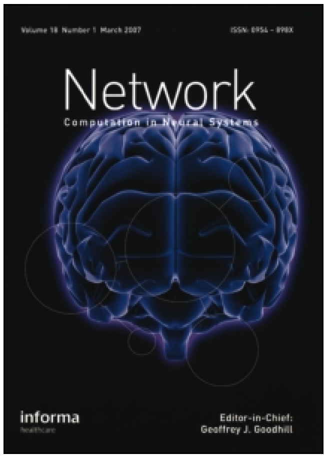

Network: Computation in Neural Systems
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713663148

V1 non-linear properties emerge from local-to-global non-linear ICA

Jesús Malo [a]; Juan Gutiérrez [b]
[a] Dept. d'Òptica, Facultat de Física, Universitat de València,
[b] Dept. d'Informàtica, Escola Técnica Superior d'Enginyeries, Universitat de València,

Online Publication Date: 01 March 2006
To cite this Article: Malo, Jesús and Gutiérrez, Juan (2006) 'V1 non-linear properties emerge from local-to-global non-linear ICA ', Network: Computation in Neural Systems, 17:1, 85 - 102
To link to this article: DOI: 10.1080/09548980500439602
URL: http://dx.doi.org/10.1080/09548980500439602

Network: Computation in Neural Systems
March 2006; 17: 85–102

V1 non-linear properties emerge from local-to-global non-linear ICA*

JESÚS MALO[1] & JUAN GUTIÉRREZ[2]

[1]*Dept. d'Òptica, Facultat de Física, Universitat de València, and* [2]*Dept. d'Informàtica, Escola Técnica Superior d'Enginyeries, Universitat de València*

(Received 30 May 2005; revised 4 October 2005; accepted 23 October 2005)

Abstract
It has been argued that the aim of non-linearities in different visual and auditory mechanisms may be to remove the relations between the coefficients of the signal after global linear ICA-like stages. Specifically, in Schwartz and Simoncelli (2001), it was shown that masking effects are reproduced by fitting the parameters of a particular non-linearity in order to remove the dependencies between the energy of wavelet coefficients. In this work, we present a different result that supports the same efficient encoding hypothesis. However, this result is more general because, instead of assuming any specific functional form for the non-linearity, we show that by using an unconstrained approach, masking-like behavior emerges directly from natural images. This result is an additional indication that Barlow's efficient encoding hypothesis may explain not only the shape of receptive fields of V1 sensors but also their non-linear behavior.

Keywords: *Human vision, natural image statistics, ICA, wavelets, non-linear systems, incremental thresholds, gain control, divisive normalization*

Introduction

The different specific formulations of Barlow's idea about the match between perception and the statistics of natural signals (Barlow 1961, 2001) have explained a variety of perception facts (Simoncelli & Olshausen 2001; Simoncelli 2003). In its original formulation; Barlow (1961) suggested that the goal of low-level perception mechanisms is to obtain a set of statistically independent responses. Recently, Barlow (2001) generalized the relevance of statistics in perception to include more abstract (Bayesian) representations of information about the environment. Nowadays, there is a a very productive debate about the generality of the original efficient encoding hypothesis (Barlow 2001; Simoncelli 2003).

However, despite the (eventually) restricted applicability of the original efficient encoding formulation, it has had remarkable success in explaining the specific shape of linear V1 receptive fields (Olshausen & Field 1996). Moreover, in Schwartz and Simoncelli (2001), the same idea has been used to obtain the parameters of a specific functional form of the non-linearities found in V1: the divisive normalization (Heeger 1992).

In this work, we show that when applying a particular (local-to-global) non-linear ICA, which assumes no specific functional form for the non-linearity, masking-like non-linearities

*This work was partially supported by the projects CICYT TIC2003-01504 and Grupos04-08 (GV).
Correspondence: Jesús Malo, Dept. d'Òptica, Facultat de Física, Universitat de València. Dr. Moliner 50, 46100 Burjassot, València, Spain. Tel: +34 963544717; Fax: +34 963544715; E-mail: Jesus.Malo@uv.es

ISSN: 0954-898X print / ISSN 1361-6536 online
DOI: 10.1080/09548980500439602

are also reproduced. This suggests that the statistical effect of the non-linearities would be to adapt the global representation of the previous linear stage (global linear ICA) to the local features of the signal around each observed (masking) stimulus. These results are an additional indication that the original efficient encoding formulation of Barlow's ideas may explain not only the linear behavior of V1, but also its non-linearities. The difference in the result presented here with regard to previous implementations of the efficient encoding idea (Schwartz & Simoncelli 2001; Kayser et al. 2003) is that no explicit functional form is assumed for the non-linearity, thus revealing more clearly that it emerges from the specific probability density function (PDF) of natural images.

The structure of the paper is as follows. In the next section, we review the relations between local linear ICA representations and the more general non-linear ICA representations (Lin 1999; Karhunen et al. 2000; Hyvärinen et al. 2001; Jutten & Karhunen 2003), the specific local-to-global non-linear ICA used in this paper is derived from these relations. Following this, we apply the above concepts to derive an integral expression for the non-linear part of the current V1 model. This expression assumes no specific functional form for the non-linearity. After this, we show how to use the proposed framework to reproduce response curves as derived in psychophysics. There after, we briefly review the set of masking non-linear effects to be statistically reproduced. Then, we show that when using the proposed framework with natural images, different masking behaviors are reproduced. Finally, we present the conclusions of the work.

Local-to-global non-linear ICA

The basic data model used in (global) linear ICA assumes that the n-dimensional samples of the signal, the column vectors $\mathbf{x}$, come from a mixture of m independent sources, column vectors $\mathbf{r}$, given by the (global) mixing matrix, $\mathbf{A_g}$ (Hyvärinen et al. 2001), where the subscript $\mathbf{g}$ stands for *global*, i.e., using all the data:

$$\mathbf{x} = \mathbf{A_g} \cdot \mathbf{r} = \sum_{j=1}^{m} \mathbf{A}_{\mathbf{g}_j} \mathbf{r}_j. \tag{1}$$

The column vectors in the mixing matrix, $\mathbf{A}_{\mathbf{g}_j}$, $j = 1, \cdots, m$, are the basis vectors of ICA. Assuming that the mixing matrix of the model is unknown, the independent components are found by determining a linear inverse mapping (separating matrix), $\mathbf{W_g}$, so that the m-vector

$$\mathbf{r} = \mathbf{W_g} \cdot \mathbf{x} \tag{2}$$

is an estimate of the independent component vector. A number of different algorithms have been proposed to compute $\mathbf{W_g}$ (Hyvärinen et al. 2001). For simplicity, the number of independent components, m, is usually assumed to be equal to the dimension of the data, so the mixing matrix is simply the inverse of the separating matrix, $\mathbf{A_g} = \mathbf{W_g}^{-1}$.

In spite of its usefulness, the basic data model assumed in Equation 1 is often too simple to describe real-world data. First, the standard model is *linear*, while a suitable *non-linear* model is usually needed to adequately represent the data. Second, the ICA model in Equation 1 is *global*, i.e., it tries to describe all the data using the same global features (ICA basis vectors, $\mathbf{A}_{\mathbf{g}_j}$). However, natural data often has different *local* features in different parts of its global PDF. This is why non-linear ICA methods (Hyvärinen et al. 2001; Jutten & Karhunen 2003) and local linear ICA methods (Karhunen et al. 2000) have been developed.

General non-linear ICA techniques (Hyvärinen et al. 2001; Jutten & Karhunen 2003) assume that the data vectors, $\mathbf{x}$, depend non-linearly on some statistically independent components, $\mathbf{r}$,

$$\mathbf{x} = F(\mathbf{r}) \tag{3}$$

and thus, the problem is to obtain the inverse mapping

$$\mathbf{r} = F^{-1}(\mathbf{x}) \tag{4}$$

that estimates the independent components. The problems with this generic formulation are, (1) it is computationally expensive, and (2) it is highly non-unique (Jutten & Karhunen 2003): i.e., multiple inverse mappings that fulfill the condition of giving rise to vectors $\mathbf{r}$ with independent components may be obtained.

Local linear methods fall in between global linear techniques that are too rigid; and non-unique non-linear techniques that are too demanding. Local linear approaches start by finding clusters in the global PDF, and then, a different *local* linear model is applied to the data in each cluster. Local PCA (Kambhatla & Leen 1997) and local ICA (Karhunen et al. 2000) have been used within this pre-clustering framework. The clustering part of these techniques is the way to break down the (difficult) global non-linear problem into a set of (easier) local subproblems where the linear approach is valid enough, however, as a result, an arbitrary number of disconnected local linear representations is obtained.

In this section, we propose a method to solve this problem by integrating the local descriptions of the data into a single global description. If the non-linear mapping in Equation 3 is smooth, the Jacobian of the non-linear mixing mapping at a particular point, $\mathbf{r}'$, (or equivalently, at the point $\mathbf{x}' = F(\mathbf{r}')$) is the local linear mixing matrix (Lin 1999; Karhunen et al. 2000), i.e., $\nabla F(\mathbf{r}') = \mathbf{A}_\ell(\mathbf{x}')$. The subscript ℓ stands for *local* and the local ICA is computed by using the data in a neighborhood of $\mathbf{x}'$.

From this result, we propose a local (differential) approach considering the Taylor expansion of F up to first order, which is valid for increments that are small enough (differential), i.e.,

$$d\mathbf{x}' = \nabla F(\mathbf{r}') \cdot d\mathbf{r}' = \mathbf{A}_\ell(\mathbf{x}') \cdot d\mathbf{r}' \tag{5}$$

accordingly,

$$d\mathbf{r}' = \nabla F^{-1}(\mathbf{x}') \cdot d\mathbf{x}' = \mathbf{W}_\ell(\mathbf{x}') \cdot d\mathbf{x}'. \tag{6}$$

A single global representation can be obtained by integrating these differential increments

$$\mathbf{r} = \mathbf{r_0} + \int_{\mathbf{x}_0}^{\mathbf{x}} \nabla F^{-1}(\mathbf{x}') \cdot d\mathbf{x}' = \mathbf{r_0} + \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{W}_\ell(\mathbf{x}') \cdot d\mathbf{x}' \tag{7}$$

where the local separating matrix at each point is computed using standard linear ICA techniques, e.g., the symmetric fixed-point algorithm (Hyvärinen 1999; Hyvärinen et al. 2001), using the K closer neighbors of the point $\mathbf{x}'$ (e.g., using Euclidean metrics in the spatial domain).

As different local separating matrices are computed at different points, a consistent criterion to order their basis vectors is needed. For instance, in order to obtain a consistent set of non-linear sensors that are tuned to specific qualitative features, a reference matrix can be taken and the basis vectors of all the other local matrices can be ordered accordingly (e.g., computing angles between the features).

Note that the proposed approach substantially differs from previously reported local approaches (Kambhatla & Leen 1997; Karhunen et al. 2000) because no clusters are used: in our case, the local neighborhoods are constructed differentially, as the response is integrated along the path. Therefore, in the proposed approach, a single (global) non-linear representation is obtained instead of an arbitrary number of disconnected local linear representations.

The local-to-global non-linear representation (a specific non-linear ICA) which is given by Equation 7 is useful only if (1) the final components of $\mathbf{r}$ are more independent than the global ICA components, and (2) the result of the integral is independent of the integration path so that no particular path has to be specified. In Appendices A and B, we show that this is the case for the problem of interest: the PDF of natural images.

V1 non-linearities and local-to-global non-linear ICA

The current model of V1 sensors involves a *linear stage* and a *non-linear stage* (Watson & Solomon 1997; Schwartz & Simoncelli 2001). First, the input images, $\mathbf{x}$, are analyzed by a set of *unit-norm* linear sensors:

$$\mathbf{c} = \mathbf{T} \cdot \mathbf{x}. \tag{8}$$

Then, a non-trivial non-linearity is applied to the outputs of the linear stage:

$$\mathbf{r} = R(\mathbf{c}). \tag{9}$$

The entire set of transforms (with the corresponding inverses) can be summarized as follows:

$$\mathbf{x} \underset{\mathbf{T}^{-1}}{\overset{\mathbf{T}}{\rightleftarrows}} \mathbf{c} \underset{R^{-1}}{\overset{R}{\rightleftarrows}} \mathbf{r}. \tag{10}$$

The challenge of statistically based vision models is to derive the experimental behavior of these transforms from the efficient encoding hypothesis together with the statistical properties of natural images.

The linear sensors can be obtained from data by looking for a linear transform that maximizes the independence between the transformed coefficients (Olshausen & Field 1996). In this work, we take this approach by computing the linear transforms from natural images, performing a global linear ICA and normalizing the basis functions to obtain unit norm sensors, i.e., $\mathbf{T} = \mathbf{W_{gu}}$ and $\mathbf{T}^{-1} = \mathbf{A_{gu}}$. We use the notation $\mathbf{W_{gu}}$ and $\mathbf{A_{gu}}$ for the global linear transforms hereafter. The subscript $\mathbf{u}$ (unit-norm) is added to stress the difference with regard to standard (global) linear ICA techniques that give rise to basis functions of different norm (Hyvärinen 1999). Note that this normalization does not modify the independence properties of the original global linear ICA transform.

As stated previously, global linear ICA techniques do not generally give rise to truly independent coefficients. Assuming the original formulation of Barlow's ideas, the role of the non-linear transform, R, would be to remove the remaining relations between the coefficients of the vector $\mathbf{c}$. Following this hypothesis, Schwartz and Simoncelli (2001) fitted the parameters of a specific model for the non-linearity (the divisive normalization (Heeger 1992)) to obtain independent responses in the vector $\mathbf{r}$. By doing so, they reproduced the experimental masking-like non-linearities. Similarly, Kayser et al. (2003) fitted the exponent of another non-linear summation model (Adelson & Bergen 1985) maximizing the temporal coherence of each neuron and the decorrelation between neurons.

In this paper, we follow the same general idea (obtaining the non-linearity, R, by looking for independent components in $\mathbf{r}$), but we assume *no specific functional form* or non-linear constraint in R by using the local-to-global ICA method described in the previous section.

To this end, we take a differential approach to locally approximate the non-linear response, R, by its Jacobian[1]. Then, given a masking stimulus, $\mathbf{x}'$, the increment in the response of the sensors given a variation of the stimulus in the pixel domain is:

$$d\mathbf{r}' = \nabla R(\mathbf{c}') \cdot d\mathbf{c}' = \nabla R(\mathbf{W_{gu}} \cdot \mathbf{x}') \cdot \mathbf{W_{gu}} \cdot d\mathbf{x}' \qquad (11)$$

so the (global) response to any stimulus, $\mathbf{x}$, is just:

$$\mathbf{r} = \mathbf{r_0} + \int_{x_0}^{x} \nabla R(\mathbf{W_{gu}} \cdot \mathbf{x}') \cdot \mathbf{W_{gu}} \cdot d\mathbf{x}'. \qquad (12)$$

Assuming that the role of the cascade of linear and non-linear transforms in Equation 10 is to obtain independent components in the coefficients of the response vector $\mathbf{r}$, the local behavior of the non-linearity could be obtained by identifying Equations 6 and 11. From this identification, we get

$$\nabla R(\mathbf{W_{gu}} \cdot \mathbf{x}') = \mathbf{W}_\ell(\mathbf{x}') \cdot \mathbf{A_{gu}} \qquad (13)$$

which when applied to Equation 12, gives rise to Equation 7. Equation 13 states that the statistical effect of V1 non-linearities would be to adapt the global representation of the previous global linear stage to the local features of the signal around each observed (masking) stimulus.

Reproducing response curves in experimental-like conditions

The non-linear properties of a sensor are (psychophysically) revealed by measuring contrast incremental thresholds of patterns that ideally isolate the response of the sensor. In experiments of this kind, the observer has to detect the presence of an incremental stimulus, $\Delta\mathbf{x}$, shown on top of a background (or masking stimulus) $\mathbf{x_0}$.

In order to analyze the response of a specific sensor, i, optimal incremental stimuli should be ideally designed to isolate its response. Therefore, the stimuli should lie in the direction that modifies a single coefficient of the response vector. In these experiments, the contrast of the incremental stimulus in the optimal direction is increased until the distortion is just noticeable. The contrast of this just noticeable optimal incremental stimulus for the i sensor, $\Delta\mathbf{x}^{(i)}(\mathbf{x_0})$, is referred to as contrast incremental threshold of the sensor i adapted to the masking stimulus $\mathbf{x_0}$. Mathematically, $\Delta\mathbf{x}^{(i)}(\mathbf{x_0})$, is the stimulus that modifies the response by $\Delta\mathbf{r}(\Delta\mathbf{x}^{(i)}) = \tau \cdot \boldsymbol{\delta}^{(i)}$, where $\delta_j^{(i)} = 0 \ \forall j \neq i$, and $\delta_i^{(i)} = 1$, and τ is the scale factor that defines the threshold size in the response domain.

Auto-masking experiments analyze the response of a specific sensor i (i.e., explore the incremental thresholds in the direction $\Delta\mathbf{x}^{(i)}$), using masking stimuli, $\mathbf{x_0}$, which only excite the same sensor, i. Cross-masking experiments explore the response of a sensor i using masking stimuli that also excite different sensors $k \neq i$.

To reproduce the non-linear effects as revealed by psychophysics, the expression of the optimal incremental stimulus that isolates the response of a single non-linear sensor can be used. Assuming the definition of the optimal distortion in the direction i, and using

Equations 11 and 13, we have

$$\Delta \mathbf{x}^{(i)}(\mathbf{x_0}) = \mathbf{A_{gu}} \cdot \nabla R^{-1}(c_0) \cdot \tau \cdot \boldsymbol{\delta}^{(i)} = \tau \cdot \mathbf{A}_\ell(\mathbf{x_0}) \cdot \boldsymbol{\delta}^{(i)}. \tag{14}$$

As expected from Equation 13, given a certain masking stimulus, $\mathbf{x_0}$, the non-linear sensor, i, is optimally excited by the column, i, of the local-linear ICA matrix, $\mathbf{A}_\ell(\mathbf{x_0})$.

The above definition of optimal incremental stimuli is convenient to simulate auto-masking and cross-masking results: note that starting from the selected masking image, if this stimulus is progressively changed according to the local optimal incremental stimulus for a particular sensor, constant increments are obtained in this single sensor, thus reproducing the ideal conditions of masking experiments. The increment in the response when modifying the stimuli in this way is $\tau \cdot \boldsymbol{\delta}^{(i)}$, as can be seen using Equation 14 as $d\mathbf{x}'$ in Equation 7. Therefore, the Michelson contrast, ΔC, of the computed incremental stimulus is the necessary contrast increment to produce a threshold variation, τ, in the response. The (variable) increments in contrast, ΔC, and the (constant) increments in the response, τ, can be used to reproduce the non-linear response of the selected sensor versus the Michelson contrast of its optimal stimulus without explicitly using Equation 7. All the response curves shown in the section 'Numerical experiments and results' were constructed in this experimentally inspired simplified way.

Perception facts to be statistically reproduced

In this section, we review the basic behavior of V1 sensors as revealed by threshold and supra-threshold psychophysics (Daugman 1980; Legge & Foley 1980; Legge 1981; Watson 1983; Harvey & Doan 1990; Foley 1994; Watson & Solomon 1997).

The first relevant fact is that V1 receptive fields are narrow band-pass filters that analyze the frequency content of the images in a log-polar wavelet-like way (Daugman 1980; Watson 1983; Harvey & Doan 1990). Beyond this linear filter-bank stage, the non-linear behavior of V1 sensors gives rise to different relevant psychophysical facts:

- *Auto-masking*: Contrast incremental thresholds of sensors that are tuned to specific frequency bands increase with the amplitude (or contrast) of the stimulus (Legge & Foley 1980; Legge 1981). This is equivalent to saying that (1) the sensitivity of the mechanisms decreases with the contrast of the optimal stimulus, or (2) the response of the mechanisms when excited by their optimal stimulus is a saturating non-linearity.
- *Cross-masking I, General behavior*: The sensitivity of a particular sensor to its optimal stimulus changes (decreases) when this stimulus is superimposed on a high contrast background of different frequency content (cross-masking) (Foley 1994; Watson & Solomon 1997). This effect increases with the contrast of the background.
- *Cross-masking II, Fine details, relative influence between frequencies*: Moreover, the above effect (or perceptual interaction between frequencies) decreases as the distance between the frequencies (in modulus or orientation) increases (Watson & Solomon 1997).

Figure 1 illustrates this set of non-linear effects as summarized in the divisive normalization model (Heeger 1992; Watson & Solomon 1997). For this particular illustration, we used the divisive normalization in the block-DCT domain with psychophysically inspired parameters (Malo et al. 2006).

In this work, we reproduce the basic wavelet-like shape of V1 receptive fields from data as in Olshausen and Field (1996). Therefore, the main contribution of this work is to reproduce the non-linear effects from natural images *assuming no parametric model.*

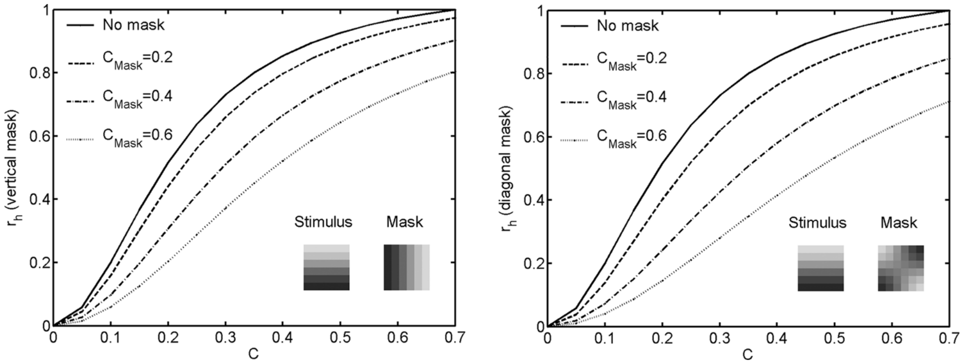

Figure 1. Illustration of non-linear effects using a divisive normalization model: compressive non-linearity in auto-masking conditions and reduction of the response in cross-masking situations.

Numerical experiments and results

In this section, we apply Equation 13 and their consequences (Equation 7 and 14) to a set of natural images. The results obtained show that the response of the sensors constructed in this way reproduces the basic behavior of V1 sensors as summarized by psychophysical parametrical models (e.g., divisive normalization in Figure 1).

Implementation details

The computation of the response for an arbitrary stimulus, $\mathbf{x}$, involves solving the definite integral in Equation 7 between a point of known response, $\mathbf{x_0}$, and the selected point. To do so (for instance, in Appendices A and B), we used a 4th order Runge–Kutta integration (Press et al. 1992) along the straight line defined by $\mathbf{x_0}$ and $\mathbf{x}$, with a number of (uniformly distributed) steps proportional to the Euclidean length of this line in the spatial domain. We used 50 integration steps for the most distant image from the mean luminance image. Note that a 4th order Runge–Kutta integration implies four Jacobian computations (and hence four local ICAs) per integration step.

In the simulations to reproduce the non-linear response of a sensor in auto-masking or cross-masking conditions, the responses for stimuli that lie in very specific directions must be computed. In these cases (e.g., the simulations in this section), we used the experimentally inspired contrast incremental threshold technique described previously using a scale factor $\tau = 1$. The computation of the response curve stops when the Michelson contrast of the current stimulus plus the incremental stimulus is greater than 1. Using $\tau = 1$, the number of iterations to arrive to the maximum contrast stimulus depends on the sensor to a certain extent, but it is typically about 30. This implies 30 local-ICA computations per response curve. This number would increase for lower values of τ. Note that using this latter procedure (valid in experimental-like conditions) one local ICA computation per iteration is needed instead of the four ICAs per step required by the Runge–Kutta algorithm.

In any of the above cases, the computational bottleneck is the speed of the algorithm to compute the local ICA basis. We are using the Matlab implementation of FastICA (Hyvärinen 1999), which converges slowly for high-dimensional vectors.

Accordingly, in the numerical experiments, we restrict ourselves to low-dimensional examples because they are a good illustration of the basic concept and appropiately suggest its possibilities at reasonable computational cost. On one hand, 6×6 image blocks (36D response vectors) were used for the (more relevant) response curve simulations. On the other hand, 6×1 one-dimensionally collected image samples were used in the (more technical and more computationally expensive) illustrative experiments shown in the appendices.

The response simulations were carried out using $2 \cdot 10^6$, 6×6 image samples randomly taken from a well known natural image database (van Hateren & van der Schaaf 1998). The original images of the database were scaled to fall in the range $[0, 255]$. The local (differential) neighborhoods around each point were determined by the 4% closest neighbors in the Euclidean sense in the spatial domain (i.e., $8 \cdot 10^4$ samples per local ICA).

In every case (response reproduction and appendices), the basis vectors (the features) in each local neighborhood were sorted according to the local basis around the zero response point (the average image) using the inner product in the spatial domain as a similarity criterion. Both sets of vectors (the reference basis and the considered basis) were normalized before inner product computation for a fair feature comparison.

Global linear features: linear V1 receptive fields

The global linear features, $\mathbf{A_{gu}}$, and their corresponding frequency content are shown in Figure 2. The functions in $\mathbf{A_{gu}}$ were sorted according to the variance of the coefficients in the transformed domain, which is equivalent to sorting them according to the norm of the basis functions in $\mathbf{A_g}$. This set of linear sensors that comes from global linear ICA would account for the empirical linear filter-bank result quoted above (Olshausen & Field 1996).

Local linear features

The local basis around the average (taken as reference to sort the other local bases) is shown in Figure 3.

Figure 4 shows the local basis around a number of illustrative points. Auto-masking and cross-masking behavior can be anticipated from the general behavior of the local basis displayed in this example. The stimuli in the top row were selected to excite just the second – low frequency, horizontal – sensor. The stimuli in the middle row were selected to excite just the third – low frequency, vertical – sensor. The stimuli in the bottom row were selected to

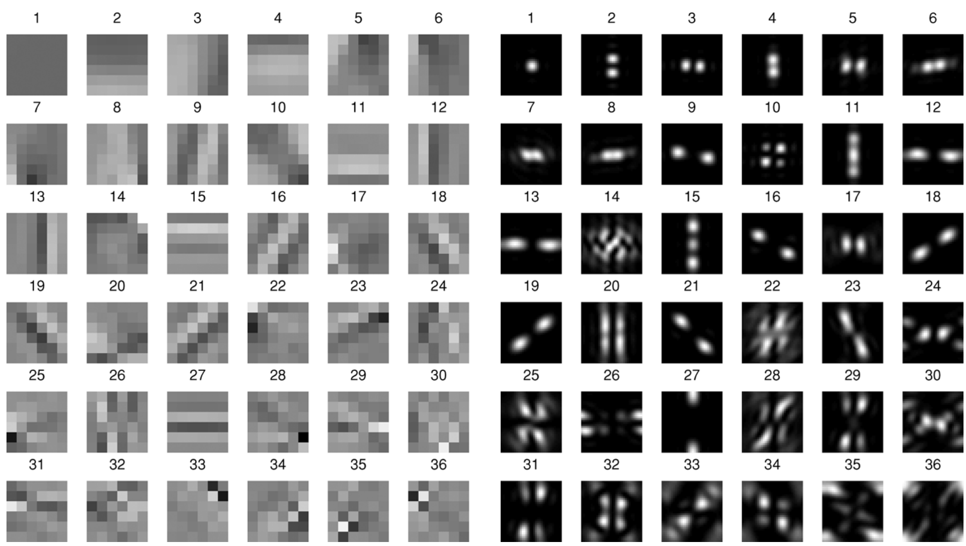

Figure 2. Basis functions of the global ICA transform in the spatial domain (left) and their frequency content (right).

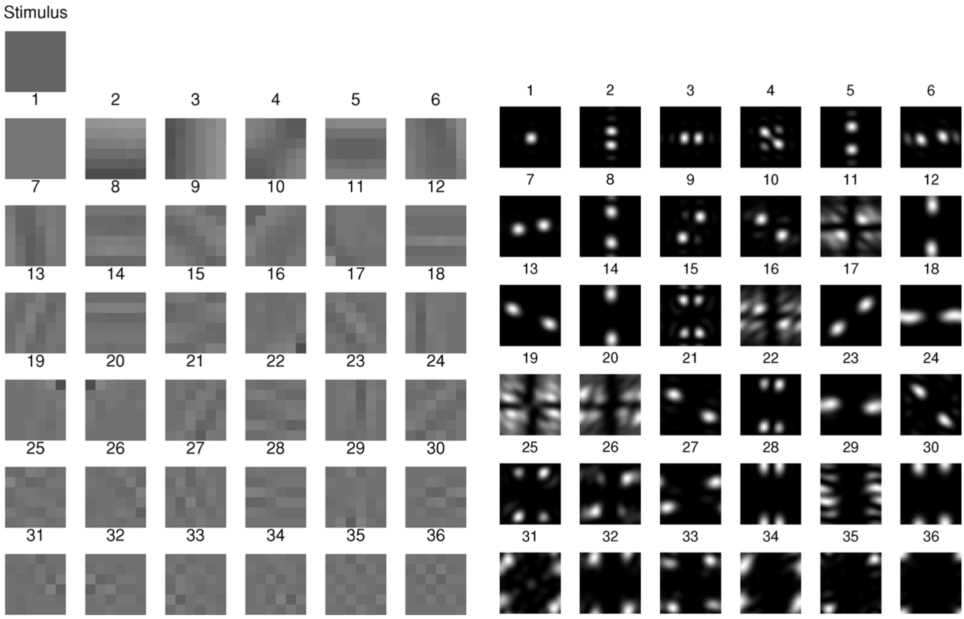

Figure 3. Local ICA basis functions around the average stimulus in the spatial domain (left) and their spectra (right). The contrast of the basis functions in the spatial domain is very small because they are represented in the same scale as the local basis functions in Figure 4. Note that the different contrast of the basis as a function of the stimulus contrast is the key issue in reproducing the non-linearities (see text).

excite the second and the third sensors simultaneously. In the last case, the stimuli were formed by adding optimal stimuli for the second (horizontal) sensor on top of masking stimuli that optimally excite the third (vertical) sensor, in such a way that the combination matches the Michelson contrast of the above stimuli, namely $C = 0.2, 0.4, 0.6$. The key issue here is the contrast of the local basis functions. As stated above, these functions are the optimal stimuli for the corresponding sensors, so their contrast is the contrast incremental threshold of the sensors.

Note that as we increase the contrast of the stimulus starting from zero (Figure 3) in a direction that is optimal for a specific sensor (e.g., top and middle rows in Figure 4), the contrast of the optimal stimuli or basis functions (the contrast incremental thresholds) also increase. This is true not only in the selected direction (which implies auto-masking), but also in all the other directions (which implies cross-masking). This is also the case in more general situations, e.g., when simultaneously exciting two sensors (bottom row in Figure 4). In this case, the contrast increment in these directions is even larger (see basis functions 2 and 3 in the example).

Reproducing auto-masking

In this section, we analyze the contrast of the optimal incremental stimuli in the direction i shown on top of stimuli that only excite that sensor using the procedure described earlier.

Figure 5 shows the auto-masking results for a number of sensors that are tuned to horizontal, vertical and diagonal stimuli of low, medium, and high frequency (according to the convention used in Figure 3, the sensors 2, 8, 14; 3, 6, 18; and 4, 10, 13). In each plot, the responses were normalized to their maximum. In each case, we show three examples

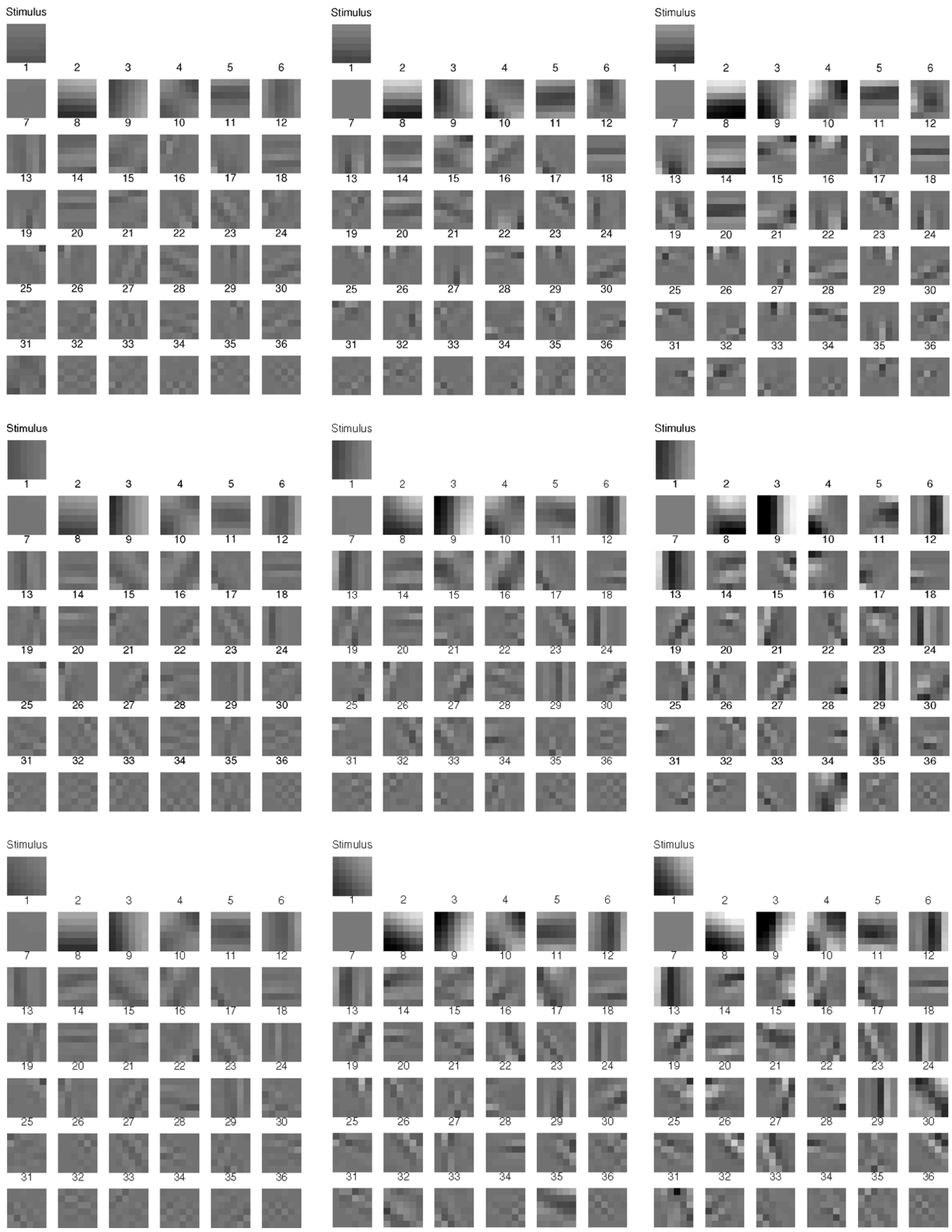

Figure 4. Local ICA basis around different stimuli in the spatial domain (see text for details).

of the optimal stimulus at different contrasts, C, and the corresponding optimal incremental stimulus, of incremental contrast, ΔC. As expected from the general behavior shown in Figure 4, these examples consistently show that the contrast incremental thresholds of the non-linear sensors increase with the contrast of the stimulus giving rise to non-linear responses as explicitly shown in the first subplot of Figure 5.

These results suggest that the proposed local-to-global non-linear ICA sensors reproduce the basic V1 non-linear behavior in auto-masking conditions (see the solid curves given by

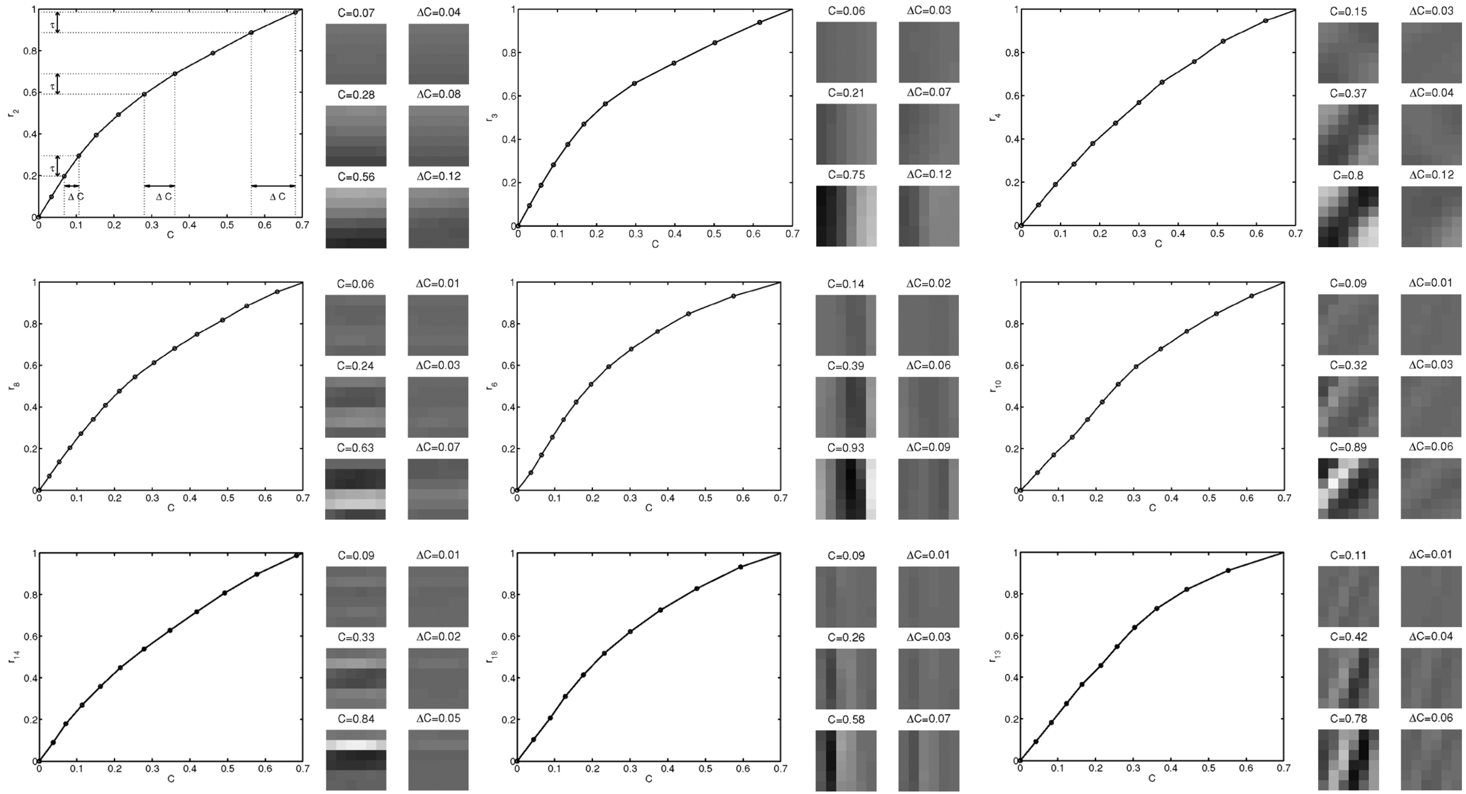

Figure 5. Non-linear responses of different sensors in auto-masking conditions. At different Michelson contrast levels, C, in the optimal direction, different contrast incremental thresholds, ΔC, are needed to generate the same increment τ.

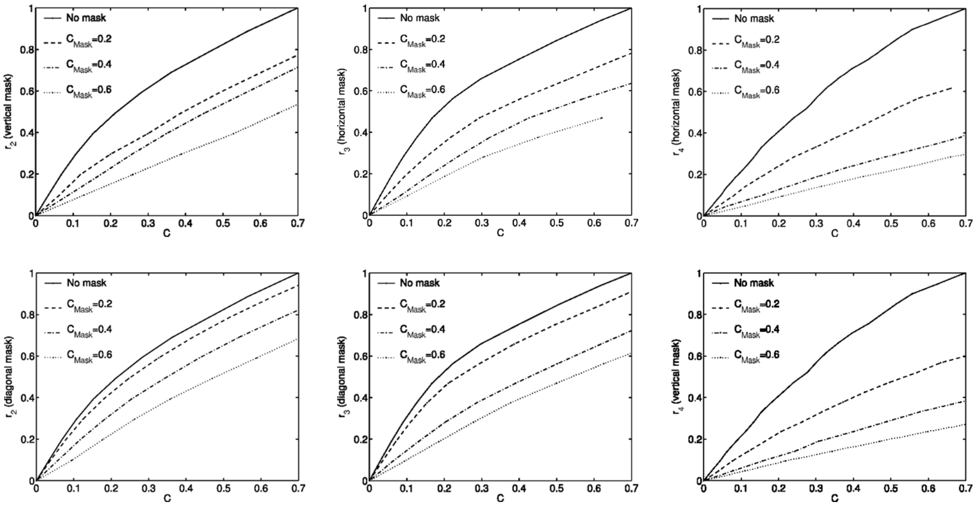

Figure 6. Non-linear response of the low-frequency horizontal (first column), vertical (second column) and diagonal (third column) mechanisms (2nd, 3rd and 4th sensors) when using different masking stimuli with different contrasts.

the psychophysically inspired parametric model in Figure 1). The proposed local-to-global non-linear ICA learns a saturating non-linearity because the radius of the neighborhood increases to compensate for the decrease in the PDF for high contrast images, giving rise to local ICA bases of different lengths.

Reproducing cross-masking

In this case, we analyze the contrast of the optimal incremental stimuli in the direction i shown on top of stimuli that excite a different sensor $k \neq i$ using the procedure described in an earlier section.

Figures 6 and 7 show the responses of different sensors in cross-masking conditions: low frequency tuned to different orientations (2nd, 3rd and 4th sensors) and horizontal sensors tuned to different frequencies (2nd, 8th and 14th sensors). Different contrasts (namely 0.2, 0.4 and 0.6) for the masking stimuli were used. The auto-masking behavior is also included for reference purposes.

The resulting responses for different masking contrast were scaled using the corresponding auto-masking response as reference. In some cases, the response does not cover the whole contrast range because the next computed stimulus (current stimulus plus incremental stimulus) exceeded the [0, 255] range.

These results show that auto-masking and the general trends of cross-masking emerge from natural data using the proposed algorithm (see the general correspondence of these figures with the examples obtained with a parametrical model, Figure 1): specifically, the response of all sensors when excited by their optimal stimulus on a uniform background (auto-masking) is a saturating non-linearity. Therefore, if these stimuli are shown on a background of different frequency (exciting a different sensor), the sensitivity is reduced and this masking effect increases with the contrast of the background stimulus.

However, the fine details of cross-masking, namely the relative interaction between frequencies (Watson & Solomon 1997), are not always reproduced in these 36D experiments:

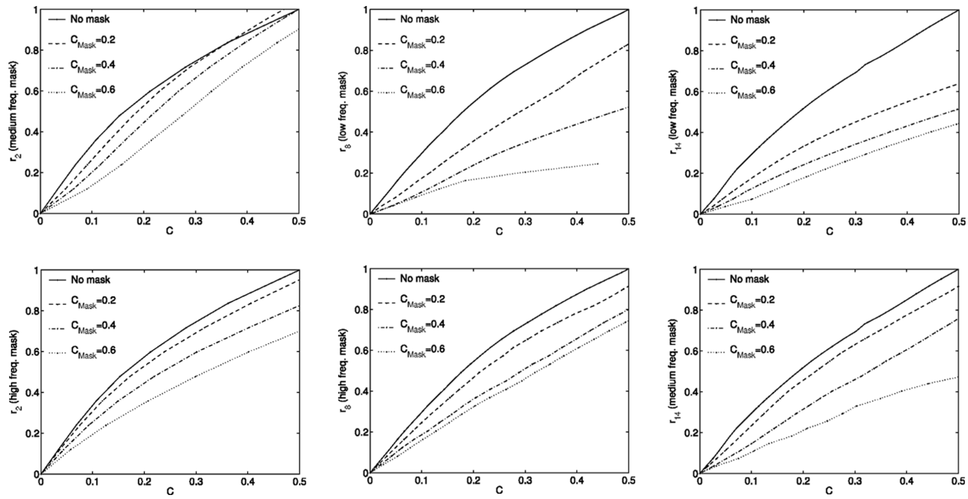

Figure 7. Non-linear response of the mechanisms tuned to horizontal orientations and low (first column), medium (second column), and high (third column) frequency (2nd, 8th and 14th sensors) when using different masking stimuli with different contrasts.

note, for instance, that in the results presented the diagonal mask induces lower response reduction in the horizontal sensor than the vertical one (see the discrepancy in the relative amount of masking between the first column in Figure 6 and the equivalent result using the parametric psychophysical model, Figure 1). This may be due to the low dimensionality of the samples considered in the experiments: the limited spatial extent of the samples gives rise to small resolution in the 2D frequency domain in such a way that the effective distance between the all linear sensors is small so the relative interaction between them is quite similar.

Conclusions and further work

In this work, we derived a non-parametric expression for the non-linear V1 behavior based on a novel local-to-global non-linear ICA. The proposed differential framework allows us to obtain the global non-linearities of V1 directly from the statistical properties of natural images without an explicit expression (such as the divisive normalization). In the proposed scheme, the aim of the local behavior of the non-linear sensors would be to capture the local features of natural images (local ICA axes) around each masking stimulus.

The statistically derived sensors exhibit a non-linear behavior that is comparable to that found in established models for visual neurons. These results, which are consistent with those presented in Schwartz and Simoncelli (2001), are an alternative indication that Barlow's efficient encoding hypothesis may explain not only the shape of receptive fields of V1 sensors but also their non-linear behavior. Our result is stronger since we are not assuming an (already non-linear) parametric model.

Further work should extend the proposed procedure to higher dimensions in order to address the fine details of cross-masking (relative influence between linear sensors). In this case, the use of local-PCA could be considered instead of local-ICA in the proposed framework in order make the method computationally affordable.

Acknowledgements

The authors thank Prof. H.B. Barlow for asking the right question at the GRC Meeting on Sensory Coding and the Natural Environment Malo et al. 2004: *ok, but can you obtain the non-linear response from here?* The authors also thank J. Rovira for his help in running the experiments in the appendices.

Appendix A

Efficiency of local-to-global non-linear ICA for natural images

In this appendix, we show how the proposed local linear ICA increases the independence of the coefficients obtained by a standard global linear ICA technique (Hyvärinen 1999). This result is consistent with similar results obtained for the specific non-linearity used in the divisive normalization model (Schwartz & Simoncelli 2001; Malo et al. 2006).

We compare the independence of coefficients in the standard global linear ICA domain and in the proposed local-to-global non-linear ICA domain using two measures, (1) bow-tie relations in the conditional PDFs (Schwartz & Simoncelli 2001; Hyvärinen et al. 2003), and (2) relative mutual information (Cover & Tomas 1991; Malo et al. 2006). The results for the spatial domain representation are also included for reference purposes.

In this experiment, we computed the response for 10^4 6D image samples which were randomly chosen from the $2 \cdot 10^6$ data set used above. These vectors were first linearly transformed using the global ICA separating matrix that was computed from the whole data set. Then, the non-linear response for all these points was computed using Equation 7. In the computation of the non-linear responses, all the data ($2 \cdot 10^6$ points) were taken into account to set the differential neighborhoods. According to the considerations in the 'Implementation details', this experiment involves about $6 \cdot 10^5$ local ICA computations. This is why we used a smaller dimensionality in this example.

Statistical dependence between the coefficients in a representation can be qualitatively revealed by analyzing the conditional histograms of pairs of samples of the signal in that domain (Schwartz & Simoncelli 2001; Hyvärinen et al. 2003). Figure 8 show examples of these conditional histograms for two pairs of coefficients in the spatial domain, the global linear ICA domain and in the proposed domain.

In the spatial domain case, these results indicate a strong correlation between neighboring luminance values: the average value of x_i changes with the value of its neighbor x_j. In the global linear ICA case, the 2nd order correlation has been removed; note that the average in c_i is independent of the given value of c_j. However, the variance (or energy) of global linear ICA coefficients, c_i, still depends on the value of the neighbors c_j (Schwartz & Simoncelli 2001; Hyvärinen et al. 2003). In the proposed representation, these bow-tie dependencies are almost removed, thus revealing a weaker statistical dependency between coefficients.

These qualitative results are consistent with quantitative measures that are based on mutual information. The mutual information of a set of variables, $c_1, \cdots, c_n$, is defined as the Kullback–Leibler divergence between their joint PDF and the product of their marginals. It can be computed from the marginal entropies, $H(c_i)$, and the joint entropy, $H(c_1, \cdots, c_n)$, of the variables (Cover & Tomas 1991):

$$I(c_1, \cdots, c_n) = \sum_{i=1}^{n} H(c_i) - H(c_1, \cdots, c_n). \tag{15}$$

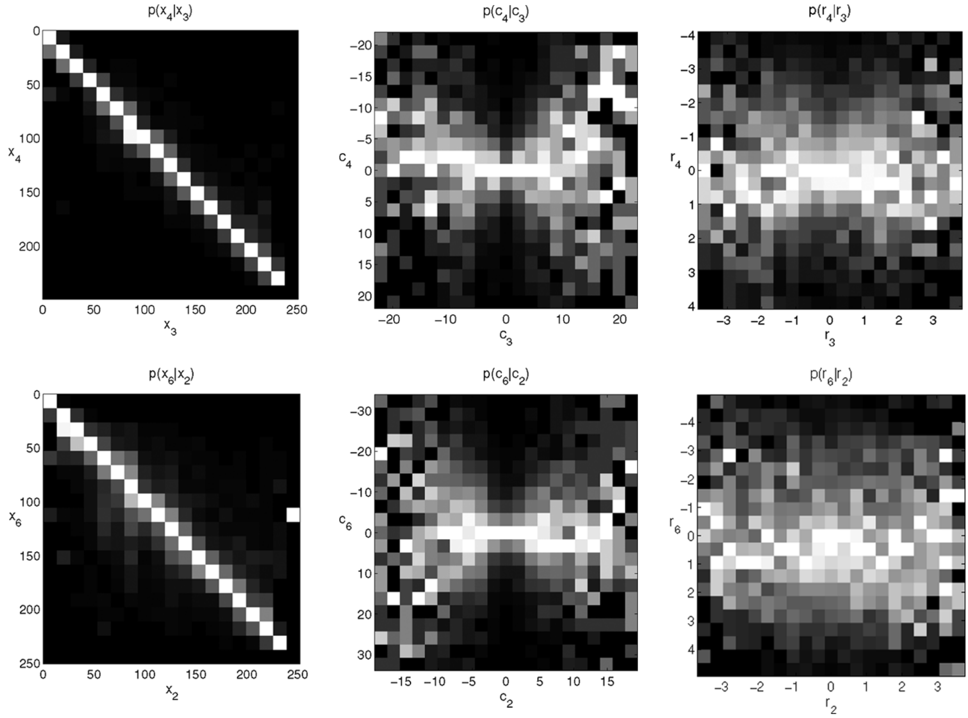

Figure 8. Examples of conditional probabilities of two pairs of coefficients in the considered domains: Spatial domain (first column), Global linear ICA (second column), local-to-global non-linear ICA (third column).

Mutual information is very useful for determining how much information (in bits) the neighbors can tell us about a specific coefficient, and hence, how strong the statistical interactions are in a specific domain. However, a specific value of mutual information can have different interpretations in terms of statistical dependency as a function of the entropy of the considered variables. This is because mutual information is affected by the shape of the marginal PDF (or, equivalently, the marginal entropy) of the coefficients of the representation. A straightforward use of mutual information, I, to assess the statistical dependency between variables in different representations will be biased in favor of representations with highly non-uniform marginal PDFs, i.e., representations where the coefficients have a small entropy.

Therefore, it is better to define a relative measure that takes the different marginal entropies into account (Malo et al. 2006). We compute the relative mutual information as

$$I_r(c_1, \cdots, c_n) = \frac{\frac{1}{(n-1)} I(c_1, \cdots, c_n)}{\frac{1}{n} \sum_{i=1}^{n} H(c_i)}. \tag{16}$$

Note that $I_r = 1$ when the c_i are fully redundant (e.g., identical), and $I_r = 0$ when they are independent.

Figure 9 shows the relative mutual information between pairs of coefficients in the three considered domains as a function of the distance between them. As expected, global linear ICA achieves a large reduction in relative mutual information with regard to the values in

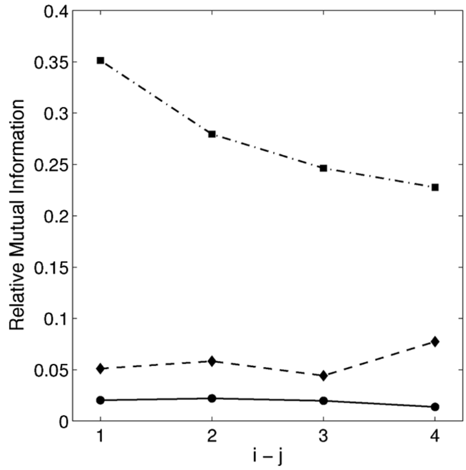

Figure 9. Relative mutual information of pairs of coefficients, (i, j), in the three considered domains as a function of the distance between coefficients, $i - j$. The spatial domain ($I_r(x_i, x_j)$, dash-dot), the global ICA domain ($I_r(c_i, c_j)$, dashed), the proposed local linear ICA domain ($I_r(r_i, r_j)$, solid).

the spatial domain. However, the proposed representation achieves an even larger reduction revealing a weaker statistical dependence between the local-to-global non-linear ICA coefficients.

These two facts (removing bow-tie dependencies and mutual information reduction) demonstrates that the proposed method increases the independence of the coefficients obtained by a standard global linear ICA technique.

Appendix B

Independence of the integration path

In this appendix, we analyze the independence of the response integral (Equation 7) on the integration path. The underlying idea of the illustration presented here is based on the standard Stokes theorem used in classical electrodynamics (Jackson 1998).

The proposed local linear ICA associates a set of vectors, $\mathbf{W}_\ell(\mathbf{x_0})$, to each point, $\mathbf{x_0}$. Thus, Equation 7 can be seen as the integration of a vector field. The Stokes theorem states that the integration of a vector field is independent of the integration path (i.e., the field is conservative) if the integral in closed paths vanishes (Jackson 1998).

In our case the results show that integration in several closed paths in different planes is almost zero. In these experiments, we used the same dimensionality, global data set, integration method and integration resolution as in Appendix A. Figure 10 shows two examples of the selected planes defined in the global ICA representation. The response was computed in two closed integration paths for each plane. The difference between these two integration paths is that the first one (closer to the origin) goes across low-contrast images in highly populated regions while in the second one the contrast is greater and there are fewer image samples per unit of volume. The response across these rectangular integration paths can be computed by adding the responses in each segment of the path, $\mathbf{r}_{ABCDA} = \mathbf{r}_{AB} + \mathbf{r}_{BC} + \mathbf{r}_{CD} + \mathbf{r}_{DA}$, as illustrated in Figure 10a.

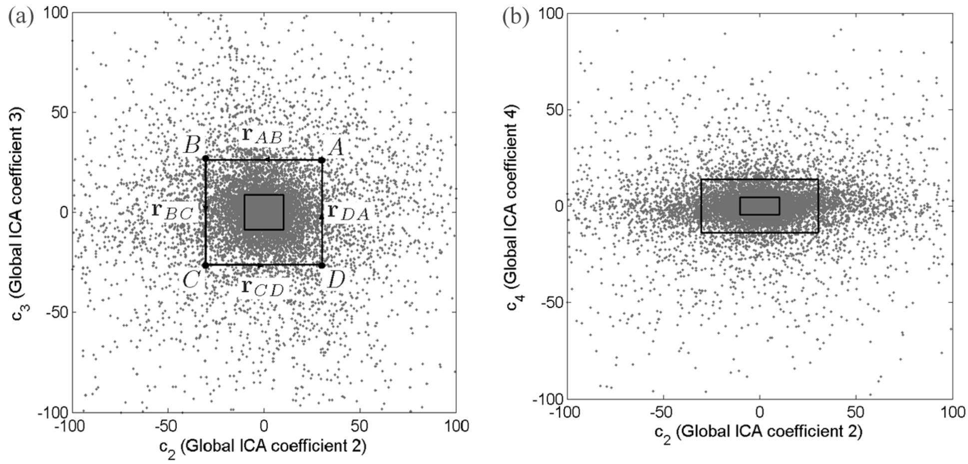

Figure 10. Examples of integration paths in different planes of the image space.

The vectors obtained for each segment in the specific path highlighted in Figure 10a, defined in the plane (c_2, c_3), are

$$\mathbf{r}_{AB} = \begin{bmatrix} -0.43 \\ -2.20 \\ -1.10 \\ 0.01 \\ -0.18 \\ -0.39 \end{bmatrix} \quad \mathbf{r}_{CD} = \begin{bmatrix} 0.59 \\ 2.00 \\ 1.10 \\ -0.05 \\ 0.19 \\ 0.33 \end{bmatrix} \quad \mathbf{r}_{BC} = \begin{bmatrix} -0.52 \\ 1.40 \\ -1.50 \\ 0.33 \\ 0.25 \\ -0.69 \end{bmatrix} \quad \mathbf{r}_{DA} = \begin{bmatrix} 0.47 \\ -1.40 \\ 1.60 \\ -0.07 \\ -0.45 \\ 0.56 \end{bmatrix} \tag{17}$$

There are a number of interesting features for the non-linear response: (1) when moving along the AB direction (reduction of the c_2 component) the greatest activity is obtained in the second sensor $\mathbf{r}_{AB_2}$; (2) however, this is not the only active sensor, revealing cross coefficient interactions in the non-linear responses; (3) when moving in the opposite direction, CD, the response vector approximately reverses, $\mathbf{r}_{CD} \approx -\mathbf{r}_{AB}$. These trends also hold for the other two segments in such a way that the final sum almost cancels.

In order to show that these responses are close to zero, we measured the norm of the response across the closed path relative to the sum of the norms along the segments:

$$n = \frac{|\mathbf{r}_{ABCDA}|}{|\mathbf{r}_{AB}| + |\mathbf{r}_{BC}| + |\mathbf{r}_{CD}| + |\mathbf{r}_{DA}|}. \tag{18}$$

Table I shows the results of n for the different planes and integration paths considered.

These results show that the norm of the response in closed paths is small (about 4%–10% of the average response norm). This suggests that the result of Equation 7 is fairly independent of the integration path, so there is no need to specify the integration path.

Table I. Relative norms of the responses in the closed paths.

	(c_2, c_4)	(c_2, c_3)	(c_3, c_4)	(c_3, c_5)	(c_4, c_6)
Low Contr.	0.04	0.05	0.04	0.03	0.10
High Contr.	0.12	0.10	0.09	0.05	0.09

End note

1. The preliminary results of the differential approach used in this paper were presented at the GRC meeting (Malo et al. 2004).

References

Adelson E, Bergen J. 1985. Spatio-temporal energy models for the perception of motion. J Opt Soc Am A 2:284–299.

Barlow H. 1961. Possible principles underlying the transformation of sensory messages. In: Rosenblith W. editor: Sensory Communication, Cambridge, MIT Press, 217–234.

Barlow HB. 2001. Redundancy reduction revisited. Network: Computation in Neural Systems 12:241–253.

Cover T, Tomas J. 1991. Elements of Information Theory. New York: John Wiley & Sons.

Daugman J. 1980. Two-dimensional spectral analysis of cortical receptive field profiles. Vision Research 20:847–856.

Foley J. 1994. Human luminance pattern mechanisms: Masking experiments require a new model. J Opt Soc Am A 11:1710–1719.

Harvey L, Doan V. 1990. Visual masking at different polar angles in the two-dimensional fourier plane. J Opt Soc Am A 7:116–127.

Heeger DJ. 1992. Normalization of cell responses in cat striate cortex. Visual Neurosci 9:181–198.

Hyvärinen A. 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans Neur Nets 10:626–634.

Hyvärinen A, Hurri J, Väyrynen J. 2003. Bubbles: a unifying framework for low level statistical properties of natural images sequences. JOSA A 20(7):1237–1252.

Hyvärinen A, Karhunen J, Oja E. 2001. Independent Component Analysis. New York: John Wiley & Sons.

Jackson JD. (1998). Classical Electrodynamics. NY: John Wiley and Sons.

Jutten C, Karhunen J. 2003. Advances in nonlinear blind source separation. Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003), pp. 245–256.

Kambhatla N, Leen T. 1997. Dimension reduction by local PCA. Neural Computation (9):1493–1500.

Karhunen J, Malaroiu S, Ilmoniemi M. 2000. Local linear independent component analysis based on clustering. Intl J Neur Syst 10(6):439–451.

Kayser C, Körding K, König P. 2003. Learning the nonlinearity of neurons from natural visual stimuli. Neural Comput. 15:1751–1759.

Legge G. 1981. A power law for contrast discrimination. Vision Res., 18:68–91.

Legge G, Foley J. 1980. Contrast masking in human vision. J Optical Soc Am 70:1458–1471.

Lin J. (1999). Factorizing probability density functions: Generalizing ICA. Proc. of the First Intl. Workshop on ICA and Signal Separation, pp. 313–318.

Malo J, Epifanio I, Navarro R, Simoncelli E. 2006. Non-linear image representation for efficient perceptual coding. 15:1. IEEE Transactions on Image Processing.

Malo J, Gutiérrez J, Rovira J. 2004. Perturbation analysis of the changes in V1 receptive fields due to context. GRC Sensory Coding and the Natural Environment.

Olshausen BA, Field DJ. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.

Press W, Flannery B, Teulosky S, Vetterling W. 1992. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge.

Schwartz O, Simoncelli E. 2001. Natural signal statistics and sensory gain control. Nature Neurosci. 4(8):819–825.

Simoncelli E. 2003. Vision and the statistics of the visual environment. Curr Opinion Neurobiology 13:144–149.

Simoncelli E, Olshausen B. 2001. Natural image statistics and neural representation. Annu Rev Neurosci 24:1193–1216.

van Hateren J, van der Schaaf, A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. Proc R Soc Lond B 265:359–366.

Watson A. 1983. Detection and recognition of simple spatial forms. In Braddick O, Sleigh A, editors, Physical and Biological Processing of Images, volume 11 of Springer Series on Information Sciences, pp 100–114, Berlin. Springer Verlag.

Watson A, Solomon J. 1997. A model of visual contrast gain control and pattern masking. J Opt Soc Am A 14:2379–2391.