Variance normalization: a key mechanism for temporal adaptation in natural vision?

Marco Buiatti , Carl van Vreeswijk

Neurophysique et Physiologie du Système Moteur - CNRS UMR 8119, Université René Descartes, Paris, France

Abstract

Recent experiments suggest that retinal neurons adapt their response not only to the temporal mean but also to the temporal variance of the visual input. Inspired by these results, we propose a simple model in which temporal adaptation can be achieved by a transformation consisting of a linear filtering followed by a variance normalization. We show that such transformation efficiently adapts to the temporal statistics of natural time series of intensities by removing almost all of its redundancy, while linear transformation alone cannot achieve the same goal. Results reproduce important features of temporal adaptation in the retina.

Key words: Adaptation, Temporal coding, Natural images, Redundancy

1 Introduction

The natural input of the visual system is characterized by light intensities that span intermittently a very wide range of values. A basic problem that the visual system faces is how to code such a widely varying input with the finite dynamic range of its neurons. Neurophysiological studies have shown that the visual system attempts to solve this problem with a variety of mechanisms of adaptation at all levels of the visual pathway [12]. Nevertheless, one can ask: is there a global computational strategy underlying all these adaptation mechanisms? In a recent revisitation of his work, Barlow [3] proposed that the visual system attempts to build an efficient representation of its input by converting the hidden redundancy abundant in natural scenes into a manifest, explicit, immediately recognizable form by separating it from the unpredictable part of the input. Following this theoretical approach, many

Email address: Buiatti@biomedicale.univ-paris5.fr (Marco Buiatti).

groups have obtained interesting results by calculating the (usually linear) transformation that optimally decorrelates the spatial structure of the natural visual input, and comparing it with the experimentally measured response of the visual system [1, 9, 10].

The purpose of this work is to apply this theoretical approach to investigate the optimal coding strategy in the time domain. The main idea, previously introduced under the name of *predictive coding* [9], is that neurons can filter out the redundant part of their input by predicting it from the history of the same input in the past. In this way, they only have to transmit the unpredictable part of the input, which spans a much narrower range of values and is therefore easier to encode.

In order to train our model on realistic natural data, we use a set of 12 time series of natural intensities (TSNI) recorded by van Hateren [11] and freely accessible on his web site (http://hlab.phys.rug.nl/tslib/index.html). Each time series lasts 45 minutes and was recorded by a photodetector that has a spectral sensitivity and an angular resolution comparable to that of human vision, and a temporal resolution of 1.2 kHz. The optical system was mounted on a headband and worn by a freely walking person. To correct for the real eyes' dynamics, the subject wore marked glasses, and was told to keep the markers at a fixed position in the visual field. Recordings were made by different subjects walking in various environments under various weather conditions.

As expected from the variability of natural visual scenes, TSNI span a range of more than 8 log-units in an intermittent way: intensities may lie within a narrow range for minutes and then suddenly jump to another range. This behavior is reflected in the variability of shape and width of the log-intensity distributions among different time series. Moreover, TSNI are very redundant: their autocorrelation function is characterized by a slow, piece-wise power-law like decay. Such long-range correlations reflect a well-known property of natural images: they are highly redundant both in time and space [5].

2 The model: theory and results

The first step to build an optimal coding mechanism consists in transmitting the difference between the current log-intensity, $x(t)$ [1], and its best *linear* prediction estimated over its past values. The transmitted signal is then the

[1] We consider as input to our model the logarithm of the light intensity because we know that this is what photoreceptors approximately signal to the rest of the retina. Time t is a discrete variable taking the values $t_k = k\Delta t$, where k is an integer and $\Delta t = 0.8333$ msec is the time interval between two consecutive data recordings.

prediction error, $y(t)$, given by

$$y(t) = x(t) - \sum_{k=1}^{T_a} a(k)x(t - k\Delta t), \tag{1}$$

where $a(k)$ is the optimal linear filter and T_a its length. We estimate $a(k)$ by minimizing the average square prediction error, $\epsilon = \langle y^2(t) \rangle$, on the TSNI. This minimization can be performed analytically and, for $T_a \to \infty$, it removes second-order correlations of $y(t)$ by construction [7]. Figure 1 shows that the filter obtained by minimizing ϵ over all time series is characterized by the first few components, while it essentially vanishes for $k\Delta t \geq 10$ msec. This means that the removal of second-order correlations requires only a short time of integration, of the order of 10 msec.

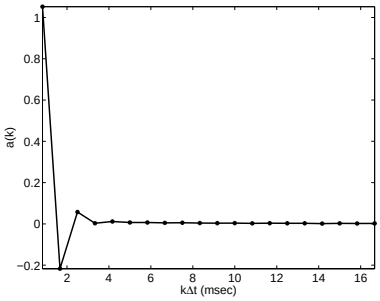

Fig. 1. Linear filter resulting from minimization of the average square prediction error on all time series. The length of the filter is 40 time steps corresponding to $T_a \simeq 33$ msec.

As it is evident from Fig. 2B, the distribution $P(y)$ is characterized by a high narrow central peak and long tails. This is not a very good distribution if the dynamic range of the output and its resolution are finite: most of the range would be devoted to code for a very small part of the input signal. A possible explanation of the sparseness of $P(y)$ could be the fact that the local variance is not constant, but changes over time. Thus, a second coding step removing correlations in the variance of y could solve the problem of adaptation.

Following this suggestion, we write y as $y(t) = \sqrt{D(t)} \cdot \xi(t)$, where $D(t)$ indicates the local variance and $\xi(t)$ is an uncorrelated variable with zero mean and unit variance. Starting from this hypothesis, in order to remove the residual redundancy, we implement the second step of our transformation by normalizing the response by an estimate of the local standard deviation

$$r(t) = \frac{y(t)}{\sqrt{D_{est}(t)}}. \tag{2}$$

The estimate $D_{est}(t)$ is obtained by performing a weighted sum of the square of the filtered signal $y(t')$ at times $t' < t$

$$D_{est}(t) = \sum_{k=1}^{T_b} b(k) \cdot y^2(t - k\Delta t), \tag{3}$$

where T_b is the length of the filter $b(k)$. The optimal filter $b(k)$ is obtained by minimizing the kurtosis of r, defined as $\kappa_r(t) = \langle r^4(t)\rangle / \langle r^2(t)\rangle^2 - 3$. Setting a constraint on the power of the output activity ($\langle r^2(t)\rangle = 1$), it is straightforward to show that kurtosis' minimization amounts to minimizing the difference between $D(t)$ and its estimate $D_{est}(t)$. Minimization was implemented by gradient descent on the components of the filter $b(k)$. Since the time scale of the optimal filter turned out to be much longer than the one of the first filter, to avoid overfitting we parameterized the filter as $b(k) \propto (k\Delta t)^{-\beta} \cdot e^{-\gamma k \Delta t}$. The filter obtained by minimizing the kurtosis over all time series is characterized by a time scale (the transition time between the power-law and the exponential decay) of about 350 msec, significantly longer than the characteristic time scale of the linear filter. This means that a much longer time is needed to estimate the local variance that optimally normalizes the input.

Figure 2A shows the histograms of the natural input x and of the responses y and r for three segments of 1 minute each drawn from the first time series . As expected, the histograms of the natural input vary considerably in width and shape. The histograms of the response y all have a similar shape, but the width is remarkably different. Moreover, they are characterized by a high peak and long tails, suggesting that the variance varies even within just such a short period of time. In contrast, the distributions of the variance-normalized response r almost overlap, and their shape is almost Gaussian: the output r continuously adapts to the local input by coding it into a compact, invariant distribution. As shown in Fig. 2B, the distribution $P(r)$, unlike $P(y)$, remains almost Gaussian even after several hours of TSNI.

In order to evaluate whether variance normalization has effectively removed high-order correlations present in the data, we measure the mutual information between the values of the data at two different times separated by an interval τ

$$I(r, \tau) = \sum P(r(t), r(t-\tau)) \log \frac{P(r(t), r(t-\tau))}{P(r(t)) \cdot P(r(t-\tau))}, \tag{4}$$

and similarly for x and y. To correct for the non-uniform shape of the distribution of the three variables over all the time series, the data are binned in K bins of variable width chosen such that a data point has equal probability to end up in any of the bins. Moreover, we correct for the finite sample effect by

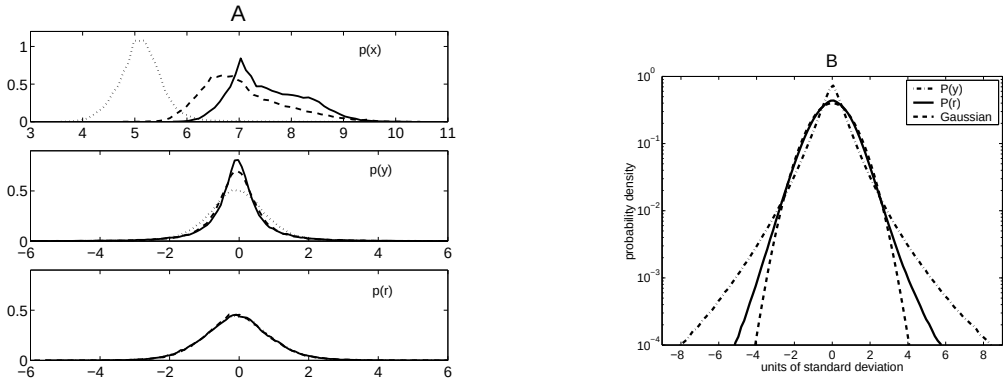

Fig. 2. A: Probability density of x, y and r for three segments of 1 minute extracted from the first time series. Segments correspond to minute 15 (solid lines), 33 (dashed lines) and 43 (dotted lines) of the first time series. The abscissa is in log-intensity units for x, and in units of standard deviation for y and r. B: Semi-logarithmic plot of the probability density of y and r for all time series, compared to a Gaussian with zero mean and unit variance.

subtracting from the mutual information estimated from the data the mutual information from randomly reshuffled data.

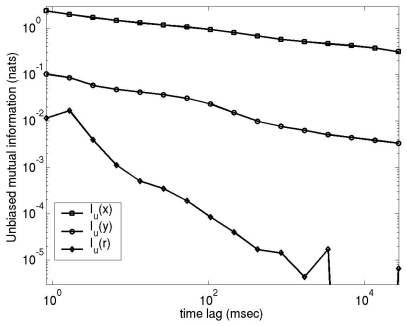

Fig. 3. Log-log plot of the unbiased mutual information versus time lag computed over all time series. The number of bins is $K = 300$.

Figure 1 shows the log-log plot of the resulting unbiased mutual information I_u for the three variables. As expected, $I_u(x, \tau)$, is very high, and decays very slowly as a power law. $I_u(y, \tau)$ is much lower, but it still decays very slowly, scaling again approximately as a power law. However, $I_u(r, \tau)$ becomes indistinguishable from zero in less than one second. The suggestion proposed above seems to be right: variance normalization allows adaptation because it largely removes not only the correlations between the fluctuations in the input, but most of the pair-wise redundancy at any order.

3 Discussion

Despite its generality, the model accounts for some major features of adaptation in the retina. In particular, it reproduces the main features of the activity of LMCs in the fly in response to the same set of natural time series, as shown in van Hateren [11]. The output distribution is almost invariant with respect to the input, its shape is very close to a Gaussian (compare the Fig. 2A with Fig. 4 in [11]), and the redundancy is almost completely removed (the power spectrum of the LMCs' activity is almost flat). Another important result concerns the time scales of adaptation. Several works found that variance normalization occurs in the vertebrate and invertebrate retina [2, 6] finding that the response completely adapts to a change in input variance within one second. This is reasonably close to the time scale of our non-linear filter, that is of the order of 350 msec.

Variance normalization has already been suggested as an important neural mechanism in the spatial domain of the visual system and the frequency domain of the auditory system [8]. Our results in the time domain of the visual system reinforce the idea that variance normalization could be a very general and crucial strategy of neural coding of sensory stimuli. An extended version of these results has been previously published [4].

References

[1] J. Atick and A. Redlich. *Neural Computation*, 4:196–210, 1992.

[2] S. A. Baccus and M. Meister. *Neuron*, 36:909–919, 2002.

[3] H. Barlow. *Network: Computation in Neural Systems*, 12:241–253, 2001.

[4] M. Buiatti and C. van Vreeswijk. *Vision Research*, 43:1895–1906, 2003.

[5] D. Dong and J. Atick. *Network: Computation in Neural Systems*, 6:345–358, 1995.

[6] A. L. Fairhall, G. D. Lewen, W. Bialek, and R. de Ruyter van Steveninck. *Nature*, 412:787–792, 2001.

[7] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, NY, USA, 3rd edition, 1991.

[8] O. Schwartz and E. P. Simoncelli. *Nature Neuroscience*, 4:819–825, 2001.

[9] M. Srinivasan, S. Laughlin, and A. Dubs. *Proceedings of the Royal Society of London B*, 216:427–459, 1982.

[10] J. van Hateren. *Journal of Comparative Physiology A*, 171:157–170, 1992.

[11] J. van Hateren. *Vision Research*, 37:3407–3416, 1997.

[12] J. Victor. *Network: Computation in Neural Systems*, 10:R1–R66, 1999.