Dissociation between saliency signals and activity in early visual cortex

Torsten Betz
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
Bernstein Center for Computational Neuroscience Berlin, Charité–Universitätsmedizin Berlin, Germany
Berlin Institute of Technology, Berlin, Germany

Niklas Wilming
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Carsten Bogler
Bernstein Center for Computational Neuroscience Berlin, and Berlin Center for Advanced Neuroimaging Charité–Universitätsmedizin Berlin, Germany

John-Dylan Haynes*
Bernstein Center for Computational Neuroscience Berlin, and Berlin Center for Advanced Neuroimaging Charité–Universitätsmedizin Berlin, Germany

Peter König*
Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Saliency is a measure that describes how attention is guided by local stimulus properties. Some hypotheses assign its computation to specific topographically organized areas of early human visual cortex. However, in most stimuli, saliency is correlated with luminance contrast, which in turn is known to correlate with activity in these early areas. Thus, any observed correlation of local activity with saliency might be due to the area encoding luminance contrast. Here we disentangle encoding of local luminance contrast and saliency by using stimuli where the two properties are uncorrelated. First, we conducted an eye-tracking study to verify that both negative and positive contrast modifications located in individual quadrants of the visual field increase saliency. Second, subjects viewed identical stimuli while fMRI signals were recorded. We find that positive contrast modifications induce a robust increase of activity in V1–V3 and hV4. However, negative contrast modifications lead to a reduced (V1, V2) or comparable (V3, hV4) activity level compared to unmodified quadrants. Furthermore, even with linear multivariate pattern-classification techniques, it is not possible to decode the location of the salient quadrant independent of the type of the contrast modification. Instead, decoding of the contrast-modified location is only possible separately for the two modification types in V1–V3. These findings suggest that the BOLD activity in V1–V3 is dominated by contrast-dependent processes and does not include the contrast invariance necessary for the computation of feature-invariant saliency.

Introduction

The brain continuously samples information by directing covert or overt attention towards different locations in the environment. In recent years, the underlying mechanisms of attentional selection have moved towards the center of research interest. An early hypothesis suggests that the brain computes a saliency map—a topographically organized representation of the visual field that can be used to decide
where to attend to next. The more salient a position is, the more likely it will be attended to (Koch & Ullman, 1985).

This hypothesis has triggered the search for brain areas that perform the required computations and encode saliency. Several regions in the cerebral cortex and subcortical areas have been suggested as the locus of a saliency map: the superior colliculus (Kustov & Robinson, 1996), the pulvinar (Shipp, 2004), V1 (Li, 1999; Li, 2002), the parietal cortex (Bisley & Goldberg, 2010; Geng & Mangun, 2009; Gottlieb, Kusunoki, & Goldberg, 1998; Serences et al., 2005), V4 (Mazer & Gallant, 2003), and frontal eye fields (Serences & Yantis, 2007; Thompson & Bichot, 2005). The term *saliency* is often used for task-dependent as well as stimulus-driven processes that might occur in different areas, an ambiguity that may lead to confusion. Here, we are only concerned with the latter. Li Zhaoping (Li, 2002; Zhang, Zhaoping, Zhou, & Fang, 2012; Zhaoping, 2011) argues that V1 creates a purely stimulus-driven (bottom-up) saliency map that relays information to higher areas.

In a network of interacting areas, it is difficult to identify and locate computations of saliency. Because V1 projects directly or indirectly to all the areas listed previously, it is reasonable to assume that saliency information would also be observable in these other areas if it is already computed in V1 (Shipp, 2004; Zhang et al., 2012). Any higher cortical area that receives information from many other areas is therefore likely to exhibit saliency-map-like properties because saliency-related information propagates up the hierarchy. To identify where saliency information is first made explicit, as opposed to simply received, it is therefore important to identify the exact contribution of early visual areas, like V1, to this computation.

A complication arises from the fact that early areas are usually considered to encode certain features of the stimulus (e.g., oriented edges in V1; Hubel & Wiesel, 1959), and these features in turn contribute to the saliency map (Itti & Koch, 2001). Thus, a strong response to a salient stimulus in an area need not reflect the explicit computation of saliency but potentially only the encoding of a feature that also influences the computation of saliency. For example, high-contrast edges strongly activate V1 neurons, and high-contrast edges correlate with saliency. But encoding of high-contrast edges in V1 does not necessarily form an explicit representation of saliency.

In this study, we investigate the contribution of early visual areas of human cortex (V1–V3 and hV4) to the computation of visual saliency. We address the dependency of contrast and saliency by exploiting a finding by Einhäuser & König (2003): Under certain conditions, a local reduction of luminance contrast leads to an increase in saliency. A brain region that explicitly encodes saliency would show an increased activity in response to local contrast reductions (saliency-encoding hypothesis). However, a brain region that encodes contrast would show reduced activity (contrast-encoding hypothesis). Hence, we created stimuli in which the luminance contrast in one of the four quadrants was either increased or decreased. An eye-tracking study confirms that both contrast modifications increased saliency. We then present these stimuli in an fMRI experiment and record blood oxygenation level dependent (BOLD) responses. The type of representation is characterized by analyzing the mean BOLD activity and by multivariate pattern classification in functionally defined regions of interest (ROI V1–V3, hV4). This allows differentiation between encoding of luminance and encoding of saliency.

### Methods

#### Stimuli

We used a set of pink-noise images as stimuli to avoid the influence of high-level factors and still retain some of the statistics of natural stimuli. We generated these stimuli by randomizing the phases in Fourier-transformed natural images, which removes all image structure but leaves the power spectrum untouched (Einhäuser et al., 2006). Twenty-seven source images were chosen randomly from the categories “natural” and “manmade” (Açık, Sarwary, Schultz-Kraft, Onat, & König, 2010). Their luminance histograms were flattened, and contrast increases and decreases were applied to each quadrant. Contrast modifications were computed as described before (Açık, Onat, Schumann, Einhäuser, & König, 2009; Einhäuser & König, 2003). These modifications were chosen because they tended to attract fixations in the studies mentioned previously. Our stimuli differed from these earlier studies in that the modified area was significantly larger, extending over an entire quadrant. This change was necessary in order to ensure a strong quadrant-specific response in the fMRI experiment. Furthermore, the modification includes spatial frequencies sufficiently low to influence saliency. Modified luminance values were given by

\[ I = (1 + zK)I_0 - |I_0 - M| + M, \]

where \(I_0\) denotes the source image, \(M\) the mean image, and \(z\) the peak modification level. The asterisk operator is a pointwise multiplication of matrices. The value of \(z\) was set to 0.9 for contrast increases and to −0.9 for contrast decreases. \(I, K,\) and \(M\) are matrices of the same size as \(I_0\). \(M\) contains the local mean...
Figure 1. (A–B) Examples of pink-noise stimuli with high and low contrast modifications. Note that the change in contrast is much more gradual and less visible if stimuli are viewed at their original size. (C) Time course of one eye-tracking trial. Each trial started with a central fixation on a gray screen, after which one pink-noise stimulus was shown until three to eight saccades were completed. In 49 out of 243 trials, a patch-recognition task was performed after stimulus offset. (D) Time course of an fMRI trial. In each trial, one image was presented repeatedly for 200 ms with a 200-ms gap. During a gap, the screen switched to a gray background. Between successive trials there was a variable interstimulus interval of 1 to 5 s. Observers had to detect the opening of one side of a central square. This task ran continuously throughout the session and independently of the pink-noise stimulation.

luminance values and is computed by convolving \( I_0 \) with a Gaussian kernel with a full width at half maximum of 6.53°. \( K \) describes how the modification level changes as a function of the distance to the peak modification. Here, \( K \) is given as the cosine of the square of the distance to the peak modification location:

\[
K(x, y) = \left( \cos\left(\frac{(x^2 + y^2)}{s}\right) + 1 \right)/2.
\]

This function has its maximum value of 1 at \( x = y = 0 \) and smoothly drops to 0 at \( (x^2 + y^2)/s = \pi \). The value of \( s \) was chosen such that the zero-crossing occurred at the edge of the modified quadrant in the horizontal direction. In the vertical direction, the modification slightly leaks into the adjacent quadrant, but this only corresponds to 0.74% of the kernel’s mass. To modify a specific quadrant, \( K \) was centered in the respective quadrant. All in all, we generated 243 different stimuli (27 stimuli \( \times \) 2 modifications \( \times \) 4 quadrants + 27 unmodified; see Figure 1 for examples).

**Eye tracking**

**Participants**

Eleven student volunteers took part in the eye-tracking experiment (four men and seven women; age range = 20–30 years, mean age = 25 years). All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye-tracking calibration with an average validation error below 0.3°. As compensation, participants received either payment or course credit. The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee.

**Apparatus**

The experimental apparatus was designed to resemble the parameters of the fMRI experiment. Screen distance was 80 cm to achieve the same stimulus size as later on in the scanner (26.6° \( \times \) 20.5°). Stimuli were presented on a 19-in. flat-screen monitor (SyncMaster 971p, Samsung Electronics, Seoul, South Korea) at a screen resolution of 1280 \( \times \) 1024 pixels and a refresh rate of 75 Hz. Eye movements were recorded with an Eyelink II system (SR Research Ltd., Mississauga, Ontario, Canada) at a sampling rate of 500 Hz. The system is capable of tracking both eyes, but only the eye that gave a lower validation error after calibration was recorded. A chin rest was used to minimize head movements. The experiment was conducted in a darkened room.

**Procedure**

The stimuli were presented in three blocks of equal length. In the break between blocks, we encouraged participants to rest and remove the eye tracker. Before each stimulus onset, drift correction was performed, requiring participants to fixate on the center of the screen. Subsequently, each stimulus was presented until a random number of saccades between three and eight had been performed (see Figure 1C). Each stimulus was presented once to each subject (243 trials per subject, 2,673 trials in total). The stimulus order was randomized, but the number of stimuli from each condition was the same in all blocks. The task was to recognize whether a patch of size 250 \( \times \) 250 pixels, presented after stimulus offset, was taken from the image just shown. The probability that the patch actually came from the previously seen image was 50%. Participants responded by pressing either the up arrow (for yes) or the down arrow (for no) on a regular keyboard. To shorten the duration of the experiment and to avoid fatigue or demotivation, the patch-recognition task was only presented after 49 randomly selected trials out of the 243. A test run, consisting of five images, was
performed in order to let the subjects gain experience with the task.

fMRI

Participants

Fourteen observers who did not know the purpose of the study participated in the fMRI study (eight women, six men; age range = 22–33 years, mean age = 27 years). They reported normal or corrected-to-normal vision and received payment for their participation. The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee. For all participants, detailed anatomical brain images were available from previous studies. The data of two observers had to be excluded from analysis. One fell asleep during the experiment, and for the other, retinotopic mapping was not successful.

Apparatus

Images were presented with a Sanyo Xtra Pro (Sanyo Electric Co., Ltd., Osaka, Japan) with a resolution of 1024 × 768 pixels. Only the central 800 × 600 pixels (26° × 20°) were in clear view, so images were resized to this resolution with bicubic interpolation. A Siemens 3T Magnetom (Siemens AG, Erlangen, Germany) was used to acquire functional magnetic resonance (MR) echo planar image (EPI) volumes, with 36 slices at an isotropic resolution of 3 mm³ (repetition time [TR] = 2000 ms; echo time [TE] = 24 ms; 36 axial slices; field of view [FOV] 192 × 192 × 108 mm; α = 0°). Structural images were acquired with a T1-weighted 3-D magnetization prepared rapid gradient echo (MPRAGE) with selective water excitation and linear phase encoding. Magnetization preparation consisted of a nonselective inversion pulse. The imaging parameters were inversion time (TI) = 650 ms, TR = 1300 ms, TE = 3.93 ms, α = 10°, spatial resolution 1 mm³ isotropic, two averages.

Procedure

The pink-noise experiment was divided into five different runs of stimulus presentations. After each run, participants could take a small break and relax their eyes. Each run lasted approximately 10 min and consisted of 108 presentations of pink-noise stimuli. Each run contained an equal number of stimuli from each condition, interleaved with 30 blank trials. In contrast to the eye-tracking experiment, a task was given that required fixating on the center of the screen. The outline of a 0.3° × 0.3° square was drawn in black in the middle of the screen. Every 1200 ms, either the left or right side of the square opened for 600 ms; participants had to indicate which side was opened by pressing one of two buttons with the middle or index finger of their right hand. The fixation task was independent of the stimulus presentation and ran continuously during each run. Each run started with 10 s of the fixation task without pink-noise stimulation. Each stimulus presentation consisted of one pink-noise stimulus being flashed three times for 200 ms in the background, with a 200-ms gap between flashes. Between flashes, the background was gray. The rationale for flashing the stimulus was that we wanted to mimic the condition of the eye-tracking experiment just before the target of the first free saccade is chosen. There, the stimulus appears with a “flash” relative to the gray background of the drift-correction screen. We chose to flash the stimulus several times to ensure that the pink-noise stimulus was not completely ignored, despite its being irrelevant for the fixation task. Additionally, this procedure was used in a previous study where saliency-processing effects could be decoded successfully (Bogler, Bode, & Haynes, 2011). The intertrial interval was variable, at 1 s, 3 s, or 5 s (see Figure 1C). Two hundred eighty-six functional MRI volumes were acquired in each run. A 42-slice whole-brain EPI image was also acquired to facilitate spatial normalization.

Retinotopic mapping and localization runs were conducted in a separate session on a different day. The retinotopic-mapping runs consisted of the presentation of a rotating wedge (5 c per 300 s) and an expanding ring (10 expansions per 300 s). This allowed for a functional definition of early visual areas, especially V1–V3 and hV4 (Wandell, Dumoulin, & Brewer, 2007; Warneking, 2002). To localize voxels that react to visual stimulation of a quadrant, flickering checkerboard patterns that were centered in one of the four quadrants were shown (quadrant localizer). These decreased in contrast according to the same spatial function that was used for the contrast modification in the pink-noise stimuli. One localizer run consisted of the sequential presentation of localizer images for all quadrants. Each image was presented for 7.5 s and changed polarity with a frequency of 10 Hz. The order of stimulation was upper left, upper right, lower left, then lower right; it was repeated 10 times. All in all, the mapping-and-localization session consisted of eight different runs in the following order: rotating wedge twice, expanding ring, rotating wedge twice, expanding ring, quadrant localizer twice. One hundred fifty-five volumes were acquired in each run, but no stimulation was present during the last 10 s (five volumes). Participants had the same fixation task as in the pink-noise experiment, with the only difference being that the fixation spot changed every 1000 ms.
Data processing

Functional brain scans were preprocessed with SPM2 (http://www.fil.ion.ucl.ac.uk/spm). The first five volumes of each experimental run were discarded to allow for magnetic relaxation effects. All volumes acquired during one experimental session were motion corrected, realigned to the initial scan of the experiment, and coregistered to the high-resolution anatomic image of the participant. For subsequent statistical analyses of the pink-noise experiment, a general linear model (GLM) with event-based and haemodynamic response function (HRF)-convolved regressors was estimated separately for each voxel. For every run, nine regressors were used that encoded stimulation onsets: one for no modification, four for a high contrast modification in one of the quadrants, and four for a low contrast modification in one of the quadrants. In addition, a constant regressor for each run was included. All analyses were carried out based on SPM parameter estimates for these regressors.

Definition of functional regions of interest

Visual areas V1–V3 and hV4 were functionally defined using well-established retinotopic mapping procedures (Wandell et al., 2007; Warnking, 2002). First, we segmented gray matter using FreeSurfer (Dale, Fischl, & Sereno, 1999). Next, the cortical surface was flattened with mrGray (Wandell, Chial, & Backus, 2000). Custom Matlab (MathWorks, Natick, MA) scripts were used to generate the flattened angular phase maps (Heinzle, Kahnt, & Haynes, 2011). Finally, we identified visual areas V1–V3 and hV4 by locating phase-reversal boundaries on these maps. The precise definition of V4 in humans is still debated (Goddard, Mannion, McDonald, Solomon, & Clifford, 2011). Here we use the definition proposed by Wandell et al. (2007). Figure 2 shows the outlines of these areas on a flattened cortex for one participant and one hemisphere. Note that some smoothing in Figure 2 is due to the visualization and was not present when determining ROI boundaries.

The quadrants of the visual field were localized by fitting a GLM with one event-based and HRF-convolved regressor for each quadrant stimulation onset. Quadrant ROIs contained voxels that were exclusively active when one particular quadrant was stimulated ($t$ test $p < 0.001$ uncorrected) but not during stimulation in one of the other three quadrants. Figure 2 shows outlines of quadrant-specific regions for one participant and one hemisphere. Note that voxels which showed specificity for more than one quadrant were discarded. Across subjects, 44%, 44%, 37%, and 31% (V1, V2, V3, and hV4) of the voxels that showed selectivity for some quadrant were selective for only that quadrant.

Multivoxel pattern analysis (MVPA)

The contrast-encoding hypothesis predicts that early areas show increasing activation with increasing contrast. This implies that a classifier can distinguish if a set of quadrant-specific voxels was stimulated with low, baseline, or high contrast. The saliency-encoding hypothesis predicts that the neural response to stimulation with equally salient low and high contrast modifications is comparable. This implies that a classifier can distinguish if a quadrant was stimulated with a low or high contrast modification, even if the training data do not contain information about the modification type.

To test these predictions, we trained support vector machines (SVMs) for classifying whether a quadrant received a contrast modification. The SVMs were trained to predict, based on the activation of voxels in one ROI (e.g., V1 Q1), whether the quadrant corresponding to the ROI (Q1) or another quadrant was stimulated with modified contrast. This implicitly compares activation within a ROI when it is stimulated with a contrast modification with the activation when it is stimulated with baseline contrast. We conducted three separate classification analyses: first, a high-contrast classifier that only received training and test data from conditions with a contrast increase in one quadrant; second, a low-contrast classifier that only received low-contrast stimulation-parameter estimates; third, a saliency classifier that received data from both conditions (twice as many individual data points as the modification-specific classifiers), based on the rationale that neurons encoding saliency should show similar responses to both modifications.

SVM classification for each region of interest was based on three pairwise comparisons, separating contrast modification in the quadrant corresponding to the ROI (stimulated) from modification in one of the other quadrants (not stimulated). For example, the high-contrast classification accuracy for the ROI V1 Q1 would be the mean of three accuracies: high contrast (hc) in Q1 versus hc in Q2; hc in Q1 versus hc in Q3; and hc in Q1 versus hc in Q4. Training and evaluation of the SVMs was performed in a leave-one-out cross-validation scheme. In each cross-validation step, SPM model-parameter estimates from four of the five experimental runs were used to train a classifier that predicted the location of the modified quadrant in the fifth run. This procedure was carried out for each participant and each ROI individually. The cost parameter was set to 1.

We also performed an analysis of the separation hyperplanes created by the SVMs in order to gain a better understanding of the representation of contrast in the different visual areas. If the neural response to stimulation with low and high contrast is the same, the SVMs should learn the same separating hyperplane for
comparing low contrast versus baseline and high contrast versus baseline. The saliency-encoding hypothesis therefore predicts that the normal vector to the separating hyperplane points in the same direction for comparing low contrast versus baseline and high contrast versus baseline. Conversely, the contrast-encoding hypothesis predicts that the two normal vectors point in opposite directions. This is illustrated geometrically in Figure 5B. We therefore computed the weight vectors (the normal vector to the separating hyperplane) for pairs of SVMs that predicted low contrast versus baseline and high contrast versus baseline stimulation based on data from the same runs. These weights were then averaged over the five runs and the three different non-ROI quadrants. In a next step, we computed the angle between those weight vectors. If the two weight vectors are completely independent—for example, if weight vectors are not consistent across runs of the experiment,—the expected angular value is 90°. Angles significantly above 90° indicate that the contrast response is greater than a potential saliency response, and angles significantly below 90° indicate a stronger saliency than contrast response.

**Results**

**High and low contrast modifications increase saliency**

The primary goal of our study was to disentangle computations of luminance contrast and saliency. We created images on which luminance contrast in a
quadrant was either decreased or increased. The attentional effect of these modifications was first investigated in an eye-tracking study. We recorded how often the first free fixation on an image fell into each quadrant. Analysis was restricted to the first fixation because its target is selected while the retinal stimulation is identical to the central fixation in the fMRI task. The distribution of fixations across the different quadrants is shown in Figure 3A. Each quadrant attracts more fixations in each of the two modification conditions than when it is unmodified (Figure 3B). This is backed by a two-factor repeated-measures analysis of variance with quadrant and modification as factors. Only the modification factor is significant (modification: \( p < 0.001 \); quadrant: \( p > 0.3 \); interaction: \( p > 0.5 \)). Importantly, both modifications are significantly different from baseline (high contrast versus baseline: \( p < 0.001 \); low contrast versus baseline: \( p < 0.001 \); \( t \) test). We conclude that both increases and decreases in local luminance contrast increase saliency in the modified quadrant by a comparable amount. Thus, these stimuli are suitable for disambiguating between the retinotopic processing of luminance contrast and saliency.

**Mean BOLD activity increases with contrast**

We analyzed how contrast modifications affected the mean BOLD response to the modified image regions in brain regions that process the visual input. We extracted GLM parameter estimates from all voxels in 16 functionally defined regions of interest (ROIs) corresponding to the four quadrants of the visual field in V1–V3 and hV4 (see Methods and Figure 2). The contrast-encoding hypothesis predicts low activity in quadrants stimulated with reduced contrast and high activity for high-contrast stimulation. The saliency-encoding hypothesis, in its strongest form, predicts increased activity for both types of modification compared to baseline. We analyzed activity averaged across quadrants in individual areas in the high-contrast condition, in the low-contrast condition, and for the unmodified images (Figure 4). A repeated-measures ANOVA with condition (high contrast, baseline, low contrast) and area (V1–V3 and hV4) as factors reveals significant main effects of both factors (\( p < 0.001 \)) and a significant interaction (\( p < 0.05 \)). Single-factor ANOVAs computed on the data of individual areas show that the effect of condition is significant throughout all areas (\( p < 0.001 \), Holm–Bonferroni corrected). We assessed the source of the significant effect with post hoc pairwise \( t \) tests. The differences between high contrast and baseline and between high contrast and low contrast are significant in all areas (\( p < 0.01 \)). The difference between low contrast and baseline, the latter inducing higher activity than the former, is only significant in V1 (\( p < 0.01 \)) and V2 (\( p < 0.05 \), all values Holm–Bonferroni corrected). In summary, high contrast leads to an increase in activity compared to both baseline and low contrast, but low
contrast, although salient, does not likewise lead to increased activity. To the contrary, if there is any difference between low contrast and the baseline, it is not in the direction predicted by the saliency-encoding hypothesis.

**MVPA supports the contrast-encoding hypothesis**

In principle, it is possible that neurons in V1–V4 encode saliency but that this information is represented in these areas in a way not accessible to an analysis of the activity level in the form of an averaged BOLD response. We used multivoxel pattern analysis (Haynes & Rees, 2006; Kriegeskorte, Goebel, & Bandettini, 2006) to test if information about the most salient quadrant can be decoded from the activity patterns in our ROIs. The contrast-encoding hypothesis predicts that the stimulation of a visual-field quadrant with a certain level of contrast leads to a specific pattern of activity in the ROI corresponding to that quadrant. It should thus be possible to decode whether the stimulus was modified in the quadrant corresponding to a ROI (see Figure 2). For example, given an activation pattern from ROI V3 quadrant 1 (Q1), induced by high contrast stimulation in either Q1 or a different quadrant, it should be possible to infer if Q1 or another quadrant was modified (see Figure 2). The same should hold for low contrast modifications. The saliency-encoding hypothesis furthermore predicts similar activation patterns for low and high contrast modifications, since both make a quadrant more salient (Figure 3). A classifier trained on both types of patterns combined should therefore be able to generalize and infer if a quadrant was modified even without knowledge of the modification type. Figure 4B shows the mean decoding accuracies above chance level (50%) achieved for the three different analyses (high contrast only, low contrast only, both contrasts mixed [saliency]) in V1–V3 and hV4. In areas V1 through V3, decoding accuracies were significantly above chance level for the high-contrast-only and low-contrast-only analyses ($t$ test across 12 subjects, $p < 0.05$). However, the decoding accuracy for the saliency analysis did not reach significance in any ROI. Since the difference between a significant result and a nonsignificant one is not necessarily itself significant (Gelman & Stern, 2006; Nieuwenhuis, Forstmann, & Wagenmakers, 2011), we also directly analyzed the differences in accuracy between the contrast classifiers and the saliency classifier. Here, we find that in areas V1 through V3, decoding accuracy is significantly higher for the contrast classifiers. Results for hV4 are not significant, but the trend goes in the direction predicted by the contrast-encoding hypothesis.

The analysis of the weights of the classifiers shows that on average, voxels have a positive weight for high contrast and a negative weight for low contrast (see Figure 5A; V1: 9 out of 12 subjects in the lower right quadrant; V2: 8 out of 12; V3: 7 out of 12; hV4: 6 out of 12). For areas V1–V3, the normal vectors for the two
hyperplanes (high contrast within ROI versus high contrast outside ROI and low contrast within ROI versus low contrast outside ROI) tend to point in opposite directions—that is, the angle between these vectors is significantly greater than 90° (Figure 5C; individual t test across 12 subjects, p < 0.05). All of these results are expected and corroborate the univariate analysis.

To ensure that the failure to decode saliency from early visual areas is not due to specific parameters chosen, we performed additional analyses. First, we trained SVMs on activation patterns from whole areas instead of only single quadrants. Second, we used anatomical ROI definitions from the SPM anatomy toolbox (Amunts, Malikovic, Mohlberg, Schormann, & Zilles, 2000; Eickhoff et al., 2005; Rottschy et al., 2007) instead of the functional ones. Neither of these changes, nor combinations thereof, affected the pattern of results reported above.

In summary, it is possible to decode whether a quadrant was modified when modification type is given. However, without this information it is not possible to decode whether a quadrant is salient. This suggests that V1–V3 and hV4 do not make the abstraction away from absolute changes in contrast to changes in saliency.

**Discussion**

We showed that low and high contrast modifications in pink-noise stimuli decouple saliency and contrast. Our eye-tracking data indicate that both types of modification increase saliency. This decoupling pro-
vides a tool for investigating saliency processing in fMRI BOLD responses in early topographically organized visual areas (V1–V3, hV4). The behavioral increase in saliency for the low contrast modifications is not mirrored in fMRI data. Instead, we found that the activity patterns of V1–V3 monotonically relate to stimulus contrast, not saliency.

In order to encode saliency for these stimuli, the visual system would have to increase its response to contrast deviations from the mean in both directions. Gardner and colleagues (2005) have shown that such a rectification operation may happen in hV4 during temporal-contrast adaptation. It might have been suspected that a similar mechanism for spatial variations in contrast is responsible for the behavioral saliency effect observed in our stimuli. We do not find evidence for this. Low contrast stimulation did not lead to an increased activity level compared to baseline, and saliency could not be decoded in hV4. However, contrast could also not be decoded in hV4, which might be indicative of a low signal-to-noise ratio. The question of whether saliency is encoded in hV4 can therefore not be conclusively addressed with our data.

The “V1 saliency hypothesis” (Li, 2002) states that activity in V1 creates a bottom-up saliency map. Specifically, the highest evoked V1 response of each visual-field location (i.e., a max operation over all features encoded for this location) gives the relative saliency of this location. There is psychophysical (Zhaoping & May, 2007) as well as physiological (Zhang et al., 2012) evidence supporting this hypothesis. At first sight, these data appear to be in conflict with the present results. However, their stimuli are not natural stimuli, but arrays of oriented bars or simple conjunctions of bars. It is known that the receptive fields of V1 neurons are highly tuned to such bars. Under these conditions, it is therefore plausible that processing in V1 contributes to a saliency map. Given the restricted stimulus set, focusing on oriented line elements, the intermediate results might be indistinguishable from the final saliency map. In contrast, we used stimuli with a power spectrum that is comparable to that of natural scenes. Recent work demonstrates that such stimuli induce qualitatively different dynamics in visual cortex than do gratings (Onat, König, & Jancke, 2011). Hence, our more complex stimuli might explain why we find that V1 BOLD activity only contributes one processing step on the way to a final saliency map. This is consistent with recent results on experimental blindsight in monkeys (Yoshida et al., 2012). Interestingly, there is even evidence for salient orientation pop-out stimuli which are represented in V4 rather than in V1 (Bogler, Bode, & Haynes, 2013). These results are not compatible with the predictions of a general saliency map localized in primary visual cortex.

It should be noted that our results do not rule out contributions of V1–V4 to the computation of saliency even in the low-contrast-modification condition. It might, for example, be that subpopulations of neurons in these areas compute saliency and that the activity of these subpopulations is swamped by the contrast-dependent activity changes of the majority of neurons. However, Zhang and colleagues (2012) do find an explicit attention-driven signal in BOLD responses in V1 regions even for stimuli that were not consciously perceived. Thus, it does not seem that the proposed V1 saliency map is in principle not discoverable with fMRI. This concern is further reduced by our use of decoding techniques. It has been shown that MVPA analyses can be successfully used to decode the activity of neuronal subpopulations below the spatial resolution of individual fMRI voxels (Haynes & Rees, 2006). But this is of course still no guarantee that decoding would have been successful in our case if saliency is encoded by a small set of neurons in early visual areas. However, more explicit representations of saliency are observable in higher visual areas with fMRI (Bogler et al., 2011). In summary, the most dominant feature in V1–V3, according to our analysis, is clearly luminance contrast and not saliency.

In conclusion, we report a case of behaviorally observable saliency that is not linearly driven by stimulus contrast. Our findings do not support the hypothesis that a saliency map, in the sense of an explicit representation of the most likely fixation target regardless of specific stimulus features, is found in V1–V3. It is conceivable that higher areas have to integrate feature-specific saliency information encoded in early processing stages.

Keywords: visual attention, saliency, visual cortex, functional MRI, eye tracking

Acknowledgments

This work was supported by two grants from the German Federal Ministry of Education and Research (BMBF grant 01GQ1001C, BMBF grant 01GQ0851) and by the EU through the project eSMCs (FP7-IST-270212) and ERC-2010-AdG #269716 - MUlti-SENSE, the Deutsche Forschungsgemeinschaft (GRK1589/1), and the Max Planck Society.

*J-DH and PK share senior authorship of this article. +TB and NW share first authorship of this article.

Commercial relationships: The authors wish to declare, for the avoidance of any misunderstanding, the following competing interests: Niklas Wilming, Torsten Betz, and Peter König hold stock in WhiteMatter Labs.
GmbH, who markets and sells predictions of a visual attention model. The authors believe that the reported results do not influence the fortune of the company (positively or negatively), and thus believe that there is no conflict of interest.

Corresponding author: Torsten Betz.
Email: torsten.betz@bccn-berlin.de.
Address: Modelling of Cognitive Processes, Technische Universität Berlin, Berlin, Germany.

References


