y) = H(X) + H_{\text{trans}}, \quad (8)$$

where H_{trans} , which represents the entropy change in the transcoding, is

$$H_{\text{trans}} = \int_{-\infty}^{+\infty} p_X(x) \log[\varphi'(x)] dx. \quad (9)$$

Expressions (8) and (9) for the transformation of the entropy by the neuron are a special instance of the change of entropy which occurs in a change of coordinates, as presented in general form by Shannon (1948). In the present neuronal context, with a given input p.d.f. $p_X(x)$, $H(X)$ is fixed, and maximization of $H(Y)$ by the neuron is equivalent to maximization of the H_{trans} of (9).

3. Adaptive maximization of the output entropy

A simple form of neuron transfer function $Y = \varphi(X)$ was considered by Chapeau-Blondeau (1994). We consider here the general form

$$\varphi(x) = f[\beta(wx - \theta)], \quad (10)$$

where f is any monotonically increasing and bounded function, for instance a logistic sigmoid $f(u) = y_{\text{max}}/[1 + \exp(-u)]$ or an error function $f(u) = y_{\text{max}} \operatorname{erf}(u)$. We note that the natural conditions for a neuron output interpreted as a firing rate are $y_{\text{min}} = 0$ (no firing) and $y_{\text{max}} > 0$ (maximum firing). Equation (10) incorporates 'natural' parameters to further shape the neuron transfer function. We assume first in this section that $\beta > 0$ is a fixed gain for the nonlinearity. The threshold θ and the efficacy w of the synapse which applies the input onto the neuron (or equivalent synaptic efficacy for a homogeneous bundle of pathways carrying the same input) are both adaptable parameters.

Now, with the transfer-function model of (10), we wish to adjust the naturally adaptable parameters θ and w in order to maximize the output entropy $H(Y)$ in the presence of any given input p.d.f. $p_X(x)$. Equivalently, we seek to adapt θ and w to maximize the

$H_{\text{trans}}(\theta, w)$ which results from the combination of (9) and (10). We choose to perform a gradient ascent along $H_{\text{trans}}(\theta, w)$, where the parameters are adapted according to the dynamic scheme (with $\eta > 0$)

$$\frac{d\theta}{dt} = \eta \frac{\partial H_{\text{trans}}}{\partial \theta}, \quad \frac{dw}{dt} = \eta \frac{\partial H_{\text{trans}}}{\partial w}. \quad (11)$$

The computation of the components of the gradient yields

$$\frac{\partial H_{\text{trans}}}{\partial \theta} = - \int_{-\infty}^{+\infty} p_X(x) \beta \frac{f''[\beta(wx - \theta)]}{f'[\beta(wx - \theta)]} dx, \quad (12)$$

$$\frac{\partial H_{\text{trans}}}{\partial w} = \frac{1}{w} + \int_{-\infty}^{+\infty} p_X(x) \beta x \frac{f''[\beta(wx - \theta)]}{f'[\beta(wx - \theta)]} dx. \quad (13)$$

An illuminating simplification arises in these adaptation laws when one adopts for the transfer function f a form which is very frequently used in neural modeling (Antón *et al.*, 1992), that is, when one chooses the logistic sigmoid $f(u) = y_{\text{max}}/[1 + \exp(-u)]$, whose derivatives verify $f'(u) = f(u)[1 - f(u)/y_{\text{max}}]$ and $f''(u) = f'(u)[1 - 2f(u)/y_{\text{max}}]$. In such conditions, (12) and (13) reduce to

$$\frac{\partial H_{\text{trans}}}{\partial \theta} = -\beta \left(1 - \frac{2}{y_{\text{max}}} \bar{Y} \right), \quad (14)$$

$$\frac{\partial H_{\text{trans}}}{\partial w} = \frac{1}{w} + \beta \left(\bar{X} - \frac{2}{y_{\text{max}}} \bar{X}\bar{Y} \right), \quad (15)$$

where, for an arbitrary function $g(X)$ of X , we introduce the mean $\bar{g} = \int_{-\infty}^{+\infty} g(x) p_X(x) dx$.

According to (14), the maximum of H_{trans} is associated with $\bar{Y} = y_{\text{max}}/2$. Thus, at the maximum of the output entropy $H(Y)$, the mean value of the output Y is located in the middle of the interval of representation [$y_{\text{min}} = 0, y_{\text{max}}$]. This outcome is a natural property, which comes here as a direct consequence of our principle of entropy maximization. If we introduce the fluctuations about the means $\tilde{X} = X - \bar{X}$ and $\tilde{Y} = Y - y_{\text{max}}/2$, the adaptation laws (11) of the neural parameters take the form

$$\frac{d\theta}{dt} = \eta \beta \frac{2}{y_{\text{max}}} \overline{\tilde{Y}}, \quad (16)$$

$$\frac{dw}{dt} = \eta \left(\frac{1}{w} - \beta \frac{2}{y_{\text{max}}} \overline{\tilde{X}\tilde{Y}} \right). \quad (17)$$

It is then remarkable to observe that the adaptation equations (16) and (17) have very simple forms. In particular, the synaptic adaptation law (17) falls in the class of Hebbian adaptation laws, since it involves the concomitant activities of both the input and output. Such Hebbian adaptation laws have been studied for a long time to account for adaptation mechanisms in the nervous systems and they have received experimental groundings (Brown *et al.*, 1990). It is noticeable that such a law, with familiar form, arises here as a consequence of an abstract information-theoretic principle instead of being introduced in

an ad hoc manner. More specifically, the present adaptation law (17) is of the covariance type, since it involves input and output signals through the average product of the deviations from their means, and it can be referred to as anti-Hebbian because of the minus sign in $-\overline{XY}$. Comparable anti-Hebbian laws have already been obtained as a consequence of other general information-processing principles, for instance, to decorrelate output signals in a network of noisy linear neurons with lateral inhibition (Plumbley, 1993). Anti-Hebbian adaptation arises here in another context, as a consequence of an entropy-maximization principle, especially with a nonlinear neuron model. The adaptation law itself comes also under a different and nonlinear form for the adapted parameter w . Such outcomes highlight the usefulness of general information-theoretic principles in contributing to an understanding of neural systems, by assigning roles to known basic mechanisms in relation to the global information-processing performances of these neural systems.

In the case where f is a logistic sigmoid, it can be easily proved by explicit examination of its first and second derivatives that $H_{\text{trms}}(\theta, w)$ of (9)–(10) is a convex function with a single maximum. The adaptation laws of (16) and (17), derived from a gradient ascent, converge to this maximum, which also achieves the maximization of the output entropy $H(Y)$. The maximum entropy so reached is the maximum achievable entropy in the presence of the constraint of a neuron transfer function under the form (10) with f being a logistic sigmoid. This constrained maximum will generally differ from the absolute maximum entropy of (3), which would require, in the presence of a given $p_X(x)$, the transfer function φ of (6). The difference can, however, be expected to be small in general, for not too special $p_X(x)$, because both the absolute φ of (6) and the optimal logistic sigmoid are monotonically increasing functions onto the same interval $[y_{\min}, y_{\max}]$. It will generally be possible to shape the optimal logistic sigmoid to come close to the absolute φ of (6), and to realize an entropy maximization under the constraint of a specific type of transfer function, coming close to the absolute maximum entropy.

For illustration, we consider the case of a Gaussian input p.d.f. $p_X(x)$ of mean m_X and variance σ_X^2 . With this input, to reach the absolute maximum of $H(Y)$ of (3) requires, according to (6), a neuron transfer function $\varphi(x) = y_{\max} \text{erf}[(x - m_X)/\sigma_X]$ (in this paper,

$$\text{erf}(u) = \frac{\int_{-\infty}^u \exp(-u'^2/2) du'}{\sqrt{2\pi}}$$

defines the error function), and this achieves the absolute maximum $H(Y) = H_{\text{max}} = \log(y_{\text{max}})$. The neuron transfer function φ of (10) with a logistic sigmoid for f cannot be exactly shaped, through adaptation of θ and w , to perfectly reproduce the error function which saturates $H(Y)$. Yet, adaptation of θ and w according to (16) and (17) brings φ very close to the optimal error function realizing the absolute maximum $H_{\text{max}} = \log(y_{\text{max}})$, as is attested by the results of Fig. 1. In Fig. 1, with $y_{\text{max}} = 1$, one has $H_{\text{max}} = 0$ bit, and the adaptive procedure of (16)–(17) reaches $H(Y) = -0.0135$ bit. In the present context, the numerical values of $H(Y)$ are interpretable as an index of the departure of $p_Y(y)$ from uniformity. For $[y_{\min} = 0, y_{\text{max}} = 1]$, uniform $p_Y(y)$ yields $H(Y) = H_{\text{max}} = 0$ bit. Suppose that, over $[y_{\min} = 0, y_{\text{max}} = 1]$, $p_Y(y)$ was constant only over $[0, 0.9]$ and zero over $]0.9, 1]$, one would then have $H(Y) = -0.152$ bit. Such quantitative results provide a basis to assess the value $H(Y) = -0.0135$ bit obtained by the adaptive procedure, which can be considered very close to the absolute maximum H_{max} . Figure 1 shows the correspond-

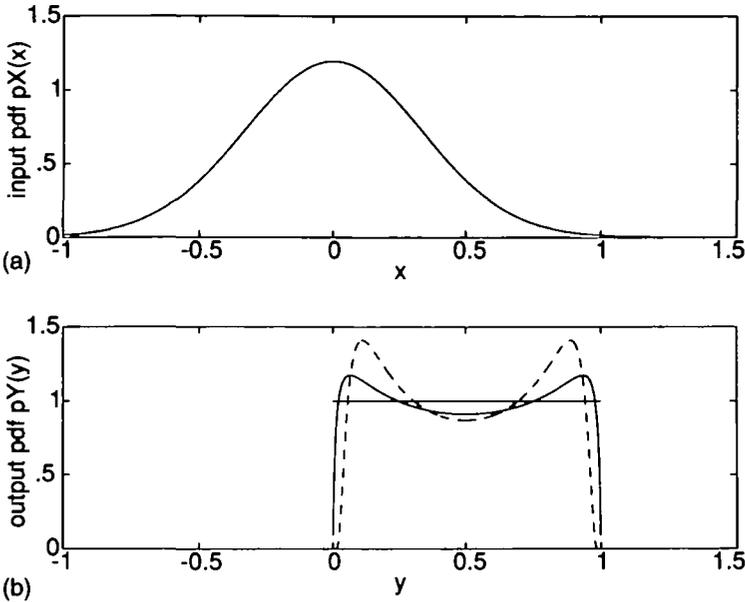


FIG. 1. Neural transcoding of a Gaussian input. (a) The Gaussian p.d.f. $p_X(x)$, with mean $m_X = 0$ and standard deviation $\sigma_X = 1/3$, on the neuron input; and (b) the output p.d.f. $p_Y(y)$ after transmission onto $[y_{\min} = 0, y_{\max} = 1]$ by the neuron of gain $\beta = 1$, with three different transfer functions $Y = \varphi(X)$. The solid straight line is the uniform p.d.f. obtained when $\varphi(x)$ is the error function resulting from (6) with the Gaussian input; this p.d.f. achieves the absolute maximum $H(Y) = H_{\max} = 0$ bit for the output entropy. The solid curve is obtained with $\varphi(x) = f(w_{\text{opt}}x - \theta_{\text{opt}})$ when f is the logistic sigmoid $f(u) = 1/[1 + \exp(-u)]$ and ($\theta_{\text{opt}} = 0, w_{\text{opt}} = 5.247$) is the fixed point of the anti-Hebbian gradient ascent of (16)–(17); this p.d.f. achieves $H(Y) = -0.0135$ bit which represents the maximum which can be reached for $H(Y)$ when a Gaussian input is transformed by a logistic sigmoid. The dashed curve is obtained with $\varphi(x) = f(w^*x - \theta^*)$ when $f(u) = 0.5(1 + u/\sqrt{1+u^2})$ and ($\theta^* = 0, w^* = 2.751$) is the zero point of (16)–(17), which, in this case, only approximates the gradient ascent along H_{trans} ; this p.d.f. achieves $H(Y) = -0.1082$ bit, which comes very close to the maximum $H(Y) = -0.1065$ bit that can be reached when a Gaussian input is transformed by the present f .

ing (small) departure of $p_Y(y)$ from uniformity. We note that this same good performance in maximizing $H(Y)$ is obtained by (16)–(17), irrespective of the parameters m_X and σ_X of the Gaussian input X ; this means that the transcoding is able to obtain the same efficiency in the output representation of any Gaussian input X . For the present example, the optimal values that maximize $H(Y)$ can be computed to be $w_{\text{opt}} = 1.749/(\beta\sigma_X)$ and $\theta_{\text{opt}} = w_{\text{opt}}m_X$, and they are accurately reached with (16) and (17).

The results of Fig. 1 illustrate that a typical neuron transfer function such as (10) with a logistic sigmoid f is very efficient at adaptively maximizing the output entropy $H(Y)$ of a neural transcoding; this occurs through simple adaptation according to (16) and (17) of naturally adjustable neuron parameters.

Now, for a function f that is not a logistic sigmoid, the adaptation laws of (16) and (17) do not achieve an exact maximization of $H(Y)$ or H_{trans} . For this one has to come back instead to the more complicated forms of (12) and (13), whose biological implementation

seems more difficult to conceive. However, for reasonable monotonically increasing functions f that are not logistic sigmoids, the simple adaptation laws of (16) and (17), used in place of (12) and (13), allow us to come very close to the maximum of $H(Y)$. This property is illustrated in Fig. 1, with the Gaussian $p_X(x)$ and $f(u) = 0.5(1 + u/\sqrt{1 + u^2})$, where (16) and (17) arrive at 0.0017 bit below the exact achievable maximum. Such results suggest that adaptation laws (16) and (17) are also efficient at increasing the entropy (improving the efficiency) of the transcoding by any neural (monotonically increasing and saturating) transfer function.

Adaptation laws (14) and (15) or (16) and (17) are driven by mean values of various functions of the input X , such as \bar{X} , \bar{Y} , or \bar{XY} . In the theoretical analysis, these mean values are defined and computed from knowledge of the input p.d.f. $p_X(x)$. In practice, with an ergodicity hypothesis, these mean values can be estimated from sample averages, performed online over one temporal realization of the signals. An online estimator will estimate the averages over a time scale T_{average} . Significant changes (say they occur over a time scale $T_{\text{adaptation}}$, controlled by η) of the adapted parameters θ and w have then to be slow relative to T_{average} . Finally, the stationarity time scale $T_{\text{stationarity}}$ of the input $X(t)$ has to be long relative to $T_{\text{adaptation}}$. The performance of an online implementation of the adaptation laws will thus depend on the quality of the estimator for the averages and on the ability to satisfy the constraints on the time scales $T_{\text{average}} < T_{\text{adaptation}} < T_{\text{stationarity}}$. For illustration, we have considered a stationary ($T_{\text{stationarity}} \rightarrow \infty$) discrete-time process $X(t) = X(n\Delta t)$ (where n is an integer), with a logistic sigmoid f for the neuron transfer function. The time averages were estimated with a first-order low-pass linear filter of time constant $\tau = 25\Delta t$, yielding $T_{\text{average}} \sim \tau$. We took an adaptation rate $\eta = 0.5$. Figure 2 then shows the resulting evolution of the neuron output Y toward maximum entropy. As can be seen in Fig. 2, this evolution is relatively fast, since after a time of $T_{\text{adaptation}} = 200\Delta t$ the theoretical maximum entropy is almost exactly reached. This verifies the possibility of a simple online realization of the adaptation laws.

It should be noted that the time scales for the present online implementation of the adaptation laws conform with the orders of magnitude that can be expected with actual sensory neurons. For actual neurons, the time step Δt can be interpreted as the smallest time at which significant changes can occur in a neural input; it can be assigned an order of magnitude of milliseconds, which represents the fastest firing period for a neuron. The time scale T_{average} can be expected to be of a few tens of milliseconds, corresponding to the time constants of the membrane processes that naturally behave as low-pass filters. Significant changes in the adapted parameters, like the synaptic efficacy, would then take place over periods $T_{\text{adaptation}}$ of a few hundreds of milliseconds, which correspond to the times needed for a sensory input to leave a print in a short-term memory process. Then follows a stationarity time $T_{\text{stationarity}}$ of the order of several hundreds of milliseconds, up to seconds, which is quite compatible with the application times required to efficiently process, in an adaptive manner, a sensory input.

4. Extension to a neuron with multiple inputs

We consider here a similar sensory neuron to that considered previously, but this time with a number $N > 1$ of inputs X_i , each one of which is mediated by the synaptic efficacy w_i .

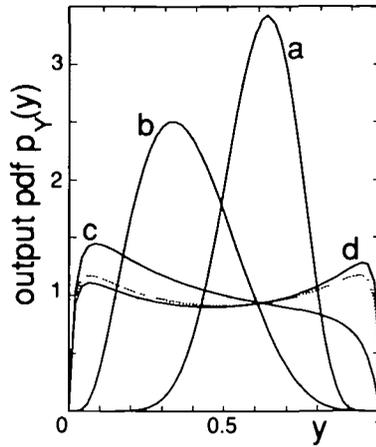


FIG. 2. Evolution of $p_Y(y)$ toward maximum entropy during an online implementation of adaptation laws (16) and (17). The neuron transfer function is with f being a logistic sigmoid onto $[y_{\min} = 0, y_{\max} = 1]$; the input p.d.f. $p_X(x)$ was chosen to be Gaussian with a mean $m_X = 1.5$ and a standard deviation $\sigma_X = 0.5$. (a) The initial state at $t = 0$ with initial values $\theta = 1$ and $w = 1$, yielding the output entropy $H(Y) = -1.125$; (b) at $t = 50\Delta t$, $H(Y) = -0.728$; (c) at $t = 100\Delta t$, $H(Y) = -0.067$; and (d) at $t = 200\Delta t$, $H(Y) = -0.015$ almost realizes the maximum entropy. (···) The theoretical realization of the exact maximum entropy $H(Y) = -0.013$ achievable with an arbitrary Gaussian input X and a logistic sigmoidal transfer function.

The neuron output is now

$$Y = f \left[\beta \left(\sum_{i=1}^N w_i X_i - \theta \right) \right], \tag{18}$$

where, as in Section 3, f is a monotonically increasing and bounded monovariate function onto $[y_{\min} = 0, y_{\max} > 0]$.

A direct transposition of the problem of Section 3 would then be to adapt the parameters w_i and θ in order to maximize the output entropy $H(Y)$. However, with multiple inputs X_i , this maximization is not sufficient to specify the w_i , because, in general, there are infinitely many N -tuples (w_1, w_2, \dots, w_N) that maximize $H(Y)$. This can be illustrated with the case where f is a logistic sigmoid, and the membrane potential $V = \sum_i w_i X_i$ is a Gaussian variable of mean $\bar{V} = \sum_i w_i \bar{X}_i$ and variance $\text{var}(V) = \bar{V}^2 - \bar{V}^2$, with $\bar{V}^2 = \sum_i \sum_j w_i w_j \bar{X}_i \bar{X}_j$. From the first example of Section 3, we can infer that the conditions on V that maximize $H(Y)$ are $\theta = \bar{V}$ and $\text{var}(V) = (1.749/\beta)^2$. When the statistics of the X_i are known, these two conditions are generally insufficient to uniquely determine the w_i . For instance, with centred mutually independent X_i , the second condition is $\sum_i w_i^2 \text{var}(X_i) = (1.749/\beta)^2$, which can be verified by infinitely many N -tuples (w_1, w_2, \dots, w_N) , even with several, and possibly up to $N - 1$, of the w_i set to zero. In general, the maximization of $H(Y)$ only constrains the statistics of V and not those of the individual $w_i X_i$.

Therefore, in general, with multiple inputs, the maximization of the output entropy $H(Y)$ through individual adaptation of the synapses w_i does not appear to be an efficient

procedure for sensory processing, since it can succeed even when the transmission over many input pathways is set to zero. We note, however, that the maximization of $H(Y)$ remains a desirable feature in this context, because it still corresponds to an optimal use of the finite interval of representation on the output. An alternative possibility, for which the present treatment is useful, is to operate with the w_i fixed, and adapt the neural gain β and the threshold θ in order to maximize $H(Y)$. For this maximization, we perform a gradient ascent along $H_{\text{trans}}(\theta, \beta)$. For the case of the logistic sigmoid f , we obtain

$$\frac{\partial H_{\text{trans}}}{\partial \beta} = \frac{1}{\beta} - \frac{2}{y_{\text{max}}} \overline{VY}, \quad (19)$$

where $\overline{V} = V - \bar{V}$.

As before, (19), with its simple form, can be used for an adaptation law, in the style of (11) and (16)–(17), to adjust β for maximum output entropy (together with (16)). This adaptation law keeps the character of anti-Hebbian adaptation, since it is dependent upon average input and output neuron activities through the product $-\overline{VY}$.

There exists another possibility, with multiple inputs, which adapts the synapses w_i for maximum output entropy, while avoiding the possibility of several input pathways being set to zero. This consists of assuming that all the synapses are kept the same that is, $w_i = w$ for all i , or in other words that the synapses undergo the same evolution from comparable initial states. This specification ensures the same relative importance for each operational input pathway, which is a functionally reasonable property for sensory processing. Under such a condition, the membrane potential factorizes into $V = w \sum_i X_i$. With the same approach as that used in Section 3 for the maximization of $H(Y)$, we are led to an adaptation law for the synapses $w_i = w$, which replaces (17) by the form

$$\frac{dw}{dt} = \eta \frac{1}{w} \left(1 - \beta \frac{2}{y_{\text{max}}} \overline{VY} \right). \quad (20)$$

Such an adaptation again resembles what can be described as a global anti-Hebbian synaptic plasticity law.

Finally, for multiple-input neurons, we have considered the application of adaptation laws (16) and (17) in fully interconnected neural networks which receive no additional inputs from the exterior of the network. Discrete-time dynamics are used for the evolution of neuron i , in the form

$$Y_i(t + \Delta t) = f \left[\beta \left(\sum_{j=1}^N w_{ij} Y_j(t) - \theta_i \right) \right] \quad (i = 1 \text{ to } N), \quad (21)$$

where f is a logistic sigmoid.

Now N^2 initial values for the w_{ij} are randomly drawn with uniform probability in $[-1, 1]$. In such networks, the time evolutions of the outputs $Y_i(t)$ can have several forms: they can converge to fixed points or limit cycles, or they can display unsteady activities of chaotic appearance, where the $Y_i(t)$ keep wandering over a subregion of their possible range of variation (and where sensitive dependence on the initial conditions can be observed) (Chapeau-Blondeau, 1993; Chapeau-Blondeau & Chauvet, 1992).

Then, for each neuron i , (16) was used to adapt the threshold θ_i . Also, (17) was used to adapt the synaptic efficacy w_{ij} on neuron i from neuron j , with Y_j playing the role of the input signal. The averages of the type \bar{X} were estimated online, with a first-order low-pass linear filter with a time constant $\tau \gg \Delta t$.

In such a mode of operation, there is no guarantee that the output entropies $H(Y_i)$ will be strictly maximized. However, we observed that the result of (16) and (17) is generally to increase the output entropies $H(Y_i)$, and to have the neuron outputs Y_i cover more evenly their possible range of variation. An application of rules (16) and (17) for a short time was usually found to be sufficient to destabilize any fixed point in which the Y_i may be sitting (thus with minimal output entropies $H(Y_i)$), and to drive the network into an unsteady chaotic evolution where the Y_i wander in time, while covering large portions of their possible range of variation (thus achieving high output entropies $H(Y_i)$).

Such a result suggests that adaptation laws (16) and (17) may be efficient to 'stimulate' the collective activities in a neural network (preventing them from sitting in fixed points), and to produce chaotic evolutions. Such chaotic evolutions have been observed in neural networks, both in theoretical models and in experimental preparations, and they may play an important role in cognitive processes by generating a large variability in the neural responses, although these implications have not yet been fully elucidated (Skarda & Freeman, 1987; Freeman, 1992; Chapeau-Blondeau, 1995).

5. Discussion

The results reported here describe connections between neural processing and information theory. We have introduced a principle of information-entropy maximization, to optimize the representation of an arbitrarily varying quantity on a neuron output whose variation is confined to a fixed finite interval. We have found that a neuron nonlinearity, monotonically increasing with saturation, is especially suited to optimal transduction of signals. We have deduced adaptation laws, (16) and (17), that we propose as efficient laws for the optimization of the transcoding of an input by a sensory neuron. For a neuron with multiple inputs, similar laws, given by (19) and (20), were also deduced. These adaptation laws modify naturally adjustable neural parameters toward maximum output entropy. These laws are found to be linear in the driving term which involves an input-output covariance; this confers an anti-Hebbian character to them. At the same time, the adaptation laws are nonlinear in the adapted parameter (w or β). This nonlinearity makes them a special class of nonlinear anti-Hebbian adaptation laws, in contrast to the linear adaptation laws which are much more commonly used. Yet, these nonlinear adaptation laws keep a simple form which could be compatible with a physical implementation in biological systems.

Various applications of information-theoretic concepts to neural transmission have previously appeared (see Taylor & Plumbley, 1993, and the references therein). These addressed different problems such as optimal information preservation in the presence of noise (Linsker, 1988), or redundancy reduction through realization of statistical independence between different correlated neural pathways (Atick, 1992; Barlow, 1989). Also, when the neuron output is modelled by a continuous variable interpreted as a firing frequency, these previous works essentially developed results for linear neurons. An important contribution is the work by Linsker (1988), which uses a principle of mutual-

information maximization (the Infomax principle), between linear noisy neurons, and which deduces linear Hebbian (opposite to anti-Hebbian) adaptation laws.

In contrast, here we have considered a regime where the noise is negligible, and thus it plays no role. In such a case, the neuron input-output mutual information is infinite, and cannot serve as a characterizing tool. We have focused on output entropy instead of mutual information, and this information-theoretic tool is applied here to the problem of the optimal neural representation of analog signals. As a result, we deduced anti-Hebbian adaptation laws of a specific nonlinear form.

Another specific feature of the present work is that it deals with the more realistic situation of continuous neurons with a nonlinear response, instead of the simpler linear response. To date, there has been little application of information-theoretic concepts to nonlinear neurons. A detailed study has been presented by Nadal & Parga (1994) who discussed the relationships between different information-theoretic principles in applications to nonlinear neurons, but this study did not derive adaptive or constructive mechanisms like (12) and (13), or (16) and (17), to adjust the modifiable neural parameters in order to fulfil the information-theoretic constraints. Chapeau-Blondeau (1994) introduced only the special version of (12) and (13), where f was a logistic sigmoid, for a single-input neuron, and the case of a multiple-input neuron was not addressed. Bell & Sejnowski (1995) used the special case where f is a logistic sigmoid, and in this condition the case of multiple inputs was treated in a way which differs from that presented in Section 4; then a subsequent interesting application in signal processing was offered, for blind separation and blind deconvolution.

The present results apply a principle of entropy maximization for the optimal representation of an analog stimulus by a single neuron output. An interesting extension would be to apply this principle to the representation of an analog stimulus by a whole population of neurons, to address another coding scheme, known as population coding, which has been observed in neural systems (Zohary, 1992). In this coding scheme, the analog stimulus is coded by the global activities of all the neurons of a population, offering a wider dynamic range of representation than for single-neuron coding. One could then seek to deduce the types of interaction and adaptation between neurons which are required to obtain an efficient representation which maximizes the entropy of the population.

6. Conclusions

We have introduced an abstract information-theoretic principle of entropy maximization to characterize the optimal representation of a signal by a neuron. We have then deduced that the fulfilment of this principle can be achieved by conditions which are naturally found in neural systems, namely, nonlinear neuron transfer functions that are monotonically increasing with saturation, and adaptation laws for the modifiable neural parameters with an anti-Hebbian form. Such a modelling procedure deduces from general principles. As we have seen, it is useful to contribute to assigning roles to known neural mechanisms, and it complements adequately other types of modelling procedures which construct from basic elements.

At a global level, neural systems have to be recognized as natural information-processing systems. A full understanding of these systems has to go beyond the strictly biophysical and biochemical levels, to complement them by descriptions involving information-processing

schemes. The present results demonstrate the usefulness of general information-theoretic principles to contribute, in this direction, to the understanding of neural systems and their remarkable performances for information processing.

REFERENCES

- ANTÓN, P. S., GRANGER, R., & LINCH, R., 1992. Temporal information processing in synapses, cells, and circuits. In: *Single Neuron Computation* (T. McKenna, J. Davis, & S. F. Zornetzer, eds.), pp. 291–313. San Diego: Academic Press.
- ATICK, J. J., 1992. Could information theory provide an ecological theory of sensory processing? *Network: Comput. Neural Syst.*, **3**, 213–51.
- BARLOW, H. B., 1989. Unsupervised learning. *Neural Comput.*, **1**, 295–311.
- BELL, A. J., & SEJNOWSKI, T. J., 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.*, **7**, 1129–59.
- BROWN, T. H., KAIRISS, E. W., & KEENAN, C. L., 1990. Hebbian synapses: Biophysical mechanisms and algorithms. *Ann. Rev. Neurosci.*, **13**, 475–511.
- CHAPEAU-BLONDEAU, F., 1993. Analysis of neural networks with chaotic dynamics. *Chaos, Solitons and Fractals* **3**, 133–9.
- CHAPEAU-BLONDEAU, F., 1994. Information entropy maximization in the transmission by a neuron nonlinearity. *Comptes Rendus de l'Acad. Sci. Sér. II, Paris* **319**, 271–6.
- CHAPEAU-BLONDEAU, F., 1995. Information processing in neural networks by means of controlled dynamic regimes. *Acta Biotheoretica* **43**, 155–67.
- CHAPEAU-BLONDEAU, F., & CHAUVET, G., 1992. Stable, oscillatory, and chaotic regimes in the dynamics of small neural networks with delay. *Neural Net.* **5**, 735–43.
- FREEMAN, W. J., 1992. Tutorial on neurobiology: From single neurons to brain chaos. *Int. J. Bifurcat. Chaos* **2**, 451–82.
- HODGKIN, A. L., & HUXLEY, A. F., 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–44.
- KANDEL, E., 1991. *Principles of Neural Science*. New York: Prentice-Hall.
- LAUGHLIN, S., 1981. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch C* **36**, 910–12.
- LAUGHLIN, S., 1987. Form and function in retinal processing. *Trends Neurosci.* **10**, 478–83.
- LINSKER, R., 1988. Self-organization in a perceptual network. *IEEE Comput.* **21**, 105–17.
- NADAL, J. P., & PARGA, N., 1994. Nonlinear neurons in the low-noise limit: A factorial code maximizes information transfer. *Network: Comput. Neural Syst.* **5**, 565–81.
- PLUMBLEY, M. D., 1993. Efficient information transfer and anti-Hebbian neural networks. *Neural Net.* **6**, 823–33.
- SHANNON, C. E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 623–56.
- SKARDA, C. A., & FREEMAN, W. J., 1987. How brains make chaos in order to make sense of the world. *Behav. Brain Sci.* **10**, 161–95.
- TAYLOR, J. G., & PLUMBLEY, M. D., 1993. Information theory and neural networks. In: *Mathematical Approaches to Neural Networks* (J. G. Taylor, ed.). Amsterdam: Elsevier.
- VON DER MALSBURG, C., 1994. The correlation theory of brain function. In: *Models of Neural Networks II: Temporal Aspects of Coding and Information Processing in Biological Systems* (E. Domany, J. L. van Hemmen, & K. Schulten, eds.), pp. 95–119. New York: Springer-Verlag.
- ZOHARY, E., 1992. Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biol. Cybernet.* **66**, 265–72.
- ZUCKER, R. S., 1989. Short-term synaptic plasticity. *Ann. Rev. Neurosci.* **12**, 13–31.