Scene statistics: neural representation of real-world structure in rapid visual perception

Groen, I.I.A.

Publication date
2014

Citation for published version (APA):
Chapter 3

Local contrast statistics are diagnostic of invariance of natural textures

Texture may provide important clues for real world object and scene perception. For these clues to be reliable, they should ideally be invariant to common viewing variations such as changes in illumination and orientation. Here, we examined whether natural image statistics derived from local contrast are diagnostic of texture invariance in the human brain under common viewing variations. In a large image database of natural materials, we found textures with local contrast statistics that varied substantially under viewing variations, as well as textures that remained relatively constant. To test whether textures with constant contrast statistics give rise to more invariant visual representations compared to other textures, we selected images of natural texture categories with either high or low variance (HV/LV) in statistics and presented these to human observers. In two distinct behavioral categorization paradigms, participants more often judged HV textures as 'different' compared to LV textures, showing that textures with constant contrast statistics are perceived as being more invariant. In a separate EEG experiment, evoked responses to single texture images (single-image ERPs) were collected. The results show that differences in contrast statistics correlated with differences in occipital single-image ERP amplitude at several stages of visual processing. Importantly, ERP differences between images of HV textures were mainly driven by illumination angle, which was not the case for LV images, for which differences were completely driven by texture category. These converging neural and behavioral results show that some natural textures are surprisingly invariant to illumination changes and that local contrast statistics are diagnostic of the extent of this invariance.

Published as:
Introduction

Despite the complexity and variability of everyday visual input, the human brain rapidly translates light falling onto the retina into coherent percepts. One of the relevant features used by the visual system to accomplish this feat is texture (Bergen and Julesz, 1983; Malik and Perona, 1990; Elder and Velisavljevic, 2009). Texture - "the stuff in the image" (Adelson and Bergen, 1991) - is a property of an image region that can be used by early visual mechanisms for initial segmentation of the visual scene into regions (Landy and Graham, 2004), to separate figure from ground (Nothdurft, 1991) or to judge 3D shape from 2D input (Malik and Rosenholtz, 1997; Li and Zaidi, 2000). The relevance of texture for perception of natural images was recently demonstrated by the finding that a computational model based on texture statistics could accurately predict human natural scene categorization performance (Renninger and Malik, 2004).

In general, a desirable property for visual features is perceptual invariance to common viewing variations such as illumination and viewing angle. Previously, invariance has been defined as a property of cognitive templates (e.g. Biederman, 1987) or as a 'goal' of visual coding to be achieved by multiple consecutive transformations along the visual pathway (Riesenhuber and Poggio, 1999; DiCarlo and Cox, 2007). Here, we explore another possible interpretation of invariance, namely as a real-world property that is present to a certain degree in natural stimuli. If visual features differ in terms of invariance, the brain might compute this property to determine which features are more invariant, and therefore more reliable cues for object and scene perception.

The effects of viewing conditions on textures have been previously studied by Geusebroek and Smeulders (2005), who showed that changes in recording conditions of photographs of natural materials are well characterized as changes in contrast statistics. Specifically, two parameters fitted to the contrast histogram of natural images described the spatial structure of several different materials completely. These parameters describe the width and outline of the contrast histogram (Figure 3.1A) and carry information about perceptual characteristics of natural textures such as regularity and roughness (Geusebroek and Smeulders, 2005). Recently, we reported that for natural scenes, the same statistics explain up to 80% of the variance of event-related potentials (ERPs) recorded from visual cortex (Ghebreab et al., 2009). We proposed that the two contrast parameters reflect relevant perceptual dimensions of natural images, namely the amount of contrast energy (CE) and spatial coherence (SC) in a scene (Groen et al., 2012a). Importantly, we found that these parameters can be reliably approximated in a biologically realistic way by summatating the output of local contrast filters modeled after LGN cells (Scholte et al., 2009), suggesting that these statistics may be available to visual cortex directly from its pre-cortical visual contrast responses.
Here, we asked whether these statistics are diagnostic of the degree of invariance of texture representations in the human brain. We approximated the contrast statistics of the large set of natural textures recorded under different viewing conditions in Geusebroek and Smeulders (2005), using the LGN model (Ghebreab et al., 2009; Scholte et al., 2009). The CE and SC of a substantial amount of these textures covaried with viewing conditions. However, the statistics of some textures remained remarkably constant under these variations. If the visual system is indeed sensitive to variability in contrast statistics, differences between textures in terms of this variability should have a consequence for their perceptual processing. Specifically, textures with constant contrast statistics across differences in viewing conditions may give rise more invariant representations compared to other textures.

To test this hypothesis, we asked whether perceptual invariance (Experiment 1 and 2) and invariance in evoked neural responses (Experiment 3) to natural textures under changes in viewing conditions was associated with variance in contrast statistics. We selected multiple texture images from the same material category (e.g. wool, bread, sand) that differed in two recording conditions: illumination angle and rotation (Figure 3.1B). Based on variance in contrast statistics, textures were labeled as either high variant (HV) or low variant (LV), see Figure 3.1C. Example images for each texture category are shown in Figure 3.1D.

In Experiment 1, human observers performed a same-different categorization task on pairs of images that were either from the same or a different texture category. We tested whether variance in contrast statistics influenced categorization accuracy: we predicted that compared to HV textures, images from LV textures would be perceived as more similar (i.e., higher accuracy on same-texture trials) and be less often confused with other textures (higher accuracy on different-texture trials), indicating higher ‘perceived invariance’. In Experiment 2, we addressed the same question using another behavioral paradigm – namely an oddity task - in which participants selected one of three images belonging to a different texture category. We predicted that when presented with two texture images from the same HV category, participants would more often erroneously pick one of these images as the odd-one-out, indicating lower ‘perceived invariance’ on these trials.

In Experiment 3, event-related EEG responses (ERPs) to individually presented texture images were collected to examine potential differences in neural processing between HV and LV textures and to evaluate the contribution of each of the two image parameters to the time course of ERP amplitude. Using regression and dissimilarity analyses, we related differences in image statistics to differences in single-image responses; an avenue that more researchers are beginning to explore (Philiastides and Sajda, 2006; van Rijsbergen and Schyns, 2009; Gaspar et al., 2011; Rousselet et al., 2011). The advantage of this approach relative to traditional ERP analysis (which is based on averaging many trials within a condition or an a priori-determined set of stimuli) is that it provides a richer and more detailed impression of the data. It also allows us to examine how differences between
individual images can give rise to categorical differences in a bottom-up way, i.e. by showing how individual images cluster together in terms of neural representations to form a single category.

The results show that variance in contrast statistics correlates with perceived texture similarity under changes in rotation and illumination, as well as differences in neural responses due to illumination changes. They suggest that low-level contrast statistics are informative about the degree of perceptual invariance of natural textures.

Materials and Methods

Computation of image statistics

Contrast filtering
We computed image contrast according to the standard linear-nonlinear model (Heeger et al., 1996). For the initial linear filtering step we used contrast filters modeled after receptive fields of LGN-neurons (Bonin et al., 2005). As described in detail in Ghebreab et al., (2009), each location in the image was filtered using Gaussian second-order derivative filters spanning multiple octaves in spatial scale, following Croner and Kaplan, (1995). Two separate spatial scale octave ranges were applied to derive the two contrast parameters. For the contrast energy parameter, each image location was processed by filters with standard deviations 0.16, 0.32, 0.64, 1.28, 2.56 in degrees; for the spatial coherence parameter, the filter bank consisted of octave scales of 0.2, 0.4, 0.8, 1.6 and 3.2 degrees. The output of each filter was normalized with a Naka-Rushton function with 5 semi-saturation constants between 0.15 and 1.6 to cover the spectrum from linear to non-linear contrast gain control in the LGN (Croner and Kaplan, 1995).

Response selection
From the population of gain- and scale-specific filters outputs, one filter response was selected for each location in the image using minimum reliable scale selection (Elder and Zucker, 1998): a spatial scale control mechanism in which the smallest filter with output higher than what is expected to be noise for that specific filter is selected. In this approach a scale-invariant contrast representation is achieved by minimizing receptive field size while simultaneously maximizing response reliability (Elder and Zucker, 1998). Similar steps, differing in detail, are implemented in standard feed-forward filtering models, e.g. Riesenhuber and Poggio, (1999). As previously (Ghebreab et al., 2009; Groen et al., 2012a), noise thresholds for each filter were determined in a separate set of images (a selection of 1800 images from the ImageNet database; Deng et al., 2009) and set to half a standard deviation of the average contrast present in that dataset per scale and gain.
**Approximation of Weibull statistics**

Applying the selected filter to each location in the image results in a contrast magnitude map. Based on the different octave filter banks, one contrast magnitude map was derived for the CE parameter and one for the SC parameter.

---

**Figure 3.1** Contrast statistics of natural images. **A)** Contrast histograms of natural images follow a Weibull distribution. Three natural images with varying detail and scene fragmentation are displayed along with a filtered (non-rectified) version of the same image. The homogenous, texture-like image of grass (upper row) contains many contrasts of various strengths; its contrast histogram approaches a Gaussian. The strongly segmented image of green leaves against a uniform background (bottom row) contains very few, strong contrasts; its histogram approaches power law. Most images, however, have intermediate histograms (middle row). The degree to which natural images vary between these two extremes can be assessed using two parameters of a Weibull fit to the histogram. The first parameter describes the histogram width: it reflects the mean local contrast strength, or contrast energy. The second parameter describes the histogram shape: it reflects the amount of scene fragmentation, or spatial coherence. **B)** The texture images were photographs of natural materials (e.g. wool, sand, bread) taken while rotation and illumination angle were manipulated. The materials were placed on a turntable, and recordings were made for rotations of 0, 60, 120 and 180 degrees. For each rotation, the material was illuminated by one of 5 different light sources (L1-L5). Technical details are listed on [http://staff.science.uva.nl/~alo/public_alot/](http://staff.science.uva.nl/~alo/public_alot/). **C)** Top: The 400 texture images (20 images for each texture material, i.e. category) plotted against their contrast statistics parameters contrast energy and spatial coherence. High-variant (HV) stimuli are colored in shades of blue to green, whereas low-variant stimuli (LV) are in shades of yellow to red. Bottom: Mean and standard deviation in contrast parameters per texture category; HV images in blue, LV images in red. **D)** Example images for each of the 20 texture categories that were used for experimentation: For each of the 20 texture categories, an example is shown for a 60º change in rotation, and for a change from middle to side or top illumination angle (L2 to L1/L3).
These contrast maps were then converted into two 256-bin histograms. It has been demonstrated that contrast distributions of most natural images adhere to a Weibull distribution (Geusebroek and Smeulders, 2002). The Weibull function is given by:

\[ p(r) = ce \left( \frac{r - \mu}{\beta} \right)^\gamma \]  

(1)

where \( c \) is a normalization constant and \( \mu, \beta, \) and \( \gamma \) are the free parameters that represent the origin, scale and shape of the response distribution, respectively. The value of the origin parameter \( \mu \) is generally close to zero for natural images and averaged out. The \( \beta \)-parameter varies with the range of contrast strengths present in the image. The \( \gamma \)-parameter describes the outline of the distribution and varies with the degree of correlation between local contrast values.

As mentioned, these two parameters can be determined in a more biologically plausible way: simple summation of X- and Y-type LGN output corresponds strikingly well with the fitted Weibull parameters (Scholte et al., 2009). Similarly, if the outputs of the multi-scale, octave filter banks (Ghebreab et al., 2009) used here are linearly summed, we obtain values that correlate even stronger with the Weibull parameters obtained from the contrast histogram at minimal reliable scale (Ghebreab, Scholte, Lamme and Smeulders, unpublished observations). For the current stimuli, the approximation based on summation correlated \( r = 0.99 \) and \( r = 0.95 \) with the \( \beta \)- and \( \gamma \)-parameter of the Weibull function, respectively. For all analyses presented here, these biologically realistic approximations based on linear summation were used instead of the fitted parameters. We refer to the approximation of the \( \beta \)-parameter as contrast energy (CE) and to the approximation of the \( \gamma \)-parameter as spatial coherence (SC).

**Experiment 1: Behavioral categorization with a same-different task**

**Subjects**

In total, 28 subjects participated in the first behavioral categorization experiment. The experiment was approved by the ethical committee of the University of Amsterdam and all participants gave written informed consent prior to participation. They were rewarded for participation with either study credits or financial compensation (7 euro for one hour of testing). The data from 2 participants was excluded because their mean behavioral performance was at chance level (50%).

**Stimuli**

Texture images (grayscale, 512x342 pixels) were selected from a large database of natural materials (http://staff.science.uva.nl/~aloi/public_alot/) photographed under various systematic manipulations (illumination angle, rotation, viewing angle and illumination color). In this chapter, we will refer to these materials as 'texture
categories'. For the stimuli used in the present study, a subset of images from each texture category was used that varied only in illumination angle (5 different light sources) and rotation (0, 60, 120 or 180 degrees); viewing angle (0 degrees azimuth) and illumination color (white balanced) were held constant. This subset contained 20 different images per texture category. For all of the 250 texture categories in the database, contrast statistics were computed for this subset of images. Based on the resulting CE and SC values, textures were classified as 'high variant' (HV) if their variance was more than half a standard deviation above the median variance for all textures, and as 'low variant' (LV) if it was half a standard deviation below the median, in both parameter values.

From the resulting LV and HV texture categories, 10 categories were randomly chosen and used in the experiment. However, care was taken that the parameter values of the selected categories were representative of the range of the entire image set. The final selection thus yielded 20 texture categories, of which half formed the HV condition and the other half formed the LV condition, with each category consisting of 20 images that were systematically manipulated in illumination angle and rotation (Figure 3.1D). In total, 400 images were used for experimentation.

**Procedure**

On each trial, two images were presented from either the same or a different texture category. Stimuli were presented on a 19-inch Dell monitor with a resolution of 1280x1024 pixels and a frame rate of 60 Hz. Participants were seated approximately 90 cm from the monitor and completed four blocks of 380 trials each. A block contained four breaks, after which subject continued the task by means of a button press. On each trial, a fixation cross appeared on the center of the screen; after an interval of 500 ms, a pair of stimuli was presented simultaneously for 50 ms, separated by a gap of 86 pixels (Figure 3.2A). A mask (see below) followed after 100 ms, and stayed on screen for 200 ms. Subjects indicated if the stimuli were from the same or a different texture category by pressing one of two designated buttons on a keyboard (‘z’ and ‘m’) that were mapped to the left or the right hand.

Within one block, one stimulus from each texture category was once paired with a stimulus from another texture category (190 trials). Stimuli were drawn without replacement, such that each image occurred once in each block, but were randomly paired with the images from the other texture category on each block. For the other 190 trials, the two stimuli were from the same texture category: here, for each texture category, 10 pairs were randomly chosen, resulting in 200 trials (20 from each texture category), from which 10 were then randomly deleted (but never more than one from each category) such that 190 trials remained. The ratio of different-category vs. same-category comparisons was thus 50:50, which was explicitly communicated to the subjects prior to the test phase. In addition, subjects were shown a few example textures, and performed 20 practice trials before starting the actual experiment. The example textures and practice trials contained comparisons of both
illuminations and rotation differences between the two presented texture images. None of the examples textures and practice trials occurred in the main experiment.

Masks were composed of all texture stimuli combined. They were created by dividing each of the 400 texture stimuli up in mini-blocks of 9x16 pixels: a single mask was created by drawing equal amounts of these mini-blocks from each stimulus and placing those at random positions in a frame of 512x342 pixels. Unique masks were randomly assigned to each of the 400 trials within a block, and were repeated over blocks. Per trial, the same mask was presented at both stimulus locations. Stimuli were presented using Matlab Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Data analysis
Mean accuracy for each subject was determined by calculating percent correct over the four blocks. It was determined separately for different-category vs. same-category comparisons and for trials on which two HV categories were compared vs. trials on which two LV categories were compared (2x2 design). Different-category trials where

---

**Figure 3.2 Methods and experimental design.**

A) Experimental paradigm of Experiment 1 (behavioral same-different task). Participants performed a same-different categorization task on pairs of texture stimuli presented on a grey background that were masked after 150 ms.

B) Experimental paradigm of Experiment 2 (behavioral oddity task); participants had to choose the odd-one-out from three images on each trial.

C) Experimental set-up and analysis of Experiment 3 (single-image EEG). Subjects were presented with individual presentations of texture images while EEG was recorded. Single-image evoked responses were extracted for each electrode, after which a regression analysis of the amplitude at each time-point based on contrast statistics was performed.

D) Representational dissimilarity matrices (RDMs) were computed at each sample of the single-image ERPs recorded from channel Oz. A single RDM displays dissimilarity (red = high, blue = low, reflecting difference in ERP amplitude) between all pairs of stimuli at a specific moment in time.
HV categories were compared to LV categories were excluded from analysis. Due to a programming error, no reaction times were recorded. As a consequence, trials in which subjects may have responded too fast (for instance before 200 ms) were still included in the analysis. This results in a potential underestimation of the error rate.

**Experiment 2: Behavioral categorization with an oddity task**

**Subjects**
In total, 18 subjects participated in the second behavioral categorization experiment, which was approved by the ethical committee of the University of Amsterdam. All participants gave written informed consent prior to participation. They were rewarded for participation with either study credits or financial compensation (7 euro/ hour). The data from 2 participants was excluded because mean performance was at chance level (33%, 1 participant) or because reaction times demonstrated an outlier (> 2 standard deviations from the mean across all participants, 1 participant).

**Stimuli and procedure**
The same set of 400 texture images was used as in the first behavioral experiment. However, for this task, on each trial three images were presented: two images from the same texture category (the *same pair*), and one from a different category (the *odd-one-out*), see Figure 3.2B. Stimuli were presented on a 19-inch ASUS monitor with a resolution of 1920x1080 pixels and a frame rate of 60 Hz. The procedure was identical to the first behavioral experiment, except that now, two images were positioned adjacent to each other and the third image was located either below or above the other two (position was counterbalanced across participants). The three images were separated by equal gaps of 120 pixels. The position of the odd stimulus was randomized over trials. Subjects were instructed to indicate which image was from a different texture by pressing one of three designated buttons on a keyboard (‘1’, ‘2’ and ‘3’ on the NUM pad of the keyboard) using their right hand.

Within one block, each texture category was twice paired with every other texture category by randomly drawing an image from both categories (380 trials). For one half of the trials, the image from the first category was designated as the odd-one-out, whereas from the second category, another stimulus was drawn to form the second half of the same pair. For the other half of the trials, the procedure was reversed, such that each texture category once formed the odd stimulus, and once formed the paired stimulus. Compared to the first experiment, the trials were thus always different-category trials, but each texture category was once the odd-one-out or the same-pair. This allowed us to test whether the variance in contrast statistics of the same-pair influenced performance: we predicted that increased variance in contrast statistics of the same-pair stimuli would lead to more errors (i.e. selecting one of the same-pair as the odd-one-out).
Data analysis

In contrast to the first experiment, reaction times (RT) were now properly recorded and trials in which the participant responded before 200 ms after stimulus-onset were excluded. To allow comparison with the same-different accuracy data from the previous experiment, we first computed accuracy for only the trials on which either two HV or two LV texture categories were compared (ignoring trials on which one HV and one LV category were compared). The same comparison was done for RT. In a subsequent analysis, we included all trials but split them into two groups in two different ways: namely based on 1) whether the odd-one out was LV or HV or 2) whether the same-pair images were HV or LV. This allowed us to test whether the variance of the odd stimulus vs. the variance of the same-pair was more predictive of errors in selection of the odd stimulus.

Experiment 3: EEG experiment

Subjects

Seventeen volunteers participated and were rewarded with study credits or financial compensation (7 euro/hour for 2.5 hours of experimentation). The data from 2 subjects was excluded because the participant blinked consistently shortly after trial onset in more than 50% of the trials (1 subject) of because their vision deviated from normal (1 subject) which became clear in another experiment conducted in the same session. The study was approved by the ethical committee of the University of Amsterdam and all participants gave written informed consent prior to participation.

EEG data acquisition

The same set of stimuli was used as in the behavioral experiment. In addition, for each image a phase-scrambled version was created, which were presented randomly intermixed with the actual textures. Stimuli were presented on an ASUS LCD-screen with a resolution of 1024x768 pixels and a frame rate of 60 Hz. Subjects were seated 90 cm from the monitor such that stimuli subtended 11x7.5° of visual angle. During EEG acquisition, a stimulus was presented one at a time in the center of the screen on a grey background for 100 ms, on average every 1500 ms (range 1000 - 2000 ms). Each stimulus was presented twice, in two separate runs. Subjects were instructed to indicate on each trial whether the image was an actual texture or a phase-scrambled image: examples of the two types of images were displayed prior to the experiment. Response mappings were counterbalanced between the two separate runs for each subject. Stimuli were presented using the Presentation software (www.neurobs.com). EEG Recordings were made with a Biosemi 64-channel Active Two EEG system (Biosemi Instrumentation BV, Amsterdam, NL, http://www.biosemi.com/) using the standard 10-10 systems with two additional occipital electrodes (I1 and I2), which replaced two frontal electrodes (F5 and F6). Eye movements were monitored with a horizontal and vertical electro-oculogram
(EOG) and were aligned with the pupil location when the participants looked straight ahead. Data was sampled at 256 Hz. The Biosemi hardware is completely DC-coupled, so no high-pass filter is applied during recording of the raw data. A Bessel low-pass filter was applied starting at 1/5th of the sample rate.

**EEG data preprocessing**
The raw data was pre-processed using Brain Vision Analyzer (BVA) by taking the following steps: 1) offline referencing to earlobe electrodes, 2) applying a high-pass filter at 0.1 Hz (12 dB/octave), a low-pass filter at 30 Hz (24 dB/octave) and two notch filters at 50 (for line noise) and 60 Hz (for monitor noise), 3) automatic removal of deflections larger than 250 µV. Trials were segmented into epochs starting 100 ms before stimulus onset and ending 500 ms after stimulus onset. These epochs were corrected for eye movements by removing the influence of ocular-generated EEG using a regression analysis based on the EOG channels (Gratton and Coles, 1983). Baseline correction was performed based on the data between -100 ms and 0 ms relative to stimulus onset; artifacts were rejected using maximal allowed voltage steps of 50 µV, minimal and maximal allowed amplitudes of -75 and 75 µV and a lowest allowed activity of 0.50 µV. The resulting event-related potentials (ERPs) were converted to Current Source Density (CSD) responses (Perrin, 1989). Trials in which the same image was presented were averaged over the two runs, resulting in one event-related potential specific to each image (single-image ERP; Groen et al., 2012a).

**Regression analyses on single-image ERPs**
To test whether differences between neural responses correlated with differences in contrast statistics between images, we conducted regression analyses on the single-image ERPs (Figure 3.2C). We first performed this analysis on ERPs averaged across subjects to test whether contrast energy and spatial coherence could explain consistent differences between images. For each channel and time-point, the image parameters (contrast energy and spatial coherence) were entered together as linear regressors on ERP amplitude, resulting in a measure of model fit ($r^2$) over time (each sample of the ERP) and space (each electrode).

To statistically evaluate the specific contribution of each parameter to the explained variance for the two different image conditions (HV en LV), we next ran regressions at the single subject level. These analyses were restricted to electrode Oz. We constructed a model with four predictors of interest (constant term + LV contrast energy, HV contrast energy, LV spatial coherence, HV spatial coherence). The obtained β-coefficients for each predictor were subsequently tested against zero by means of t-tests, which were Bonferroni-corrected for multiple comparisons based on the number of time-points for which the comparison was performed (154 time samples).
Finally, to test whether each predictor contributed unique variance, we conducted a stepwise version of a regression analysis with two predictors (contrast energy and spatial coherence for LV and HV images combined in one column) for each single subject. In this analysis, a predictor was entered into the model if it was significant at $\alpha < 0.05$, and removed if $\alpha > 0.10$. In the initial model, none of the parameters were included. We then counted, for every time-point, for how many subjects the full model was chosen, only one of the predictors was included, or no parameters were included.

**Representational similarity analysis**

To better examine how variance between individual visual stimuli arises over time, and how differences between individual images relate to image variance (HV/LV) and image manipulations (rotation and illumination angle), we computed representational dissimilarity matrices (RDMs; Kriegeskorte et al., 2008) based on single-image ERPs recorded at channel Oz. We computed, for each subject separately, at each time-point, for all pairs of images the difference between their evoked ERP amplitude (Figure 3.2D). As a result we obtained a single RDM containing 400x400 dissimilarity values between all pairs of images at each time-point. Within one matrix, the pixel value of each cell reflects the difference in ERP amplitude between two images indicated by the row and column number.

**Comparison between dissimilarity matrices**

To compare the observed dissimilarities in single-image ERPs with corresponding differences in contrast statistics between the images, we computed a dissimilarity matrix based on the image parameter values CE and SC. For each pair of images, we computed the sum of the absolute differences between the (normalized) contrast energy (CE) and spatial coherence (SC) values of those two images, e.g. $(CE_{image1} + SC_{image1}) - (CE_{image2} + SC_{image2})$, resulting in one difference value reflecting the combined difference in image parameters between the images (see also Chapter 2). For each subject, this matrix was compared with the RDMs based on the ERP data using a Mantel test for two-dimensional correlations (Daniels, 1944), denoted as $r_m$.

**Computation of luminance and AIC-values**

To obtain a measure of luminance for each image, we computed the mean luminance value per image (LUM) by averaging the pixel values (0-255) of each individual image. For the EEG analysis, to compare the regression results based on LUM with those obtained with contrast statistics, we used Akaike's information criterion (AIC; Akaike, 1973). The AIC-values were computed by transforming the residual sum of squares (RSS) of each regression analysis using:

$$AIC = n*\log(RSS/n) + 2k$$  \hspace{1cm} (2)
where \( n \) = number of images and \( k \) = number of predictor variables (\( k = 2 \) for contrast statistics, and \( k = 1 \) for LUM). AIC can be used for model selection given a set of candidate models of the same data; the preferred model has a small AIC-value.

**Results**

**Experiment 1: Same-different categorization**

Categorization accuracy was calculated separately for high-variant (HV) and low-variant (LV) trials and for same-category and different-category comparisons (**Figure 3.3A**). A repeated-measures two-way ANOVA indicated a significant main effect of variance (\( F(1,25) = 298.9, p < 0.0001 \)), but not of type of comparison (\( F(1,25) = 3.6, p = 0.07 \)); however, the interaction between variance and comparison was significant (\( F(1,25) = 61.8, p < 0.0001 \); **Figure 3.3B**). Subsequent paired t-tests revealed that participants performed better for LV than HV textures at both different-category (\( t(25) = 6.1, p < 0.0001, \text{mean difference} = 6\%, \text{ci} = 4\%-8\% \)) and same-category comparisons (\( t(25) = 16.3, \text{mean difference} = 17\%, \text{ci} = 15\%-19\%, p < 0.0001 \)), but also for different-category HV comparisons relative to same-category HV comparisons (\( t(25) = 3.4, \text{mean difference} = 11\%, \text{ci} = 4\%-17\%, p = 0.002 \)). These results show that participants generally made more errors on trials in which they compared two different HV texture categories than on trials in which they compared two LV categories. In addition, they more often incorrectly judged two images from the same HV texture category as different than vice versa (two different HV images as the same category).

**Figure 3.3 Results from Experiment 1: same-different categorization**

A) Accuracy scores for individual subjects according to task conditions. Subjects compared pairs of images that were either from different (circles) or same (squares) texture categories with either low (LV) or high variance in contrast statistics (HV). Comparisons between LV and HV categories were excluded from analysis. B) Mean accuracy per condition, showing an interaction effect between texture variance (HV; blue vs. LV; red) and type of comparison (same vs. different trial). Performance was worst on HV same texture trials, suggesting subjects viewed these textures as different even though they were from the same category. C) Accuracy on same-texture trials correlates with texture category variance in contrast statistics. Error bars indicate S.E.M.
This result shows that 1) LV texture categories are easier to categorize than HV categories and 2) images from the same HV texture category are perceived as less similar. This latter conclusion is supported by an additional analysis performed on the accuracy scores, in which we correlated the specific amount of variance in contrast statistics with the average number of same-texture errors. We found that variance in contrast statistics correlated with same-texture accuracy across all texture categories (Spearman’s \( \rho = -0.73 \), \( p < 0.0001 \); Figure 3.3C). This result suggests that, for high-variant stimuli, the specific amount of variance in contrast statistics influences perceived similarity of same-texture images: more variance implies less similarity.

Because subjects always compared only two images on each trial, we cannot be certain to what degree they based their judgment on the between-stimulus differences versus the difference of these images compared to all other images in the stimulus set. To investigate this more explicitly, we conducted another behavioral experiment using an oddity task, in which each trial consisted of three images that were drawn from two different texture categories. In this task, subjects always made a difference judgment: they had to pick the most distinct stimulus (the odd-one-out) and thus actively compare differences between texture categories with differences within texture categories. If variance in contrast statistics of a texture category indeed determines its perceived invariance, we would expect that for comparisons between images from HV texture categories, it is more difficult to decide which stimulus is different.

**Experiment 2: Oddity categorization**

Categorization accuracy on comparisons of HV texture categories was significantly lower compared to comparisons of LV texture categories (\( t(15) = 14.4 \), mean difference = 17%, \( ci = 14-20\% \), \( p < 0.0001 \); Figure 3.4A). Participants were also significantly faster on LV trials compared to HV trials (\( t(15) = -3.5 \), mean difference = 27 ms, \( ci = 10-43 \) ms, \( p < 0.004 \)). If we compute accuracy across all possible comparisons of texture categories (also including HV-LV/LV-HV comparisons), and split the data either according to the variance of the odd stimulus, or to the variance of the same-pair on each trial, we see that it is specifically the variance of the same-pair images that predicts differences in accuracy (Figure 3.4B). On trials at which the same-pair was from a HV texture category, subjects more often incorrectly chose one of that pair as the odd-one-out. If we correlate the amount of variance in contrast statistics of the same-pair with accuracy, we again find a significant correlation (\( \rho = -0.75 \), \( p < 0.0001 \); Figure 3.4C), indicating that with increasing variance in contrast statistics, images from the same texture category are more often perceived as different.
Overall, the results of the two behavioral experiments indicate that low variance in contrast statistics allows observers to more accurately categorize images of natural textures. Images of a texture category with constant statistics under different viewing conditions are more often accurately recognized as coming from the same category compared to images from categories with variable statistics. Conversely, categories with variable statistics are more often seen as different even though they are from the same category, and more often confused with other categories. This suggests that texture categories with little variance in contrast statistics are perceived as more invariant.

**Experiment 3: EEG**

*Contrast statistics explain variance in occipital ERP signals*

As a first-pass analysis, we first averaged single-image ERPs over subjects, after which a simple regression model with two predictors (contrast energy and spatial coherence) was fitted based on ERPs at every channel and time-point. Despite individual differences between subjects in EEG responses (e.g. in mean evoked response amplitude, likely due to individual differences in cortical folding), this analysis revealed a highly reliable ERP waveform time-locked to the presentation of the stimulus (Figure 3.5A). This time-locked ERP nonetheless varied substantially between individual images, mostly between 100 and 300 ms after stimulus-onset. The regression results show that early in time, nearly all mean ERP variance is explained by the image parameters CE and SC (maximal $r^2 = 0.94$ at 148 ms, $p <$
0.0001 on channel Oz, Figure 3.5B). Also at later time-points and at other electrodes, there is substantial (e.g., more than 50%) explained variance. If we examine the results for all channels simultaneously (Figure 3.5C), we see that it is highest at occipital channels, extending towards parietal and lateral-occipital electrodes. This spatial pattern is similar for early and late time-points (i.e., mostly central-occipital).

This result shows that low-level image statistics can explain a high amount of variance, both early and late in time, of image-specific differences across participants. To test more precisely 1) whether these effects were present in all participants, 2) which of the two image parameters contributed most to the explained variance and 3) whether these contributions differed between the two conditions (LV/HV), we selected the electrode with the highest $r^2$-value (Oz) and conducted regression analyses at the single-subject level using a model containing four parameters (see Materials and Methods): LV contrast energy, HV contrast energy, LV spatial coherence, HV spatial coherence.

Mean explained variance across subjects peaked 156 ms after stimulus onset ($r^2 = 0.65$, mean $p < 0.0001$, Bonferroni-corrected; Figure 3.6A); maximal values for individual subjects ranged between $r^2 = 0.49$-$0.85$ at 144-168 ms after stimulus-onset and were all highly significant (all $p < 0.0001$). If we compare the time courses of the $\beta$-coefficients associated with each predictor (Figure 3.6B), we observe that CE and SC have distinct time courses. Statistical evaluation of the $\beta$-coefficients (Figure 3.6C) shows that ERP amplitude at an early time interval is mostly correlated with CE (between 136-183 ms, all $t(15) < -5.1$, max $t(15) = -9.0$, all $p < 0.0003$), which correlates again much later in time (between 305-340 ms, all $t(15) < -5.1$, max $t(15) = -6.5$, all $p < 0.0001$). SC only contributes significantly to the explained variance.
between 220-240 ms, all \( t(15) > 4.7 \), max \( t(15) = 6.1 \), all \( p < 0.003 \); again between 274-330 ms, all \( t(15) > 5.4 \), max \( t(15) = 9.0 \), all \( p < 0.0003 \). Importantly, the temporal profiles of CE and SC are similar for HV and LV images; differences between the \( \beta \)-coefficients of these two conditions are relatively small (Figure 3.6D). For both CE and SC, the difference between HV and LV \( \beta \)-coefficients appears somewhat larger at two time-intervals between 150 and 300 ms, but statistical tests of these differences were right at the threshold of Bonferroni-corrected significance (CE at 223 ms, \( t(15) = -4.8 \), \( p = 0.0002 \); SC at 285 and 289 ms, \( t(15) = 4.6 \), \( p = 0.0003 \); Figure 3.6E). We thus cannot conclude that the CE and SC values are differentially involved in processing of LV or HV textures; they play a similar role in both cases.

Finally, to test whether CE and SC explained any unique variance, we conducted stepwise regression analyses (see Materials and Methods) on the single subjects ERPs. At each time-point of the ERP, we counted for how many participants a) either the full model was chosen or b) only one predictor was included in the model (Figure 3.6F). The results show that early in time, CE alone is preferred over the full model, but that later (from \( \sim 200 \) ms), SC is included for most subjects. Especially later in time, SC thus adds additional explanatory power to the regression model.

Figure 3.6 Regression analyses on single-subject ERP data at channel Oz. A) Explained variance for each individual subject (colored lines) and averaged across subjects (black line) based on a regression model with four predictors (see Methods). B) Mean \( \beta \)-coefficient at each time-point associated with each of the four regression model predictors. CE = contrast energy, SC = spatial coherence, LV = low-variant, HV = high-variant. Shaded areas display confidence intervals obtained from a t-test of each predictor against zero across single subjects. C) Resulting t-statistic of testing the \( \beta \)-coefficient associated with each predictor against zero for every time-point of the ERP: the gray dashed line indicates significance level at \( \alpha < 0.05 \) when correcting for multiple comparisons (Bonferroni-correction). D) Mean difference in \( \beta \)-coefficients between LV and HV images for each image parameter. Shaded areas display confidence intervals obtained from a t-test between LV and HV coefficients. E) Resulting t-statistic of the difference between LV and HV coefficients: gray dashed line indicates Bonferroni-corrected significance level at \( \alpha < 0.05 \). F) Results of the single-subject stepwise regression analysis: displayed are the number of subjects for which either only contrast energy (black solid line), only spatial coherence (dotted black line) or both parameters (green dashed line) were included in the model at each time-point of the ERP.
These regression results show that differences in neural activity are strongly related to differences in contrast statistics. Early ERP amplitude differences between individual images correlates with variance in contrast statistics for both high-variant and low-variant textures to a similar degree. Whereas CE explains most variance early in time, at later time intervals both parameters become significantly correlated with ERP amplitude. Variance in contrast statistics thus leads to extensive and temporally sustained modulations of evoked neural activity, suggesting that the brain is sensitive to differences between texture images in this property. However, they do not reveal how these differences are related to texture category, or variations in recording conditions. We investigated this in the next section.

Dissimilarities between images map onto contrast statistics

To examine the origin of the variance between images in both contrast statistics and evoked activity, we computed (for each subject separately) representational dissimilarity matrices (RDMs) based on differences in evoked responses between individual images (see Materials and Methods and Figure 3.2D). In brief, to build an RDM, we compute for each possible combination of individual images the difference in evoked ERP amplitude, and convert the result into a color value. The advantage of this approach is that RDMs allow us to see at once how images are (dis)similar to all other images, and how this relates to texture category membership.

To present the results of this analysis, we selected the RDM at the time-point of maximal explained variance for the subject-averaged ERP regression analysis (148 ms after stimulus-onset; see above) for each subject and simply averaged the resulting matrices across subjects. In this RDM (Figure 3.7A), every consecutive 20 rows/columns index all images from one specific texture category. The categories are sorted according to their mean contrast energy and spatial coherence values (i.e. distance from zero in the image statistics space in Figure 3.1C). If we simply visually examine the RDM, we see that differences between HV images (lower right quadrant) occur at different locations in the RDM than for LV images (upper left quadrant). Specifically, HV stimuli have large differences within textures, whereas for LV stimuli, the differences are largest between textures: within a 20x20 square, images are 'similarly dissimilar' from other textures. LV images thus cluster more by texture category than HV images, which differ strongly within a given texture.

Next, we tested to what extent these image-specific differences between images in the ERP were similar to differences in contrast statistics. We calculated another 400x400 difference matrix, in which we subtracted the parameter values of each image from the values of each other image (Figure 3.7B, see Methods). Based on visual inspection, it is already clear that the relative dissimilarities between individual images in contrast statistics are very similar to the ERP differences. Inter-matrix correlations at each time-point (Figure 3.7C) showed that the RDM of the ERP signal correlated significantly with the difference matrix based on contrast statistics for each subject; between 137 and 227 ms after stimulus onset, the correlation was
significant for all 17 subjects (range of maximal $r_m = 0.31$-0.72, all $p < 0.01$, Bonferroni-corrected). This shows that not only the variance, but also the categorical clustering of images in terms of neural activity is strongly related to contrast statistics.

**Dissimilarities between high-variant stimuli reflect illumination changes**

Presumably, the higher dissimilarities within HV textures result from variability in responses driven by changes in recording conditions. To isolate these effects, we computed demeaned versions of the RDMs, by dividing the evoked response to each image by the mean response to all 20 images of its category. As a result, we obtain RDMs that only reflect differences in single-image responses relative to the mean response to the category, ignoring differences between the category means. Analogously, for the contrast statistics RDM, we divided differences in CE and SC values between images of each category by the mean values of the category.

### Figure 3.7 Representational dissimilarity analysis.

**A)** Mean RDM of the ERP signal at the time-point of maximal dissimilarity. Each cell of the matrix reflects dissimilarity (red = high, blue = low) between two individual images, indexed on the x- and y-axis. Starting from the top row, the first 200 images are from LV textures; the bottom half are HV textures. This yields a 'LV quadrant' (top left) and a 'HV quadrant' (bottom right). Within these quadrants, each 20 rows/columns index images of a single texture category; categories are sorted according to their mean CE and SC values (distance from 0 in Figure 3.1C).

**B)** Dissimilarity matrix based on difference in contrast statistics (CE and SC combined).

**C)** Correlation between the ERP RDM and contrast statistics RDM (Mantel test) over time. Colored lines: single subject correlations; black line: mean RDM correlation.

**D)** The LV and HV quadrants of demeaned dissimilarity matrices. Left: demeaned ERP RDM quadrants, as shown in panel A. Right, demeaned contrast statistics RDM quadrants, as shown in panel B.

**E)** Demeaned contrast statistics color-coded according to illumination angle (ignoring texture category membership). Each scatter point represents one image. Illumination angles are illustrated again in the cartoon inset. For color-coding, we collapsed over L1 and L5 as well as L2 and L4, which have the same angle but are positioned on opposite sides.

**F)** The LV and HV quadrants of the demeaned RDMs shown in panel D, but now sorted according to illumination angle, ignoring texture membership. In the HV quadrant only, an effect of illumination is present, as evident by clustering of dissimilarities by illumination angle.
As expected, dissimilarities between images in LV stimuli have completely disappeared in the demeaned ERP RDMs (Figure 3.7D, displaying only the HV-HC and LV-LV quadrants). All variance observed for the LV stimuli thus reflects differences between textures; evoked activity within LV textures is highly invariant. For HV stimuli however, the dissimilarities within textures remain after demeaning, suggesting that HV images are variable within category. Moreover, we observe a ‘plaid-like’ pattern in the RDM, which suggests that dissimilarities of individual HV images do not fluctuate randomly, but are present in a regular manner. What manipulation is driving these dissimilarities? If we consider the clustering of images based on demeaned contrast statistics (Figure 3.7E), we see that for HV stimuli, the variance from the mean is caused by changes in illumination direction: illumination changes ‘move’ stimuli to another location the contrast statistics space in a consistent manner. To demonstrate this effect in the evoked activity, we resorted all images in the RDMs based on illumination direction instead of texture category: in the resulting RDM, differences between ERPs cluster with illumination changes (Figure 3.7F), confirming that dissimilarities between high-variant stimuli result mostly from illumination differences.

These results again confirm that differences between individual images in ERP responses are correlated with differences in contrast statistics. Importantly, they show that differences between HV textures occur for other reasons than differences between LV textures. For HV images, manipulations of illumination angle drive the structure of the RDM: instead of clustering by category (evidenced by within-texture similarity and between-texture dissimilarity), images are selectively dissimilar for one illumination angle compared to another. For LV images, the pattern of results is different: stimuli do cluster by category, such that all images of a given texture are ‘similarly dissimilar’ from other textures (or similar, if the mean of the other images is very nearby in 'contrast statistics space', Figure 3.1C).

Overall, this suggests that the variance between evoked responses correlates with variance in contrast statistics resulting from variations in recording conditions, specifically illumination differences. Textures that vary little in contrast statistics appear to form a more 'invariant' representation in terms of neural activity.

**Image manipulations: rotation vs. illumination effects in behavior**

The results of the EEG experiment suggest that the high variance of HV texture images is related to higher sensitivity of these textures to changes in illumination direction: the large differences that remain within texture categories after subtracting differences between texture categories are driven by differences in illumination angle. In comparison, we observe very little effect of rotation, the other view manipulation. Given this result, it is likely that effects of illumination direction, rather than image rotations also drive the main behavioral effect of increased error rates on HV texture categories.
To address this question, we post-hoc sorted the data from Experiment 1 based on whether the two presented images differed in a) rotation only, b) illumination only, or c) both rotation and illumination, and separately computed the accuracies for each of the different conditions (same LV, same HV, different LV, different HV). Because the pairing of individual images was randomized over trials (see Materials and Methods), there were unequal amounts of manipulation differences for each subject and condition. To increase the number of trials per condition and to be able to compare across conditions, we collapsed over same angles from different sides, as in the EEG RDM analysis (see Figure 3.7E). As a result, we obtained four different trial-types: 1) same illumination, same rotation (SI, SR), 2) same illumination, different rotation (SI, DR), 3) different illumination, same rotation (DI, SR) or 4) different illumination, different rotation (DI, DR).

Across all trial-types, accuracy was lower for HV than for LV stimuli (Figure 3.8A; main effect of variance on both same-category and different-category trials, all F(25) > 27.9, p < 0.0001). However, as predicted, on same-category trials most errors were made when illumination was different compared to when rotation was different and illumination the same (main effect of illumination, F(25) > 262.9, p < 0.0001). Importantly, this effect was much larger for HV textures than for LV textures (significant interaction between variance and illumination, F(25) = 162.1, p < 0.0001). Thus, the influence of illumination differs for HV vs. LV textures: LV textures are more invariant to this viewing variation. Interestingly, the effect is reversed for different-category trials (more errors for same illumination trials; main effect of illumination, F(25) = 20.6, p < 0.0001) suggesting that in this case, illumination changes 'help' to distinguish different texture categories from another.

Figure 3.8 Post-hoc analysis of the effect of viewing manipulations on the results of Experiment 1. A) Accuracies per condition (HV/LV) and type of comparison (same/different) were computed separately for trials in which the two images were photographed under same illumination and same rotation (SI, SR; green dashed line), same illumination and different rotation (SI, DR; green solid line), different illumination and same rotation (DI, SR; purple dashed line), or different illumination and different rotation (DI, DR; purple solid line). Participants always made more errors on HV texture categories than LV texture categories. However, most errors were made for same-category, different illumination HV trials, not different rotation trials. Error bars indicate S.E.M. B) Effect of illumination angle (L1-L5, see Figure 3.1B) on same-illumination (SI, diagonal values) and different-illumination (DI, off-diagonal values) trials, collapsed over rotation, for each condition and type of comparison. Most errors on same-category HV trials are made for images with a large difference in illumination angle (i.e., L3 vs. L1/L5); most errors on different-category HV trials are made for angle L3. Color codes for angles correspond to Figure 3.7E.
Most importantly, these results support the conclusion that the extent to which a given texture category is sensitive to illumination changes can be predicted from contrast statistics CE and SC. To demonstrate the effects of illumination changes more clearly, we sorted the DI accuracies based on the exact illumination angle (L1-L5, see Figure 3.1B and Fig. 3.7E) of each of the two stimuli on a given trial. Given the small effect of rotation, we now collapsed over same rotation and different rotation trial-types, i.e. SR and DR trials. The results of this analysis are displayed as confusion matrixes in Figure 3.8B (diagonals represent the SI trials). On same-category HV trials (lower right matrix), most errors are made when the change in illumination angle was large (e.g., a pairing of L3 and L1/L5). For same-category LV trials (lower left matrix), this effect is much weaker, indicating that illumination changes have less effect on categorization of LV textures. Interestingly, the confusion matrix of the different category HV trials (upper right panel) shows that most errors were made for angle L3, suggesting that in general, images from different texture categories become more similar when the light is coming right from above; likely, this is due to higher saturation of the image (overexposure) under this illumination angle. This effect is however again absent for the different-category LV trials (upper left panel), suggesting that saturation effects occur less often for LV images.

**Luminance statistics**

The behavioral and EEG results suggest that the contrast statistics of the texture categories are diagnostic of perceived variance under changes in illumination. The behavioral results further suggest that the amount of illumination change is directly related to the perceived similarity on same-texture HV trials, and that the specific illumination angle used on different-texture HV trials may also influence the observed similarity (Figure 3.8B). Does this mean that a simple description of differences in luminance between images (i.e. brightness), rather than contrast statistics, explain this data equally well? To test this, we computed the mean luminance (LUM, see Materials and Methods) of each image and tested whether differences in luminance were correlated with differences in behavioral categorization and EEG responses.

First of all, LUM values of individual images were indeed correlated with CE ($\rho = -0.69, p < 0.001$), and somewhat lower but significantly correlated with SC ($\rho = -0.38, p < 0.001$). However, if we split the texture categories into high and low variance based on LUM values per category, we do not find the same categories in each condition as in the original division based on contrast statistics; see Figure 3.9A. In fact, about half of the LV categories are high-variant in LUM. Correlations of the variance in LUM and accuracy were either not significant ($\rho = -0.32, p = 0.15$, for the oddity experiment) or lower compared to the correlation with contrast statistics ($\rho = -0.51, p = 0.02$ for the same-different experiment). This suggests that variance in brightness rather than contrast is not an alternative explanation for the finding that observers perceive images from LV texture categories more often as invariant.
Could it be that not the variance in luminance, but the average luminance of a texture category affects the degree to which it is seen as invariant? A majority of LV texture categories indeed have low mean luminance, but this is not the case for all categories, suggesting that HV texture categories are not systematically brighter than LV categories. Correlations of accuracy and mean LUM are again more inconsistent compared to contrast statistics: they are significant in the oddity experiment ($\rho = -0.62$, $p < 0.005$), but not in the same-different experiment ($\rho = -0.35$, $p = 0.12$). As can be seen in Figure 3.9B, the correlation with behavioral accuracy can be explained by partial overlap of the LV/HV categories and the low/high LUM values.

In the EEG data, differences in LUM explained less variance (peak mean $r^2$ across subjects = 0.44, between 0.27-0.72 for individual subjects, Figure 3.10A) than differences in contrast statistics. To compare the model fits directly, we used Akaike’s information criterion (see Materials and Methods) to compute AIC-values based on the residuals of the regression analyses. In this analysis, a lower AIC-value indicates a better fit to the data, or ‘more information’.

The mean AIC-values obtained at each time-point of the ERP are shown in Figure 3.10B, where it can be seen that in the early time interval where contrast statistics correlate most strongly with the ERP (~140-180 ms), the LUM regression model has a higher AIC-value and thus worse predictive power than the contrast statistics model. Note that this is despite the fact that the contrast statistics consist of two parameters and LUM only one, which benefits the AIC-value of the latter. For most subjects, the AIC-value based on contrast statistics is consistently lower than the AIC-value from regression on LUM, also later in time (Figure 3.10C). Between the models, individual maximal $r^2$ values and corresponding time-points were highly correlated ($\rho = 0.87$, $p < 0.0001$ and $\rho = 0.67$, $p < 0.005$, respectively), suggesting that subjects with high explained variance for contrast statistics also had high $r^2$-
values for LUM and that these effects occurred around the same time. Interestingly however, the difference in AIC-value is largest for subjects with high maximal $r^2$-values (Figure 3.10D), suggesting that increased explained variance is associated with a larger difference in goodness of fit between the LUM and contrast statistics.

These analyses suggest that despite the high correlations between contrast statistics and luminance, contrast statistics better predict perceived invariance as well as differences in evoked activity. This is not surprising: from physiology, it is known that neurons in LGN effectively band-pass filter contrast values from the visual input (De Valois and De Valois, 1990). Indeed, repeated band-pass filtering of visual information seems a fundamental property of visual cortex, resulting in increasingly invariant representations (Bouvrie et al., 2009). From this perspective, contrast information is itself more invariant than luminance, and therefore more reliable. Our results however suggest that this hierarchical increase in invariance obtained by contrast filtering is not equal for all types of texture. After contrast filtering, each texture presumably becomes more invariant, but some textures become more invariant than others. Together, our results show that this difference is evident in both behavioral categorization and evoked neural activity of texture images.

*Figure 3.10* Regression analyses based on LUM values on single-subject ERP data and comparison with contrast statistics. **A** Explained variance at channel Oz for each individual subject (colored lines) and averaged across subjects (black line) from a regression model with LUM values. **B** AIC-values of the regression results based on LUM compared to contrast statistics (CE/SC); a low AIC-value indicates a better model fit. **C** Single-subject differences in AIC-value between LUM and contrast statistics over time. **D** Subject-specific differences in AIC-values at the time-point where the difference between the two models in mean AIC-values is largest (148 ms), sorted based on maximal $r^2$ values of the contrast statistics model. For subjects with higher $r^2$ values, the difference in AIC-values becomes somewhat larger, suggesting that high explained variance on contrast statistics is not coupled with increased fit of both models simultaneously, but rather with a better fit of contrast statistics compared to LUM values.
Discussion

In a large database of natural textures, we selected images with low-level contrast statistics that were either constant or variable under changes in illumination angle and orientation. In both EEG and behavior, we showed that textures with little variation in low-level contrast statistics were perceived as more invariant (Experiment 1 and 2) and led to more invariant representations at the neural level (Experiment 3). High variance in contrast statistics increased the probability that subjects judged images of the same texture as different categories, specifically if the images differed in illumination direction. Accordingly, high-variant but not low-variant textures gave rise to neural evoked responses that are modulated by illumination direction.

As indicated by higher accuracy on same-texture comparisons in the behavioral experiments, textures with low variance in contrast statistics remained more perceptually similar under different illumination (and rotation) conditions. Consistently, for low-variant textures, we observed 'clustering by texture' of dissimilarities in single-image ERPs, whereas there was clustering by illumination direction for high-variant textures. These results suggest that distance between different textures in terms of contrast statistics is more reliable for low-variant textures than for high-variant textures. This is not surprising if one examines the clustering of low-variant vs. high-variant stimuli in contrast statistics space (Figure 3.1C): as a natural consequence of the lower variance within LV textures, the differences between texture categories become more similar for images of LV textures. This distance information may be useful to an observer in order to estimate whether two stimuli are from the same or from a different texture.

This work extends recent findings that statistical variations in low-level information are important for understanding generalization over single images (Karklin and Lewicki, 2009). In addition, it has been demonstrated that behavioral categorization accuracy can be predicted using a computational model of visual processing: a neural network consisting of local filters that were first allowed to adapt to the statistics of the natural environment could accurately predict behavioral performance on an object categorization task (Serre et al., 2007). Compared to the latter study, however, we applied no training or tuning of a network on a separate set of stimuli such that statistical regularities were implicitly encoded: here, perceived texture similarity was inferred directly from explicitly modeled contrast responses.

In addition to behavioral categorization, we also tested the contribution of our two contrast parameters to evoked neural responses using EEG. It is well known that early ERP components are affected by low-level properties of (simple) visual input (Luck, 2005). Our finding that contrast energy of natural textures predicts single-image ERP amplitude around 140-180 ms is also consistent with previous reports of an early time-frame in which stimulus-related differences drive evoked responses, e.g. between face stimuli (Philiastides and Sajda, 2006; van Rijsbergen and Schyns, 2009). These authors used classification techniques on single-trial ERPs to show that
at later time intervals, inter-image differences reflect either a more refined representation of task-relevant information (van Rijsbergen and Schyns, 2009) or the actual decision made by the subject (Philiastides and Sajda, 2006). This suggests that over the course of the ERP, the visual representation is transformed from simple low-level properties towards a representation of information that is task-relevant. In this light, it is surprising that our image parameter SC is still correlated with late ERP activity - around 300 ms - and that it explains additional variance compared to contrast energy alone in this time interval.

One possible explanation of this apparent discrepancy is that the spatial coherence parameter is itself correlated with more refined or relevant features of natural images, essentially constituting a summary statistic of visual input that can be used for rapid decision-making (Oliva and Torralba, 2006). Another interesting hypothesis is that this parameter is predictive of the availability of diagnostic information, reflecting higher 'quality in stimulus information' (Gaspar et al., 2011) or less noise in the stimulus (Bankó et al., 2011; Rousselet et al., 2011), which may influence the accumulation of information for decision-making (Philiastides et al., 2006). Since our two stimulus conditions (HV/LV) were defined based on variance in both contrast energy and spatial coherence, we cannot test which one is more strongly correlated with behavioral accuracy. Also, our work is substantially different from these previous reports in that our experiments did not require formation of a high-level representation (e.g., recognition of a face/car), but merely a same-different judgment, essentially constituting a low-level task.

Another difference between our results and those reported in the face processing literature (see e.g. Rousselet et al., 2008) is the localization of our effects. Maximal sensitivity of evoked activity to faces and objects is found at lateral-occipital and parietal electrodes (PO), whereas our correlations, obtained with texture images, are clustered around occipital electrode Oz. This is consistent with the proposal that textural information is processed in early visual areas such as V2 (Kastner et al., 2000; Scholte et al., 2008; Freeman and Simoncelli, 2011).

In this paper, we specifically aimed to test whether invariance, in addition to a goal of visual processing, could be defined as a property of real-world visual features (in this case, textures). In the first scenario, one would expect the representation of the visual input to change over time to (gradually) become more invariant. Our behavioral results however indicate that variance in low-level properties of natural textures can already predict the perceived invariance by human observers under specific viewing manipulations.

Moreover, it demonstrates that there are interesting differences between natural textures in terms of this invariance: some textures appear to be surprisingly invariant. It has been argued that, in evolution, mechanisms have evolved for detecting stable features in visual input because they are important for object recognition (Chen et al., 2003). In light of the present results, a biologically realistic instantiation of such a stable feature could be 'a texture patch whose contrast
statistics do not change under viewing variations'. This natural invariance is rooted in physical properties of natural images, but is present at the level of image statistics (stochastic invariance). Such invariance may play an important role in stochastic approaches to computer vision, such as the successful bag-of-words approach (BoW; Fei-Fei et al., 2007; Jégou et al., 2011). For example, a visual scene patch with invariant contrast statistics may be a more reliable 'word' for categorization in a BoW model for scene recognition (Gavves et al., 2011).

Our results suggest that such stochastic invariances are not only reflected in occipital ERPs recorded at the scalp, but that the human visual system may actively exploit them. In our data, low-variant textures did not only give rise to more reliable differences between texture categories in evoked responses, but were also associated with more reliable judgments about similarity between different textures – i.e., with behavioral outcome. The link between contrast statistics and categorization accuracy leads to the hypothesis that in more naturalistic tasks such as object detection or natural scene processing, image elements that are stochastically invariant, i.e. reliable, may weigh more heavily in perceptual decision-making than variable, unreliable elements.

In sum, the present results show that low-level contrast statistics correlate with variance of natural texture images in terms of evoked responses, as well as perceived perceptual similarity. They suggest that textures with little variance in contrast statistics give rise to more invariant neural representations. Simply put, invariance in simple, physical contrast information may lead to a more invariant perceptual representation.

This makes us wonder about visual invariance as a general real-world property: how much of it can be derived from scene statistics? Are there other low-level visual features that differ in their degree of invariance? Next to studying top-down, cognitive invariance or transformations performed by the visual system to achieve invariance of visual input, exploring to what extent 'natural invariances' exist and whether they play a role in visual processing may provide an exciting new avenue in the study of natural image perception.

Acknowledgements
We thank Marissa Bainathsah, Vincent Barneveld, Timmo Heijmans and Kelly Wols for behavioral data collection of the same-different experiment. This work is part of the Research Priority Program ‘Brain & Cognition’ at the University of Amsterdam and was supported by an Advanced Investigator grant from the European Research Council.