

BOLD responses in human V1 to local structure in natural scenes: Implications for theories of visual coding

Jochem W. Rieger

Institute of Psychology, Oldenburg University,
Oldenburg, Germany



Karl R. Gegenfurtner

Department of Psychology, Giessen University,
Giessen, Germany



Franziska Schalk

Clinic of Neurology, Magdeburg University,
Magdeburg, Germany



Nick Koechy

Clinic of Neurology, Magdeburg University,
Magdeburg, Germany



Hans-Jochen Heinze

Clinic of Neurology, Magdeburg University,
Magdeburg, Germany



Marcus Grueschow

Laboratory for Social and Neural Systems Research,
Department of Economics, Zurich University,
Zurich, Switzerland



In this study we tested predictions of two important theories of visual coding, contrast energy and sparse coding theory, on the dependence of population activity level and metabolic demands on spatial structure of the visual input. With carefully calibrated displays we find that in humans neither the V1 blood oxygenation level dependent (BOLD) response nor the initial visually evoked fields in magnetoencephalography (MEG) are sensitive to phase perturbations in photographs of natural scenes. As a control, we quantitatively show that the applied phase perturbations decrease sparseness (kurtosis) of our stimuli but preserve their root mean square (RMS) contrast. Importantly, we show that the lack of sensitivity of the V1 population response level to phase perturbations is not due to a lack of sensitivity of our methods because V1 responses were highly sensitive to variations of image RMS contrast. Our results suggest that the transition from a sparse to a distributed neural code in the early visual system induced by reducing image sparseness has negligible consequences for population metabolic cost. This result imposes a novel and important empirical constraint on quantitative models of sparse coding: Population metabolic rate and population activation level is sensitive to second order

statistics (RMS contrast) of the input but not to its spatial phase and fourth order statistics (kurtosis).

Introduction

A prevalent theory of spatial coding in the early visual system supposes that the spatial receptive fields of neurons in the early mammalian visual cortex evolved to efficiently represent information about the natural environment by exploiting redundancies in the spatial structure of the visual world (e.g., Atick & Redlich, 1992; Bell & Sejnowski, 1997; Field, 1987, 1994; Olshausen & Field, 1996). By making this assumption the efficient coding theory puts a strong weight on the biological adaptive value of the neuronal code. In V1, coding efficiency is often measured as the ability of neurons to represent information about the natural environment with a sparse population code in which only few neurons are highly active at a time.

The sparse coding hypothesis received strong theoretical support. Field (1994) found that natural scene

Citation: Rieger, J. W., Gegenfurtner, K. R., Schalk, F., Koechy, N., Heinze, H.-J., & Grueschow, M. (2013). BOLD responses in human V1 to local structure in natural scenes: Implications for theories of visual coding. *Journal of Vision*, 13(2):19, 1–15, <http://www.journalofvision.org/content/13/2/19>, doi:10.1167/13.2.19.

photographs elicit a sparse population response when they are presented to a population of model neurons with receptive fields comparable to those found in a monkey V1. However, the sparseness of the code was greatly reduced and activation spread among larger proportion of the model neuron population when pink noise images, matching natural scenes in the shape of the amplitude spectrum, were presented instead of natural scenes. Furthermore, computational models designed to represent natural scenes with a sparse population response found localized Gabor patch like basis functions resembling localized V1 simple cell receptive fields (Bell & Sejnowski, 1997; Olshausen & Field, 1996). The analysis of the spatial structure of natural scenes suggested that edges defining the shapes of objects are the features of the visual environment that are important for the development of a sparse code with localized, receptive fields (Bell & Sejnowski, 1997; Field, 1993, 1994; Thomson, Foster, & Summers, 2000). Edges in natural scenes can be described as sparsely distributed local luminance variations that are spatially correlated over multiple spatial scales.

However, it is less clear what the adaptive value of a sparse code in V1 could be. Two approaches are currently investigated. One focuses on the efficiency in information representation and the other on metabolic efficiency. Sparse codes are considered *information efficient* as compared to dense codes. In the latter many neurons are simultaneously active. Information efficiency increases because neuronal responses are less correlated in a population employing a sparse code (Bell & Sejnowski, 1997; Földiák, 2002; Olshausen & Field, 1997; but see also Bethge, 2006). Supporting the notion of information efficient coding of natural scenes, several studies report that V1 neurons exhibit sparse responses when presented with complex natural scene stimuli (e.g., Baddeley et al., 1997; Felsen, Touryan, Han, & Dan, 2005; Vinje & Gallant 2000, 2002; Weliky, Fiser, Hunt, & Wagner, 2003; Willmore & Tolhurst, 2001). However, most of these studies investigated single neuron responses which makes it hard to draw conclusions about the behavior of the population of neurons in a brain area (Olshausen & Field, 2004). In addition to increasing information efficiency, a sparse population code might help to reduce the metabolic burden of visual processing. Metabolic costs of neuronal firing are high (Attwell & Laughlin, 2001; Lennie, 2003) and theoretical considerations suggest that a sparse code, in which only relatively few neurons in a population fire simultaneously at high rates while all others are relatively inactive, might be *metabolically efficient* (Baddeley et al., 1997; Graham & Field, 2009; Hyvärinen, Hurri, & Hoyer, 2009; Laughlin & Sejnowski, 2003; Lennie, 2003; Levy & Baxter, 1996; Olshausen and Field, 2004; Rozell, Johnson, Baraniuk, & Olshausen, 2008; Vinje and Gallant, 2002). However,

no experimental data exists so far investigating the metabolic efficiency of sparse coding in area V1 at the neuronal population level.

The extent to which the neuronal code in the early visual system can sparsely represent a visual input depends on the input's specific features. In natural scenes adjacent locations provide highly correlated information (Field, 1994; Kersten 1987) except for edges where color and luminance change rapidly over space. Several theoretical studies (Bell & Sejnowski, 1997; Field, 1993, 1994; Thomson et al., 2000) identified such local structure in natural scenes with local Fourier phase alignments over spatial scales and argued that these phase alignments are the basis for the sparseness of the responses of spatial filters resembling V1 receptive fields. Importantly, Field (1994) noted that such spatial filters cannot sparsely represent natural scene photographs with local phase alignments destroyed by increasing phase noise (i.e., pink noise images). Hence, sparsely distributed edges seem to be important features that allow V1 neurons to represent information from natural scenes with a sparse code (Bell & Sejnowski, 1997; Field, 1994). Accordingly, the neuronal code should become more distributed and thus less information efficient and energy efficient when edges are removed.

However, the prediction that energy efficiency depends on sparseness of the visual input stands in contrast to another well-accepted classical theory of visual coding in V1, the contrast energy theory (Albrecht & Hamilton, 1982; De Valois, Albrecht, & Thorell, 1982; De Valois & De Valois, 1990; Movshon, Thompson, & Tolhurst, 1978). Contrast energy theory supposes that the activation level in the neuronal population of V1 is essentially determined by local contrast at different spatial resolutions rather than by local phase alignments that produce edges. In other words, contrast energy theory predicts that the population activation level in V1 does not vary when edges are removed from the visual input, given that the global image contrast is retained. This prediction stands in sharp contrast with the presumption that metabolic efficiency of sparse neural coding critically depends on the occurrence of edges. Contrast energy theory of V1 coding (Albrecht & Hamilton, 1982; De Valois et al., 1982; De Valois & De Valois, 1990; Movshon et al., 1978) would be fundamentally flawed if the energy efficiency assumption of sparse coding theory put forward by several authors (Baddeley et al., 1997; Graham & Field, 2009; Hyvärinen et al., 2009; Laughlin & Sejnowski, 2003; Lennie, 2003; Levy & Baxter, 1996; Olshausen & Field, 2004; Rozell et al., 2008; Vinje & Gallant, 2002) proves to be correct for natural scene stimuli. The few existing experimental data addressing this contradiction between the two most successful theories of early visual coding are not

conclusive (Dakin, Hess, Ledgeway, & Achtman, 2002; Rainer, Augath, Trinath, & Logothetis, 2001).

In the approach taken in this study we use functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) to quantify the strength of the population responses in early visual areas to natural scene stimuli with constant contrast but parametrically varied amounts of local structure such as edges. Our results indicate that sparse coding theory and classical spatial energy coding theories are compatible. However, energy savings in V1 due to concentrating contrast energy in edges appear to be minimal or at least below our measurement accuracy. Our findings impose a strong constraint on sparse coding theories in that the stimulus dependent neural code must change from sparse to distributed in a population energy conserving way when the visual input becomes less sparse.

Methods

Stimuli, phase randomization, and image statistics

Color photos of natural scenes were chosen from a commercial database (Corel Stock Photo Library, 356 different images) and vignetted with a cumulative Weibull function (10 pixels wide) to ensure a smooth transition into the screen background. Fourier phase randomization was employed to gradually remove edges in the images. Edges critically depend on local phase alignments of Fourier components over spatial scales. Note that phase randomization retains the images' root mean square (RMS) contrast. Parseval's theorem (Equation 1) states that the rooted mean square (RMS) image contrast is equal to the sum over the image's 2-D power spectrum:

$$\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} p(x, y)^2 = \frac{1}{NM} \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} F(u, v)^2. \quad (1)$$

Here, $p(x, y)$ is the value of pixel p at the image location x, y . The Fourier component $F(u, v)^2$ is the real valued power (squared amplitude) of the complex valued spatial frequency component $F(u, v)$. Power is by definition phase independent. Hence, phase randomization retains image statistics up to the order of two including both RMS contrast and the shape of the amplitude spectrum $F(u, v)$.

The complex Fourier transform of real valued images is symmetric up to complex conjugation of phase angles such that $F(-u, -v) = F(u, v)^*$, where $*$ denotes complex conjugation. Phase randomization must retain this symmetry because only the symmetric Fourier matrices produce real valued images when the

inverse Fourier transform is applied. To retain this symmetry we performed phase randomization on one half of the complex valued Fourier components (e.g., the quadrants u, v and $-u, v$) and used their complex conjugates to fill in the other two quadrants in the Fourier matrix (e.g., $-u, -v$ and $u, -v$, respectively). For phase randomization a random angle α chosen from an interval $[-\theta, \theta]$ with flat density was added to the phase angles $\varphi(u, v)$ of the 2-D Fourier matrix. A new random angle was drawn for each phase angle. The amount of phase randomization was varied in five steps by choosing narrower or wider interval limits for θ : $[-30^\circ, -30^\circ]$, $[-60^\circ, 60^\circ]$, $[-90^\circ, 90^\circ]$, $[-135^\circ, -135^\circ]$, and $[-180^\circ, -180^\circ]$. Resulting angles exceeding $\pm 180^\circ$ were circularly included at the opposing end of the interval by taking the modulo over the interval 2π :

$$\varphi_{rand}(u, v) = \text{mod}(\pi + \varphi(u, v) + \alpha, 2\pi). \quad (2)$$

This circular inclusion retains the flat distribution phase angles and avoids the over-representation of certain phase angles. The latter can lead to artificial local concentrations of image contrast in the noise images (Dakin et al., 2002; Rainer et al., 2001; Tjan, Lestou, & Kourtzi, 2006).

Inverse Fourier transformation of the phase randomized images can produce pixel values exceeding the dynamic range available for the real valued photographs. We clipped out-of-range pixel values occurring after inverse Fourier transformation to the attainable range. In order to prevent excessive clipping, which may alter image contrast, we reduced the contrast of all natural scene photographs by 10% before Fourier transformation. This reduced the proportion of clipped pixels to less than 0.5%. Importantly, the flat average RMS contrast function in Figure 1B shows that our approach was successful because clipping due to phase randomization did not alter image contrast. After the inverse Fourier transformation the phase randomized images were vignetted with a cumulative Weibull function (10 pixels wide) to guarantee a smooth transition into the background. Examples of the original and the phase randomized images are shown in Figure 1A. The images' RMS contrasts within the region surrounded by the vignette are shown in Figure 1B.

We calculated the excess kurtosis (Thomson et al., 2000, Equation 3) of the image pixel values, a fourth order statistic, to assess the effects of phase randomization on image sparseness:

$$k = \frac{\frac{1}{NM-1} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (p(x, y) - \bar{p})^4}{\left(\frac{1}{NM-1} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (p(x, y) - \bar{p})^2 \right)^2} - 3. \quad (3)$$

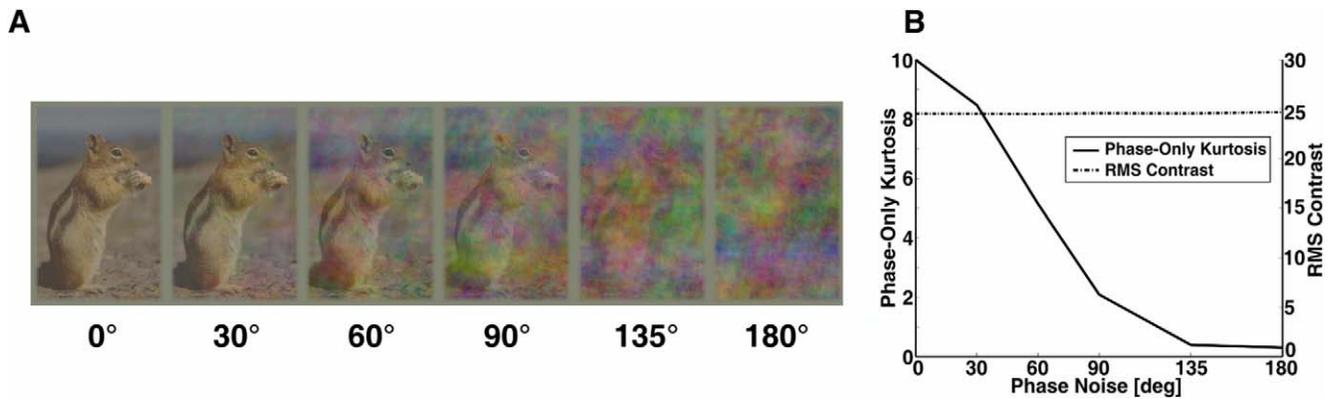


Figure 1. (A) Example of a photograph used in the experiment. The phase noise level increases parametrically from left to right. All images have identical amplitude spectra and RMS contrast. (B) The mean of the RMS contrast and the phase-only kurtosis over all images used in the first experiment. The RMS-contrast variations over the phase noise levels are less than 0.6%. Pixel kurtosis decreases with increasing phase noise. RMS contrast and kurtosis values were calculated from RGB-values (8-bit resolution) used to represent the images. Only the pixels within the vignette were used to calculate the image statistics.

Excess kurtosis of a Gaussian distribution is zero, and values above zero indicate a peaked distribution. Excess kurtosis is a pixel based measure to assess phase alignment in the natural scenes. Similar to the kurtosis of the coefficients of the log Gabor transformation of the images (Field, 1994) excess kurtosis falls off to values close to zero for fully phase randomized natural scene pictures (Figure 1B). Note that phase randomization may change the third order statistics (skew) of the image as well as kurtosis (Field, 1994; Thomson et al., 2000). However, a potential role for skew in sparse coding is less clear. Therefore, we focus on RMS contrast and kurtosis as statistical descriptors.

Subjects

Four subjects (three male) participated in the first fMRI experiment in which only the Fourier phase was manipulated. The same number of subjects (one new male) participated in the second fMRI experiment where both phase and contrast of the images were manipulated. Borders of V1 were determined with retinotopic mapping. Ten new subjects participated in the MEG experiment (five male). The data from one subject were excluded from the analysis due to excessive artifacts in the MEG recordings. All subjects had normal or corrected-to-normal visual acuity and gave written informed consent prior to the experiment. The experimental protocol was approved by the ethics committee of the Medical University of Magdeburg.

Stimulus presentation

In the MRI scanner stimuli were back projected (Sharp SX21) onto a screen 27 cm from the eyes in

the magnet bore and viewed via a mirror. The screen's background was set to the average luminance of the photographs (892 cd/m^2). Natural scene photographs and phase randomized versions thereof were presented alternately (every 3 s) to the left and right of a central fixation cross, centered on the horizontal midline. The medial edges of the images were 2° from fixation. The photos subtended 26.1° visual angle vertically and 17.4° visual angle horizontally. In the MEG experiment, the stimuli were back-projected (JVC DLA-G150CL) onto a screen placed 120 cm from the eyes. Our aim was to keep the visual stimuli as comparable as possible between the experiments. Therefore, we matched the retinal size of the images with the size in the MRI scanner and set the background luminance to the mean luminance over all natural scene photographs (377 cd/m^2). However, in the MEG experiment photos were presented in the lower visual field to increase the size of the C1 component that is thought to reflect neuronal activity from striate cortex (Martínez, Di Russo, Anllo-Vento, & Hillyard, 2001; Noesselt et al., 2002). To ensure fixation and to bind spatial attention subjects performed a demanding discrimination task at the fixation cross throughout all experimental runs. The length of the bars of the fixation cross slightly changed approximately every second and independent of the image onsets. The subjects had to detect the change and to report whether the shape of the cross was tall or wide. The mean correct performance in this task was 89.5% throughout the experiments and did not differ significantly across experimental conditions.

The RGB-to-luminance functions of both display systems were measured with a PR650 spectroradiometer (SpectraScan). A linearization lookup table was calculated and used to linearize the intact and

transformed images. This is necessary since RMS contrast in phase-randomized images is only maintained when the RGB-to-luminance function of the display system is linear.

FMRI scanning and data analysis

MRI scanning was performed in a 1.5T GE-Signa LX neuro-optimized system (General Electrics, Milwaukee, WI) with a 5" receive-only surface coil placed over the occipital pole. A BOLD sensitive gradient recalled echo planar imaging (EPI) sequence (TR = 2, TE = 40 ms, flip angle = 80°) was used to collect 23 slices (2.81 × 2.81 × 3 mm in-plane resolution, no gap, matrix 64 × 64 pixels) approximately perpendicular to the calcarine sulcus. The six phase randomization levels were presented in a pseudorandomized sequence of 36 s blocks separated by 10 s blocks in which only the fixation task displayed. Twelve blocks were presented per run (two blocks of every phase randomization level per run). Every run started and ended with the fixation condition. Every subject was scanned in seven runs, each 594 seconds long. Eye movements were monitored online with a custom-made video-based eye-tracking system (Kanowski, Rieger, Noesselt, Tempelmann, & Hinrichs, 2007). No significant eye movements were observed. For the anatomical localization of the functional images, T1 weighted anatomical images (2-D spin echo sequence) were scanned with the same slice position and thickness as the EPI images (in-plane resolution 0.76 × 0.76 mm, 256 × 256 pixels).

We used the standard approach of the statistical parametric mapping (SPM) software (Wellcome Department of Cognitive Neurology, UK) to estimate individual BOLD response amplitudes from the EPI time series. The EPIs were first 3-D motion corrected and then spatially smoothed applying a 6 mm full width at half maximum (FWHM) Gaussian kernel. We estimated the magnitude of the BOLD responses employing the regression model implemented in SPM. In short, we produced model BOLD responses for each of the seven experimental conditions (intact scenes, five levels of phase randomization, and rest) by convolving boxcar functions representing the time course of the condition with the canonical hemodynamic response. The model functions were regressed onto the EPI voxel time series using a multiple regression model. The regression weights are estimates of the BOLD responses evoked in the different experimental conditions. To make the estimates of BOLD responses more comparable between subjects we set the individual BOLD responses elicited by intact scenes to one and expressed the BOLD responses to the phase randomized versions

of the images in proportions of the BOLD response to the intact scenes.

The voxel in V1 with the most significant modulation in the comparison “intact scene > fixation only” was selected as the region of interest (ROI) for subsequent analyses. Note that due to the spatial smoothing the time series of this voxel is a weighted average of 27 neighboring voxels.

Retinotopic mapping was performed in a separate session to establish the individual borders of V1 and followed the standard protocol (Engel et al., 1994). A custom modified version of the mrVista toolbox (<http://white.stanford.edu/software>) was used to determine the borders of V1 and to confirm that the anatomically selected ROIs were within the functionally defined area borders.

MEG recordings and data analysis

The MEG was recorded with a BTI Magnes 2500 WH 148 channel whole head magnetometer system at 678.17 Hz sampling rate. Data were low-pass filtered at 40 Hz and trials containing artifacts (peak-to-peak values exceeding 3 pT within 500 ms) or drifts were discarded from further analysis. The same blocked presentation scheme used in the fMRI scans was used in the MEG experiments except that only three experimental conditions were presented (intact scenes, 90°, and 180° phase randomization) to increase the signal-to-noise ratio in each condition. We used the same attention binding task at the fixation as in the fMRI study to minimize top down attentional influence. A photodiode was attached to the presentation screen to detect the onset of the image. This signal was recorded and served as the marker for the start of the trials. On average, 322.2 trials per subject and condition were analyzed.

We performed two complementary analyses. Sensor data were analyzed using an ANOVA with the factor phase noise calculated for each sampling point and thresholded according to a standard criterion ($p < 0.01$) and at least ten consecutive samples in the time series above threshold (Johnson and Olshausen, 2003). In addition, current source density (CSD) maps were calculated with the Curry software package (Neuroscan, Charlotte, NC) on the surface of the individual brains. We extracted the CSD time courses from ROIs (radius 5 mm) placed on the individual activation maximum in the calcarine sulcus elicited by intact scenes and extracted the activation time courses between 50 ms to 100 ms after the onset of the scene. Effects on the stimulus driven activation in V1 would be expected between 50 ms and 80 ms (Martínez et al., 2001; Noesselt et al., 2002). The CSD time courses provide a better separation of the

activity from different sources outside V1 than the sensor data.

Results

The effect of phase randomization on natural scene photographs is shown in Figure 1A. Perceptually, edge sharpness declines as phase noise increases because edges require phase alignment of spatial image components over spatial scales. Consequently, only a cloudy image is left at the highest phase randomization level (180°). The effect of phase randomization on the pixel statistics of the pictures is shown in Figure 1B. As phase noise increases, excess kurtosis of the pixel values histogram decreases. Lower kurtosis indicates a decrease of pixel value redundancy in the pictures with increasing phase noise (Bell & Sejnowski, 1997; Field, 1994; van Hateren & van der Schaaf, 1998). Perceptually this is evident as reduction of uniform areas and a transition to cloudy pictures (see Figure 1A). Figure 1B shows that at the highest phase noise level excess kurtosis is close to zero, indicating that the distribution of pixel values is now close to Gaussian. In addition to pixel based excess kurtosis (Thomson et al., 2000) we calculated the kurtosis of the responses in a Gabor filter bank (Field, 1994) to quantify the sparseness of the population response of filters at different phase noise levels. The filter bank consisted of Gabor filters with 4 - 0.5 cycles/° visual angle center frequency (three steps), 1.5 octave bandwidth, and eight orientations. The kurtosis of the filter bank responses to the different phase randomization levels was slightly higher than pixel based kurtosis (phase randomization/kurtosis: 0/12.3, 30/10.6, 60/7.1, 90/3.9, 135/1.7, and 180/1.6) but followed the same trend as predicted by Field (1994): increasing phase noise decreases the pixel sparseness (kurtosis) of the images and the sparseness (kurtosis) of the population response in the Gabor filter bank. Importantly, RMS contrast remained constant over phase noise levels (Figure 1B). The reason is that second order statistics, the RMS contrast, is determined by the Fourier amplitude spectrum and thus unaffected by noise added to the Fourier phases. Conversely, kurtosis is a fourth order statistic and therefore sensitive to Fourier phase noise. According to the sparse coding hypothesis decreasing redundancy in the input images should transform the population distribution of activation in early visual areas from a sparse code with few vigorously activated neurons to a dense code with more distributed population activation (Bell & Sejnowski, 1997; Dakin et al., 2002; Field, 1994). Thus, if sparse codes are more energy efficient than dense codes, then the population activation level in V1 should increase with increasing phase random-

ization. However, the population activation should remain at a constant level if it is determined by RMS contrast and independent of the redundancy of the pixel values in the natural scene photographs. We tested these conflicting predictions in the first experiment.

We determined individual regions of interest (ROI) in each hemisphere by computing the BOLD contrast intact image versus fixation and selecting the region around the peak effect in V1 (Figure 2A). Figure 2B shows the BOLD responses obtained in the V1 ROI at different phase noise levels. We found no influence of phase coherence or image pixel sparseness on the population response level in V1. Accordingly, the parametric response function is flat (regression slope = 0.00003/° with a 95% confidence interval of -0.0005/° to 0.0006/°) and an analysis of variance (ANOVA) fails to show a significant effect of phase noise, $F(4, 12) = 2.4$; $p > 0.1$. This result was independent of the particular definition of the ROIs. We obtained the same nonsignificant regression when we included all voxels in calcarine sulcus that showed a response to pictures of any phase randomization level, regression slope = 0.00068/°; $p > 0.1$; 95% confidence interval of -0.0003/° to 0.0017/°; $F(4, 12) = 0.8$, $p > 0.1$; mean number of voxels in ROI: 61.6 with an *SE* of ± 32.1 . Figure 2C shows the results of a control experiment in which we reduced the RMS contrast of the pictures in addition to the phase manipulation. The V1 BOLD response reliably covaried with the image RMS contrast, Figure 2C, ANOVA, $F(4, 12) = 8.9$, $p < 0.005$; regression slope = 0.11/RMS with a 95% confidence interval of 0.04/RMS to 0.19/RMS. These results indicate that the population activation level in V1 assessed by the BOLD response is independent of several spatial features of the stimuli: phase coherence, presence or absence of edges, and the recognizability of the image content. Importantly, the V1 BOLD response level was also independent of the higher order pixel statistics such as kurtosis, a measure of redundancy in the pixel statistics of the stimuli. Thus, although sparseness of the neuronal population code in V1 varies with redundancy in the stimulus, as suggested by the sparse coding hypothesis (Bell & Sejnowski, 1997; Field, 1994) and confirmed by multiple single cell studies (Baddeley et al., 1997; Felsen et al., 2005; Vinje & Gallant, 2000, 2002; Weliky et al., 2003; Willmore & Tolhurst, 2001), our data indicate that the population activation level does not depend on the level of input redundancy (Field, 1994). Importantly, the constant V1 activation levels for stimuli with equal RMS contrast are predicted by contrast energy theory. According to this theory the population activation level is largely independent of the particular spatial structure of the stimulus but

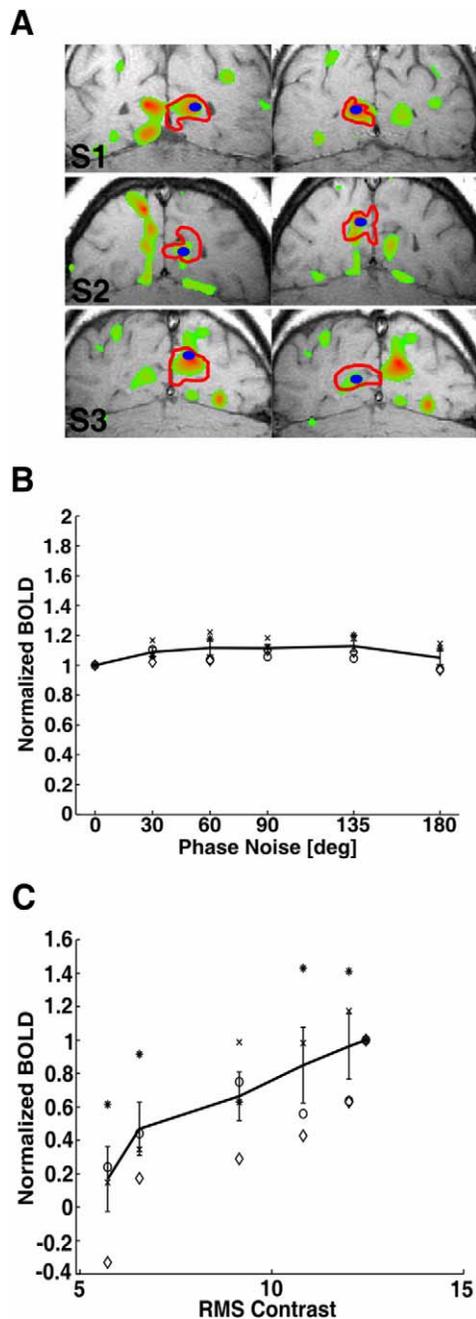


Figure 2. (A) The location of the ROIs in the left and right hemispheres of three subjects. The blue dots indicate the location of the ROI. The red lines depict the borders of V1. Red in the color map indicates higher and green indicates lower t -values in the comparison intact scene > fixation only. (B) The normalized BOLD responses elicited in human V1 by natural scene photographs with different phase noise levels. The flat response profile indicates that neither the noise in the phase spectrum nor the phase-only kurtosis had an influence on the amplitude of the V1 BOLD response. Normalization was done by dividing individual BOLD responses by the individual BOLD response obtained with intact scenes. The different symbols depict the responses for single subjects. (C) The BOLD response obtained in human V1 when the contrast was varied in addition to the phase noise. The BOLD response increases as the contrast increases. The responses were normalized to the value obtained with intact scenes which had the maximum contrast.

changes when RMS contrast, a second order stimulus statistics, changes.

In a third experiment, we aimed to exclude the possibility that the observed insensitivity of the V1 population response level to higher order spatial stimulus characteristics in the slow BOLD response is mediated by late top-down processes that compensate bottom-up response differences. MEG provides the necessary temporal resolution in the order of milliseconds. Therefore, we ran the experiment in the MEG with three phase randomization levels (0° , 90° , and 180°). Figure 3A shows the time series of significant differences between MEG activations of at least two phase randomization levels. The sensors included in this plot broadly covered early visual and temporal brain areas but also receive input from higher order visual areas in the temporal and parietal cortex. The sensor locations are indicated as small dots in the upper left inset. No activation differences are evident during the initial 80 ms after scene onset. The bottom-up activation sweep in V1 is known to occur between 50–80 ms at the scalp (Martínez et al., 2001; Noesselt et al., 2002). The lack of an effect during this interval indicates that phase randomization does not have an effect on the amplitude of the initial bottom up population response in early visual areas. The earliest significant effects of phase randomization observable in some sensors started after 86 ms and peaked circa 110 ms after stimulus onset ($p < 0.01$ in at least 10 consecutive samples). The later effects indicate the beginning of neural processing of meaningful image contents in extrastriate brain areas. A previous study showed that they predict the subject's ability to discriminate among different (intact) natural scene photographs in single trials (Rieger, Reichert et al., 2008). In short, the sensor space analysis suggests that the initial V1 population activation level is independent of phase randomization level whereas late phase randomization dependent MEG activations differences are generated in extrastriate cortex.

Sensor space analysis is valuable because it provides an overview of the sequence of effects in different brain areas. However, a problem with it is that MEG sensors receive signals from different brain areas which can be located several centimeters apart. Therefore, we aimed to increase the anatomical specificity of our time resolved analysis. We used current source density (CSD) mapping to model the spatial distribution of neural population activation on the individual gray matter sheets of our subjects' brains in order to recover the initial V1 activation elicited by image presentations from the MEG sensor data. CSD mapping provides a spatially and temporally resolved estimate of the local density of electric current flow accompanying neuronal activity in cortex. The inset in Figure 3B shows the CSD distribution at 70 ms after stimulus onset on the

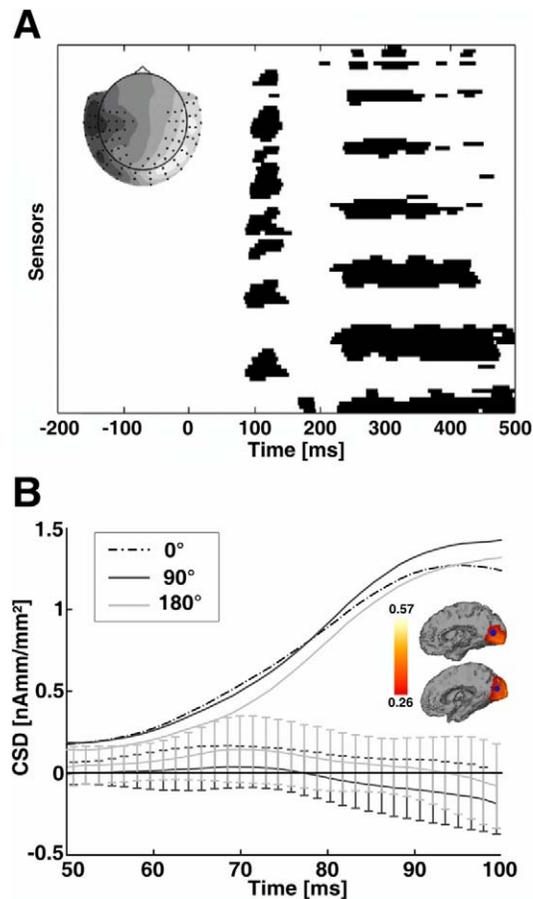


Figure 3. (A) The MEG sensor by time matrix of the time series of significant effects of phase randomization (ANOVA with three levels of phase randomization) provides an overview of the timing of statistically significant effects of phase randomization. The earliest effects begin 86 ms post stimulus and reach a maximum after circa 110 ms. The initial stimulus driven response in V1 would be expected earlier between 50–80 ms (see text). Each row shows the results of a time series of ANOVAs performed in one MEG-sensor. Time points with $p < 0.01$ in more than 10 consecutive samples are marked in black. The time series of individual MEG-sensors are stacked vertically. The inset shows the average evoked magnetic fields distribution over the head 80 ms after scene onset. Black dots show that the sensors included in the matrix cover the early activation. (B) Curves without error bars show the current source density (CSD) time courses extracted from a ROI in the calcarine sulcus at the location showing maximum activation around 80 ms after stimulus onset. The time courses are very similar for all three phase randomization levels tested. The difference curves with error bars (0° – 90° and 0° – 180° phase randomization) are close to zero. The $p < 0.01$ confidence interval for a paired t test is indicated by the errors bars. No significant differences occur in the critical time interval for initial stimulus driven V1 activation. The inset shows the CSD distribution on the individual right hemispheres of two subjects 70 ms after stimulus onset. The blue dots indicate the location of the ROIs.

medial views of the right brain hemispheres of two subjects. At this latency, the activation is mostly restricted to occipital cortex. The three activation time courses are taken from ROIs that were placed on the individual activation maximum in calcarine sulcus. The black and gray curves with error bars show the time courses of the difference between activations elicited by natural scene photographs and 90° and 180° phase randomized scenes, respectively. The error bars indicate the two sided $p < 0.01$ confidence interval for the difference at each time point. These confidence intervals contain zero over the whole time interval in indicating phase randomization has no effect on the magnitude of the population activation in V1 during the initial bottom-up response between 50 and 100 ms after image presentation onset.

In concordance with the fMRI results, the MEG results indicate that the population activation level in V1 is insensitive to phase randomization, the occurrence of edges, and the redundancy in the stimulus as assessed by higher-order image statistics. The independence of the V1 activation level of stimulus redundancy assessed by both fMRI and MEG strongly argues against the assumption that a sparse population code in V1 would be particularly energy efficient in a natural environment with high local redundancies. Our results rather suggest that the population activation level in V1 is determined by RMS contrast, as predicted by contrast energy theory.

Discussion

The two best accepted theories of visual coding in primate V1 state that (a) neurons in the early visual system code image contrast and that (b) their localized receptive fields are adapted to efficiently represent spatially sparse informative features in the natural environment in a sparse neural code. A necessary precondition for neural sparse coding is edges in the luminance profile that manifest as local phase alignment over a range of spatial scales. A potential contradiction between the two theories arises with the often stated idea that a sparse neural code is metabolically more efficient than a dense code. We tested this hypothesis by presenting carefully manipulated photographs of natural scenes with spatial structures important for neural sparse coding gradually removed but unaltered RMS contrast. We measured V1 neural activation elicited by such stimuli with two complementary imaging methods, functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG). With precisely calibrated displays we find that neither the V1 BOLD response nor the initial visually evoked fields in MEG are sensitive to

parametric manipulations of image sparseness. As controls, we quantitatively show that the applied phase perturbations decrease sparseness (kurtosis) of our images but preserve their RMS contrast and that the lack of sensitivity of the V1 population responses level to phase perturbations is not due to a lack of sensitivity of our brain imaging methods. BOLD responses in V1 proved to be highly sensitive to variations of image RMS contrast. Our results indicate that the transition from a sparse to a dense neural code in the early visual system induced by reducing image pixel sparseness has negligible consequences for population activation level and metabolic cost in V1. This result imposes a novel and important empirical constraint on quantitative models of sparse coding: Population metabolic rate and population activation level in human V1 is sensitive to second order statistics (RMS contrast) of the input but not to its spatial phase and fourth order statistics (pixel kurtosis).

Input sparseness, sparse coding in the early visual system, and metabolic efficiency

The recent decades of research on the visual system were strongly influenced by the question of potential goals of sensory coding driving the development of localized receptive fields empirically found in primate V1. Different models were developed to represent information in input images with a sparse code in which few neurons are simultaneously active. In one approach sparse coding is achieved by adapting the neurons' receptive fields, the model's basis functions, to exploit regularities in the spatial structure of natural scenes (Bell & Sejnowski, 1997; Field, 1994; Olshausen & Field, 1996). In another approach the models start from an overcomplete set of basis functions and enforce sparseness of the image representation, for example by means of lateral inhibitory interaction (Rozell et al., 2008; Schwartz & Simoncelli, 2001). We will relate our results to both approaches.

Important in the first approach was the idea that the spatial structure of the receptive fields in the early visual system is well suited to optimize information representation in a sparse neural code that exploits spatial regularities in the natural environment and reduces energy consumption (Lennie, 2003). Starting from signal processing considerations several groups (Atick & Redlich, 1992; Field, 1987) found that receptive fields of simple and complex cells in primate V1 are well adapted to the approximate $1/f$ fall-off in the amplitude spectrum of natural scenes. However, later investigations showed that the $1/f$ amplitude fall-off in natural scenes cannot explain the development of localized receptive fields in the visual system (Bell & Sejnowski, 1997; Field, 1994; Olshausen & Field, 1996).

Field (1987, 1994) showed that nonlocalized receptive fields would be better suited to represent spatial information in a world with $1/f$ amplitude fall-off than localized receptive fields. This insight highlighted the importance of local phase alignments over multiple spatial scales of natural scenes for sparse coding. These phase alignments perceptually manifest as lines and edges in natural scenes and are thought to drive the development of localized Gabor-like receptive fields in models when a sparse coding constraint is imposed (Bell & Sejnowski, 1997; Field, 1994; Olshausen & Field, 1996). Field (1994) argued that these nonrandom phase alignments in natural scenes produce higher spatial redundancy in natural scenes as compared to noise images with amplitude spectra that have the same shape as the intact images. Therefore, our goal was to keep amplitude spectra perfectly matched between phase randomization levels. In our approach, we randomized Fourier phases in photographs of natural scenes independently over spatial scales and orientations. In addition, we used different degrees of randomization to parametrically transform the highly redundant spatial statistics of our natural scene photographs to the low redundancy level of pink noise images. The success of this manipulation is indicated by the decrease of pixel kurtosis with increasing degrees of phase randomization while RMS contrast remains at the same level. Field (1994) showed theoretically and multiple single neuron recordings (Baddeley et al., 1997; Felsen et al., 2005; Vinje & Gallant, 2000, 2002; Weliky et al., 2003; Willmore & Tolhurst, 2001) showed experimentally that localized log Gabor-like receptive fields, similar to those of V1 neurons, translate high pixel redundancy in natural scenes in high redundancy in neuronal population activity. To provide further support for the biological plausibility of this theory, Olshausen and Field (1996) derived a dictionary of basis function which closely resembled the receptive fields of neurons in V1 to approximate natural scene photographs with a sparse code. To obtain these basis functions the authors traded off the veridicality of natural scene representation with the sparseness of the neural code. In sum these studies suggest that V1 neuronal population activation is expected to be sparse with many weakly activated and few strongly activated neurons when intact natural scene photographs are presented as visual stimuli. With increasing phase randomization in our images, however, the neuronal population code in V1 would become less sparse. This behavior is expected for sparse coding models based on log Gabor-like basis functions that are optimal to veridically represent natural scenes with a sparse population code. The implementations of sparse coding discussed so far do not impose constraints on the veridicality of the image representation once the basis functions are determined. The veridical sparse repre-

sensation of natural scenes is considered energy efficient with such basis functions (Baddeley et al., 1997; Graham & Field, 2009; Hyvärinen et al., 2009; Perna, Tosetti, Montanaro, & Morrone, 2008; Rozell et al., 2008; Vinje & Gallant, 2000, 2002; Willmore & Tolhurst, 2001) but because they are optimized to represent structure in natural scenes they would produce a less energy efficient population code when the goal is to veridically represent images lacking the required spatial structures. Of course it is not possible to directly measure neural population sparseness and veridicality of the image representation with the noninvasive recording methods we used in our study. However, based on the discussed theoretical considerations and experimental evidence we consider it safe to suppose that the manipulation of the pixel sparseness by increasing phase randomization of our stimuli transformed the neuronal population code in V1 from sparse to distributed as predicted by sparse coding theory. Importantly, this manipulation had no effect on the overall population activation level measured with the complementary imaging methods fMRI and MEG. This result implies that redundancy reduction in V1's neural population response does not reduce the metabolic demand of neural information coding. These are constraints that sparse coding models should meet to be biologically plausible. The independence of the population response level of Fourier phase and the dependence of the response level on Fourier amplitude we found support the hypothesis that at the population level, neural responses in V1 can be described in a good approximation as a linear shift invariant system (e.g., De Valois & De Valois, 1990; Movshon et al., 1978). This conclusion is in concordance with contrast energy theory and theoretical work aiming to derive linear basis functions (receptive fields) for sparse coding of spatial information directly from natural scenes (Bell & Sejnowski, 1997; Field, 1987; Hancock, Baddeley, & Smith, 1992; Olshausen & Field, 1997; van Hateren & van der Schaaf, 1998).

The second implementation of sparse coding considered here does not necessarily assume that the goal of sparse coding is to veridically represent images independent of spatial structure. For example, contrast normalization through inhibitory interactions in a pool of neurons receiving input from the same retinotopic location might help to make neural responses sparser (Schwartz & Simoncelli, 2001). Moreover, a recent biologically inspired signal processing approach to sparse approximation of images used a fixed dictionary of basis functions and traded off the veridicality of the image representation against the sparseness of the population activation (Rozell et al., 2008). In principle, such approaches could enforce a fixed level of sparseness of the image representation independent of the level of phase randomization in

the input images but at the cost of the veridicality of the representation. Whether a bound on sparseness that retains population activity strength for images with different levels of phase randomization can be implemented in a biologically plausible way remains to be shown.

Relationship among neural coding schemes

The literature on sparse coding distinguishes two different types of sparseness (e.g., Willmore & Tolhurst, 2001). A neuron with high *lifetime sparseness* would rarely respond to few natural scene images but vigorously if it responds. *Population sparseness* describes the situation that an image elicits vigorous responses in a few neurons and no response in most other neurons of the population. Most previous single cell studies focused on lifetime sparseness and mean firing rates because population sparseness is difficult or even impossible to measure in recordings using only one electrode (Willmore & Tolhurst, 2001). We designed our stimuli to manipulate population sparseness by carefully varying image pixel sparseness. Covariation of neural population activation with varying stimulus contrast and independence of population activation levels of image pixel sparseness indicates that V1 population activation levels represent stimulus contrast rather than sparseness of the population code. This is in concordance with the contrast energy model of visual coding that states that V1 neurons represent local visual stimulus contrast (De Valois et al., 1982; Movshon et al., 1978; Wandell, 1995) up to a nonlinearity in the output function which becomes evident at high contrasts (Albrecht & Hamilton, 1982) and a potential non-phase sensitive contrast normalization (Heeger, 1992; Tolhurst, 1972). However, sparse population coding and contrast energy coding are not necessarily mutually exclusive. This is already suggested by the fact that the early theory of sparse coding in V1 cortex was developed on the basis of linear, spatially localized log Gabor-like filter banks (Field, 1987, 1994; Olshausen & Field, 1996).

What is a sparse neural code good for?

Graham and Field (2009) discuss three potential goals of sparse codes: increase of the code's information capacity, increase of the neural population's memory capacity, and decrease of the metabolic burden of visual processing. A potential decrease of the metabolic burden of visual processing in a sparse code is one of seemingly plausible goals of sparse codes put forward by several authors (e.g., Baddeley et al., 1997;

Graham & Field, 2009; Hyvärinen et al., 2009; Perna et al., 2008; Rozell et al., 2008; Vinje & Gallant, 2000, 2002; Willmore & Tolhurst, 2001). The high metabolic cost of spiking (Attwell & Laughlin, 2001; Lennie, 2003) is thought to limit the brain's information processing capacity. A sparse code would reduce the average spike rate and thus reduce metabolic demands of information coding. However, our data suggest otherwise. We find no indication that the neural code in the early visual system is optimized for representing ecologically valid natural scene stimuli with low metabolic demands. In concordance with Olman, Ugurbil, Schrater, and Kersten (2004) and Tjan et al. (2006) we find that the metabolic demand of the population code in the early visual system is invariant to the phase structure of the visual input but highly dependent on the RMS contrast. In addition, the importance and the amount of information capacity improvement by a sparse code has been called into question (Bethge, 2006; Graham & Field, 2009).

To date no strong, empirically supported argument seems to exist for the notion that localized receptive fields in the early visual system developed as an adaptation to spatial statistics of natural scenes. It is thus conceivable that internal, instead of external, factors contribute to the development of receptive fields in their observed form. One such biological constraint might be the reduction of average wiring length in neural tissue (Chklovskii & Koulakov, 2004; Laughlin & Sejnowski, 2003). Axons have slow and variable conduction velocities. Thus shorter axons might increase computation speed and spike timing reliability in the visual system (VanRullen & Thorpe, 2002). In addition, 40%–50% of the human brain consists of white matter (Miller, Alston, & Corsellis, 1980), space in the skull unavailable for the gray matter that performs computations. Several attractive arguments exist that make the reduction of wiring length a plausible constraint that has a strong influence on the spatial functional organization of the primate visual system (Chklovskii & Koulakov, 2004). However, further theoretical and experimental research is required to investigate this possibility with respect to the structure of receptive fields.

Early visual system fMRI effects of modifications of spatial structure in the visual input

In contrast to the rather large number of studies employing animal single cell electrophysiology or theoretical approaches to investigate the idea of sparse neural coding of spatial structure in natural scenes in V1, only a few previous fMRI studies were particularly designed to investigate this issue in the BOLD response. One of the first studies by Rainer et al. (2001) raised the

possibility that phase alignment might indeed have an influence on the strength of the population response, in particular at intermediate phase randomization levels. However, Dakin et al. (2002) argued on theoretical grounds that Rainer et al.'s approach to stimulus generation may have artificially created this result. Olman et al. (2004) investigated V1 contrast sensitivities for inputs with different spatial structure, a stimulus dimension orthogonal to phase coherence. Similar to Rainer et al. they found similar BOLD response levels at the endpoints of the phase randomization scale (intact and fully phase randomized pink noise). Unfortunately, these authors did not include intermediate phase randomization levels where Rainer et al. had reported their strongest effects. Dumoulin, Dakin, and Hess (2008) used binary (black and white) versions of natural scenes to investigate the influence of spatial structure on BOLD responses. However, the study used highly artificial stimuli with kurtosis matched among stimulus conditions, potentially at the cost of modifying the shape of the amplitude spectrum. This study cannot inform about the relationship between image statistics higher than an order of two (the RMS contrast) and the BOLD response. Tjan et al. (2006) used a similar approach to phase randomization as we did, although their study was designed to relate recognizability of natural scene content to the BOLD response level in different brain areas. Interestingly, these authors found in V1 nearly constant BOLD response levels at different phase randomization levels although their subjects performed a task on the images. To our knowledge our study is the first to specifically test the effects of precisely controlled manipulations of the spatial statistics of natural scene photographs in humans combining spatially precise fMRI with temporally precise MEG recordings.

The combination of BOLD with MEG measurements is important because the interpretation of the results obtained with either method alone is limited. The MEG is dominated by post-synaptic intracellular currents in cortical pyramidal cells oriented tangentially to the surface of the skull (Murakami & Okada, 2006). The BOLD signal measured in fMRI is of complex biochemical rather than bioelectric origin but its amplitude is correlated with the amplitude of the local field potentials (Logothetis & Wandell, 2004). The BOLD response can reflect postsynaptic activity that may not be visible to noninvasive MEG due to closed fields that cancel out in the distant MEG sensors (Nunez & Silberstein, 2000). It is thus reassuring that we neither find a modulation of the initial MEG nor of the BOLD response amplitude in V1 when we varied the phase noise level in our images. The exact physiological processes linking BOLD response to neuronal activity are multifactorial (Davis, Kwong, Weisskoff, & Rosen, 2008; Griffeth & Buxton, 2011)

and currently not well understood. It is generally agreed that an increase of neural activity triggers an increase of supply of fresh blood in the activated brain area to meet the increased metabolic demands (Sokoloff, 1977) and that cerebral blood flow is closely coupled to local energy metabolism but may overcompensate the actual demand (Goense & Logothetis, 2010; Logothetis & Wandell, 2004). In this sense the BOLD response can be considered a sensitive index for local changes in metabolic demand. Thus, the fact that we neither find a change in neural activity in MEG nor an index of a change in metabolic demands in the BOLD response strongly supports our conclusions that V1's metabolic demands do not differ between neural coding of natural scenes or phase randomized versions of them.

Caveats of phase randomization

Randomization of Fourier phase is a technique that is increasingly used to parametrically degrade the visibility of image features defining shapes and outlines of complex visual stimuli (e.g., Honey, Kirchner, & VanRullen, 2008; Philiastides & Sajda, 2006; Rainer et al., 2001; Rieger, Köchy, Schalk, Grüşchow, & Heinze, 2008; Sadr & Sinha, 2004; Tjan et al., 2006). It is considered an elegant way of controlling recognizability of image content to study mid- and high-level vision without introducing amplitude spectrum shape and offset distortions as it is the case with image scrambling (e.g., Rieger, Köchy, et al., 2008; Tjan et al., 2006). Manipulations of the Fourier amplitude spectrum can be problematic in studies of mid- and high-level vision because they lead to distortions of overall image contrast and image sharpness. As a consequence, effects of low-level contrast manipulations may confound effects of mid- and high-level processing of image content. Phase randomization can perfectly retain overall RMS contrast as well as shape and offset of the image's amplitude spectrum. However, the method offers a multiple potential pitfalls. For example, Dakin et al. (2002) pointed out that some phase randomization schemes might lead to overrepresentation of certain phase angles and, as a consequence, to local concentrations of image contrast, e.g., in the image corners (Rainer et al., 2001). Observed neural effects of phase randomization might then be attributable to changes in image contrast rather than phase manipulations. Another important aspect is that phase randomization only retains the original amplitude spectrum if all transformations applied to the final retinal image are linear. Of course, the fast Fourier transform and the inverse fast Fourier transform are linear and phase manipulations in complex Fourier space leave amplitude manipulations unaltered. How-

ever, an important step that needs to be taken care of after phase randomization is the nonlinear conversion of RGB pixel values to pixel luminance of the display system. This nonlinearity typically takes the form of a power law with an exponent around two. Careful calibration of the display ensuring a linear RGB value to pixel intensity relationship is necessary to prevent severe distortions of the amplitude spectrum of the retinal image. Unfortunately only a few studies using phase randomization report that such calibration procedures were employed (Rieger, Köchy, et al., 2008; Tjan et al., 2006; Wichmann, Braun, & Gegenfurtner, 2006). This opens the possibility that several studies using image phase randomization might be flawed.

Keywords: natural scenes, sparse coding, contrast energy coding, MEG, fMRI

Acknowledgments

We want to thank Robert Fendrich, Toemme Noesselt, and an anonymous reviewer for helpful comments on the manuscript. FS, NK, and JWR were supported by grants RI 1511/1-1 and RI 1511/1-3 from the Deutsche Forschungsgemeinschaft.

Commercial relationships: none.

Corresponding author: Jochem Rieger.

Email: jochem.rieger@uni-oldenburg.de.

Address: Institute of Psychology, Oldenburg University, Oldenburg, Germany.

References

- Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, *48*(1), 217–237. [PubMed]
- Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, *4*(2), 196–210.
- Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, *21*(10), 1133–1145. [PubMed]
- Baddeley, R., Abbott, L. F., Booth, M. C. A., Sengpiel, F., Freeman, T., Wakeman, E. A., et al. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*(1389), 1775–1783, doi:10.1098/rspb.1997.0246. [PubMed]

- Bell, A. J., & Sejnowski, T. J. (1997). The “independent components” of natural scenes are edge filters. *Vision Research*, 37(23), 3327–3338. [PubMed]
- Bethge, M. (2006). Factorial coding of natural images: How effective are linear models in removing higher-order dependencies? *Journal of the Optical Society of America A*, 23(6), 1253–1268, doi:10.1364/JOSAA.23.001253.
- Chklovskii, D. B., & Koulakov, A. A. (2004). Maps in the brain: What can we learn from them? *Annual Review of Neuroscience*, 27, 369–392. [PubMed]
- Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology*, 12(14), R476–R478. [PubMed]
- Davis, T. L., Kwong, K. K., Weisskoff, R. M., & Rosen, B. R. (1998). Calibrated functional MRI: Mapping the dynamics of oxidative metabolism. *Proceedings of the National Academy of Sciences*, 95(4), 1834–1839. [PubMed]
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22(5), 545–559, doi:10.1016/0042-6989(82)90113-4. [PubMed]
- De Valois, R. L. D., & De Valois, K. K. D. (1990). *Spatial vision*. New York: Oxford University Press.
- Dumoulin, S. O., Dakin, S. C., & Hess, R. F. (2008). Sparsely distributed contours dominate extra-striate responses to complex scenes. *Neuroimage*, 42(2), 890–901. [PubMed]
- Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., et al. (1994). fMRI of human visual cortex. *Nature*, 369(6481), 525, doi:10.1038/369525a0. [PubMed]
- Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biology*, 3(10), e342. [PubMed]
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12), 2379–2394.
- Field, D. J. (1993). Scale-invariance and self-similar ‘wavelet’ transforms: An analysis of natural scenes and mammalian visual systems. In M. Farge, J. Hunt, & J. C. Vassilicos (Eds.), *Wavelets, fractals and Fourier transform: New developments and new applications* (pp. 151–193). Oxford, U.K.: Oxford University Press.
- Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, 6(4), 559–601.
- Földiák, P. (2002). Sparse coding in the primate cortex. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks*, 2nd Ed. (pp. 1064–1068). Cambridge, MA: MIT Press.
- Goense, J., & Logothetis, N. K. (2010). Physiological basis of the BOLD signal. In M. Ullsperger & S. Debener (Eds.), *Simultaneous EEG and fMRI: Recording, analysis, and application* (pp. 21–46). New York: Oxford University Press.
- Graham, D. J., & Field, D. J. (2009). Natural images: Coding efficiency. *Encyclopedia of Neuroscience* (pp. 19–27). Oxford: Academic Press.
- Griffeth, V. E. M., & Buxton, R. B. (2011). A theoretical framework for estimating cerebral oxygen metabolism changes using the calibrated-BOLD method: Modeling the effects of blood volume distribution, hematocrit, oxygen extraction fraction, and tissue signal properties on the BOLD signal. *NeuroImage*, 58(1), 198–212, doi:10.1016/j.neuroimage.2011.05.077. [PubMed]
- Hancock, P., Baddeley, R., & Smith, L. (1992). The principal components of natural images. *Network: Computation in Neural Systems*, 3(1), 61–70, doi:10.1088/0954-898X/3/1/008.
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9(02), 181–197, doi:10.1017/S0952523800009640. [PubMed]
- Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of Vision*, 8(12):9, 1–13, <http://www.journalofvision.org/content/8/12/9>, doi:10.1167/8.12.9. [PubMed] [Article]
- Hyvärinen, A., Hurri, J., & Hoyer, P. O. (2009). Sparse coding and simple cells. *Natural Image Statistics: Computational Imaging and Vision*, 39, 131–150.
- Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3(7):4, 499–512, <http://www.journalofvision.org/content/3/7/4>, doi:10.1167/3.7.4. [PubMed] [Article]
- Kanowski, M., Rieger, J. W., Noesselt, T., Tempelmann, C., & Hinrichs, H. (2007). Endoscopic eye tracking system for fMRI. *Journal of Neuroscience Methods*, 160(1), 10–15. [PubMed]
- Kersten, D. (1987). Predictability and redundancy of natural images. *Journal of the Optical Society of America A*, 4(12), 2395–2400, doi:10.1364/JOSAA.4.002395. [PubMed]
- Laughlin, S. B., & Sejnowski, T. J. (2003). Communication in neuronal networks. *Science*, 301(5641), 1870–1874, doi:10.1126/science.1089662. [PubMed]
- Lennie, P. (2003). The cost of cortical computation. *Current Biology*, 13(6), 493–497, doi:10.1016/S0960-9822(03)00135-0. [PubMed]

- Levy, W. B., & Baxter, R. A. (1996). Energy efficient neural codes. *Neural Computation*, 8(3), 531–543, doi:10.1162/neco.1996.8.3.531. [PubMed]
- Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, 66(1), 735–769, doi:10.1146/annurev.physiol.66.082602.092845. [PubMed]
- Martínez, A., Di Russo, F., Anllo-Vento, L., & Hillyard, S. A. (2001). Electrophysiological analysis of cortical mechanisms of selective attention to high and low spatial frequencies. *Clinical Neurophysiology*, 112(11), 1980–1998, doi:10.1016/S1388-2457(01)00660-5. [PubMed]
- Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of Physiology*, 283(1), 53–77. [PubMed]
- Miller, A. K., Alston, R. L. & Corsellis, J. A. (1980). Variation with age in the volumes of grey and white matter in the cerebral hemispheres of man: Measurements with an image analyser. *Neuropathology and Applied Neurobiology*, 6, 119–132. [PubMed]
- Murakami, S., & Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *The Journal of Physiology*, 575(3), 925–936, doi:10.1113/jphysiol.2006.105379. [PubMed]
- Noesselt, T., Hillyard, S. A., Woldorff, M. G., Schoenfeld, A., Hagner, T., Jäncke, L., et al. (2002). Delayed striate cortical activation during spatial attention. *Neuron*, 35(3), 575–587, doi:10.1016/S0896-6273(02)00781-X. [PubMed]
- Nunez, P. L., & Silberstein, R. B. (2000). On the relationship of synaptic activity to macroscopic measurements: Does co-registration of EEG with fMRI make sense? *Brain Topography*, 13(2), 79–96, doi:10.1023/A:1026683200895. [PubMed]
- Olman, C. A., Ugurbil, K., Schrater, P., & Kersten, D. (2004). BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Research*, 44(7), 669–683, doi:10.1016/j.visres.2003.10.022. [PubMed]
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607–609, doi:10.1038/381607a0. [PubMed]
- Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4), 481–487, doi:10.1016/j.conb.2004.07.007. [PubMed]
- Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37(23), 3311–3325. [PubMed]
- Perna, A., Tosetti, M., Montanaro, D., & Morrone, M. C. (2008). BOLD response to spatial phase congruency in human brain. *Journal of Vision*, 8(10):15, 1–15, <http://www.journalofvision.org/content/8/10/15>, doi:10.1167/8.10.15. [PubMed] [Article]
- Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, 16(4), 509–518, doi:10.1093/cercor/bhi130. [PubMed]
- Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Current Biology*, 11(11), 846–854, doi:10.1016/S0960-9822(01)00242-1. [PubMed]
- Rieger, J. W., Köchy, N., Schalk, F., Grüschow, M., & Heinze, H.-J. (2008). Speed limits: Orientation and semantic context interactions constrain natural scene discrimination dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, 34(1), 56–76, doi:10.1037/0096-1523.34.1.56. [PubMed]
- Rieger, J. W., Reichert, C., Gegenfurtner, K. R., Noesselt, T., Braun, C., Heinze, H.-J., et al. (2008). Predicting the recognition of natural scenes from single trial MEG recordings of brain activity. *NeuroImage*, 42(3), 1056–1068, doi:10.1016/j.neuroimage.2008.06.014. [PubMed]
- Rozell, C. J., Johnson, D. H., Baraniuk, R. G., & Olshausen, B. A. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural Computation*, 20(10), 2526–2563. [PubMed]
- Sadr, J., & Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science*, 28(2), 259–287.
- Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8), 819–825, doi:10.1038/90526. [PubMed]
- Sokoloff, L. (1977). Relation between physiological function and energy metabolism in the central nervous system. *Journal of Neurochemistry*, 29(1), 13–26, doi:10.1111/j.1471-4159.1977.tb03919.x. [PubMed]
- Thomson, M. G. A., Foster, D. H., & Summers, R. J. (2000). Human sensitivity to phase perturbations in natural images: A statistical framework. *Perception*, 29(9), 1057–1069, doi:10.1068/p2867.
- Tjan, B. S., Lestou, V., & Kourtzi, Z. (2006).

- Uncertainty and invariance in the human visual cortex. *Journal of Neurophysiology*, 96(3), 1556–1568, doi:10.1152/jn.01367.2005. [PubMed]
- Tolhurst, D. J. (1972). Adaptation to square-wave gratings: Inhibition between spatial frequency channels in the human visual system. *Journal of Physiology-London*, 226, 231–248. [PubMed]
- van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1394), 359–366, doi:10.1098/rspb.1998.0303. [PubMed]
- VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, 42, 2593–2615. [PubMed]
- Vinje, W. E., & Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *The Journal of Neuroscience*, 22(7), 2904–2915. [PubMed]
- Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456), 1273. [PubMed]
- Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates.
- Weliky, M., Fiser, J., Hunt, R. H., & Wagner, D. N. (2003). Coding of natural scenes in primary visual cortex. *Neuron*, 37(4), 703–718. [PubMed]
- Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vision Research*, 46(8-9), 1520–1529, doi:10.1016/j.visres.2005.11.008. [PubMed]
- Willmore, B., & Tolhurst, D. J. (2001). Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems*, 12(3), 255–270. [PubMed]