Modulation of early ERPs by accurate categorization of objects in scenes

Andrea De Cesarei

Department of Psychology, University of Bologna, Bologna, Italy ✉

Ilaria A. Peverato

Department of Psychology, University of Bologna, Bologna, Italy ✉

Serena Mastria

Department of Psychology, University of Bologna, Bologna, Italy ✉

Maurizio Codispoti

Department of Psychology, University of Bologna, Bologna, Italy ✉

The categorization of objects within natural scenes is carried out in a sequence of stages, which may build on the detection of perceptual regularities in the visual appearance of objects or may represent a more semantic level of categorization. Here, we examined the neural correlates of correct categorization of objects in scenes, using natural scenes which were equalized in color and spectral amplitude, and controlled in terms of spatial coherence. Event-related potentials (ERPs) were used to track the early stages of visual processing. Participants viewed degraded (phase-scrambled) versions of natural scenes and then categorized them as depicting animals or people. At an intermediate scrambling level, a negative-going occipitotemporal ERP modulation by categorization accuracy was observed, beginning approximately 150 ms after stimulus onset; at more degraded levels, no ERP modulation was observed. These results suggest that this early negative-going ERP modulation reflects processing of perceptual evidence which is predictive of later correct categorization, even when low-level differences in color, spectral amplitude, and spatial coherence are balanced or controlled.

Introduction

Despite the visual complexity of the world, the human visual system can usually analyze visual information and convert it into meaningful representations without feelings of overt effort. This remarkable efficiency has attracted a vast amount of research on how the visual system processes visual information.

While classic studies focused on the analysis of simple stimuli such as sinusoidal gratings, more recent studies have focused on the processes which subtend the analysis of complex natural scenes (Felsen & Dan, 2005; Kayser, Körding, & König, 2004). Since the 1970s, several studies have successfully investigated the minimum exposure time for identifying stimulus properties such as identity, categorical belonging, or features such as openness or depth (Busey & Loftus, 1994; Greene & Oliva, 2009; Intraub, 1981; Loftus, 1972; Potter, 1975). More recently, electrophysiological measures such as event-related potentials (ERPs) have complemented these behavioral studies, providing insight into the minimum time at which a differential electrocortical activity for categorized stimuli is observed, prior to or in absence of a behavioral response (Luck, 2005).

Several recent studies investigated the latency for detecting a target such as an animal in a briefly presented (24 ms) scene, and it was found that as soon as 150 ms after stimulus onset, a differential activity for targets compared to nontargets could be observed over occipital areas (Thorpe, Fize, & Marlot, 1996). Subsequent research indicated that this early differential activity was localized in the inferotemporal cortex (Codispoti, Ferrari, Junghöfer, & Schupp, 2006b; Fize et al., 2000) and that response requirements (e.g., go/no go vs. multiple choices) and target category (e.g., animals vs. means of transport) did not account for the observed effects (Antal, Keri, Kovacs, Janka, & Benedek, 2000; Codispoti, Ferrari, De Cesarei, & Cardinale, 2006a; De Cesarei, Codispoti, Schupp, &

Stegagno, 2006). Further insight into the nature of this differential ERP activity comes from a previous study (VanRullen & Thorpe, 2001) which alternated the target/distractor status of the same images in different blocks and compared ERPs depending on the status of the pictures. When a category (e.g., animals) was designated as target, it elicited a differential ERP activity starting at 150 ms after stimulus onset compared to the same category when it was designated as distractor. The observation of an ERP modulation associated with target detection at such low latencies seems to complement the behavioral observation that briefly presented complex scenes can be quickly characterized in terms of basic features (Evans & Treisman, 2005; Greene & Oliva, 2010; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001).

Which information allows for this fast image categorization? It has been suggested that the visual system continuously computes the statistical properties of the visual input, for instance concerning the distribution of amplitudes in the frequency spectrum (Oliva & Torralba, 2001, 2006; Simoncelli & Olshausen, 2001; Torralba & Oliva, 2003). These statistics provide information about the features that are regularly associated with a category, such as sharp contours for artificial compared to natural scenes (Torralba & Oliva, 2003), or about the composition of visual scenes (e.g., in terms of the fragmentation of picture layout). Eventually, the calculation of these statistics aids basic-level categorization and favors scene understanding (Crouzet & Serre, 2011; Ghebreab, Scholte, Lamme, & Smeulders, 2009; Oliva & Schyns, 2000; Oliva & Torralba, 2001, 2006; Ullman, Vidal-Naquet, & Sali, 2002). Recently, it has been shown that the gamma parameter of the Weibull fit to image contrast reflects the fragmentation of the layout of the scene and may serve as a basis for natural-image identification (Geusebroek & Smeulders, 2005; Ghebreab et al., 2009; Yanulevskaya & Geusebroek, 2009). In particular, the gamma parameter of the Weibull fit to image contrast has been dubbed the shape parameter (Geusebroek & Smeulders, 2005) or spatial coherence (Groen, Ghebreab, Lamme, & Scholte, 2012), as it describes the visual clutter that is present in a scene.

Scene statistics have been shown to be biologically relevant, as they approximate computations which are carried out during visual processing and modulate electrocortical responses (Scholte, Ghebreab, Waldorp, Smeulders, & Lamme, 2009). More specifically, several studies have indicated that the spatial coherence of a scene modulates electroencephalogram (EEG) activity (Groen, et al., 2012; Groen, Ghebreab, Prins, Lamme, & Scholte, 2013; Scholte et al., 2009). These studies indicate that most of the variance in the amplitude of single-trial event-related potentials in an early time interval (around 113 ms) is explained by spatial coherence (Groen et al., 2013; Scholte et al., 2009) or by the combined effects of spatial coherence and contrast energy (Ghebreab et al., 2009; Groen et al., 2012). Moreover, it has been shown that the presence of diagnostic colors (e.g., colors that are typically associated with a class of scenes, such as blue for the sea) can modulate early ERPs related to scene identification (Goffaux et al., 2005). Similarly, a vast amount of data (e.g., De Cesarei, Mastria, & Codispoti, 2013; Hansen, Jacques, Johnson, & Ellemberg, 2011; Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Rousselet & Pernet, 2011; VanRullen, 2011) indicates that differences in image statistics may modulate early ERPs during scene or object categorization.

The research problem

The present research aimed to extend the results of previous studies by investigating whether early ERPs are modulated by categorization accuracy. If the early ERP modulation that has been previously observed (Thorpe et al., 1996) reflects the attainment of correct categorization of objects in scenes, then a comparable modulation should be observed not only when comparing targets to distractors but also when comparing correctly categorized to misidentified trials. Due to the remarkably high accuracy that is achieved using intact images, this type of analysis was not possible in several previous studies (Antal et al., 2000; Codispoti et al., 2006b; De Cesarei et al., 2006; Thorpe et al., 1996; VanRullen & Thorpe, 2001). An interesting preliminary result in this direction comes from a previous study investigating a go/no-go categorization task (Rousselet, Thorpe, & Fabre-Thorpe, 2004), in which categorization accuracy was reduced by presenting one, two, or four scenes at the same time. In that study, similar early ERP modulation was observed when targets were correctly identified and when a distractor was incorrectly identified as a target. The present study was conceived in order to extend these results to a forced-choice paradigm, in which perceptual cues related to color and to the spectral amplitude were controlled.

Here the experimental manipulations aimed to reduce perceptual differences between images and to modulate scene identifiability. To this end, participants looked at degraded pictures of natural scenes and were asked to perform a forced-choice categorization between two target categories (animals and people). Scenes were manipulated in three aspects: They were equated in terms of color and Fourier spectrum amplitude, and their spectral phase was randomized. Perceptual differences in color and global spectral amplitude were eliminated, so that scene processing was examined in the absence of these perceptual cues. Additionally, categorization accuracy was kept low and

modulated by presenting scenes in which the phase of the frequency spectrum was randomized to varying extents. Phase scrambling is a procedure which is frequently used to decrease the identifiability of a stimulus without altering its spectral amplitude properties (Arsenault, Yoonessi, & Baker, 2011; Bieniek, Pernet, & Rousselet, 2012; Joubert et al., 2009; VanRullen, 2011). The manipulation of the phase of the frequency spectrum changes the locations in the picture where contrast changes happen; as a consequence, phase scrambling breaks the local structure of a scene. In previous studies, the distortion or randomization of the phase spectrum dampened categorization accuracy for both scenes (Gaspar & Rousselet, 2009; Joubert et al., 2009) and faces (Bieniek et al., 2012; Rousselet, Pernet, Bennett, & Sekuler, 2008). However, other studies have also emphasized an effect of spectral amplitude or an interaction of amplitude and phase in the modulation of behavioral and electrocortical correlates of categorization (Gaspar & Rousselet, 2009; Loschky & Larson, 2008; Wichmann, Braun, & Gegenfurtner, 2006).

Depending on the role of categorization and perceptual cues in the modulation of early ERPs, one of two scenarios may be expected. In the first scenario, modulation of early ERPs reflects the semantic categorization of objects in scenes. If this prediction holds true, then ERP modulation should be observed for successfully categorized compared to misidentified trials, despite the fact that perceptual differences in color and spectral amplitude are controlled. Alternatively, it is possible that the previously observed early ERP difference did not reflect semantic categorization per se but was elicited by a perceptual cue that is eliminated in the controlled conditions adopted here. This second scenario therefore predicts that no ERP modulation will be observed, as important perceptual cues (color and spectral amplitude) are controlled.

An additional problem that occurs when visual scenes are degraded is that some scenes are more affected by degradation than others, in terms of identifiability. Natural scenes differ in several aspects that are captured by image statistics; here, we equated two of these aspects, namely the color and the amplitude of the Fourier spectrum. However, this does not exclude the possibility that local differences still exist in equalized pictures, and it is possible that scenes with a more coherent local structure are more resistant to phase scrambling. To understand in which pictures objects could be correctly and incorrectly categorized, we measured the spatial coherence of all intact stimuli and analyzed results accordingly.

Finally, we manipulated the context in which scenes were analyzed. In the first condition, participants viewed scenes in a randomized order, and no prior information about the upcoming picture was present.

Alternatively, pictures could be seen in a sequence in which the most degraded version was presented first and picture phases were gradually unscrambled. It may be expected that prior coarse information about picture content may modulate successive processing. In the neural model of visual perception which was put forward by Bar (2004), coarse information about picture content is quickly analyzed and projected to the prefrontal cortex, where hypotheses regarding the content of visual input are generated; these perceptual hypotheses are then back-projected to visual areas and constrain further processing of fine-grained information. While this model describes how coarse and fine information interact during the processing of a single scene, it may be asked whether a similar facilitation can be observed on a trial-by-trial basis, when stimuli are progressively revealed. It may be expected that ERP modulation will be facilitated (earlier latency or more pronounced amplitude) in the sequential compared to the randomized condition.

Methods

Participants

A total of 20 participants (14 women) took part in the study for course credit. Age ranged from 19 to 29 years ($M = 21.6$, $SD = 3.1$). All participants had normal or corrected-to-normal vision, and none of them reported current or past neurological or psychopathological problems. The participants had no previous experience with the materials used in this experiment. The experimental protocol conformed to the Declaration of Helsinki and was approved by the Ethical Committee of the Department of Psychology at the University of Bologna.

Stimuli and equipment

A total of 240 pictures were selected for the present study from public-domain images available on the Internet and from the International Affective Picture System (IAPS) database (see Figure 1 for examples of the pictures used). Half of the pictures represented animals and half depicted people. The present data were part of a larger project on emotional response, and pictures of people could be positively, neutrally, or negatively valenced. Pictures subtended a visual angle of 28° (horizontal) by 21° (vertical). The resolution of the original pictures was 800 × 600 or higher, and all stimuli were cropped to a 4:3 ratio and scaled to 800 × 600 pixel size.
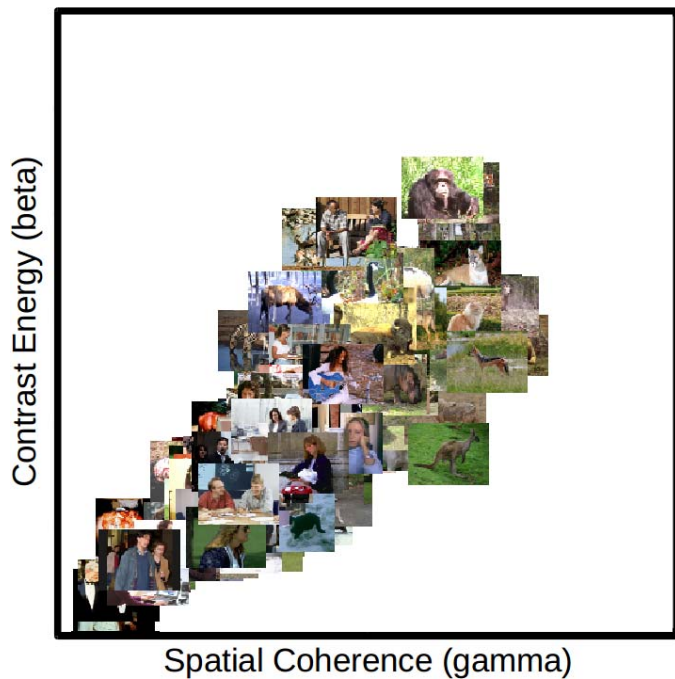
Figure 1. Scatter plot of the contrast energy (*y*) and spatial coherence (*x*) in the present picture set. Each original image is positioned at its respective values of contrast energy and spatial coherence. These parameters were calculated on the original, unequalized versions of the pictures.

For each original picture, spatial coherence (SC) was calculated using the algorithm suggested by Yanulevskaya and Geusebroek (2009). This procedure estimates the beta (contrast energy) and gamma (spatial coherence) parameters of the Weibull fit to the distribution of contrast. The results of this analysis are represented in Figure 1.

The pictures were modified in the following way: All pictures were converted to grayscale and equated to the same frequency spectra, brightness, and contrast using a MATLAB-based toolbox (Willenbockel et al., 2010). Then four phase-scrambled versions of each picture were created using a weighted mean phase algorithm (Dakin, Hess, Ledgeway, & Achtman, 2002). This procedure consists of three steps. First, the power and phase of the image spectrum are calculated. Then the phase spectrum of the original image is combined with a random phase, according to a mixing factor ranging from 100% (only the random-phase information is used) to 0% (only the original-phase information is used). In the third and final step, the resulting phase spectrum is recombined with the original spectral power, and a picture is obtained which retains the spectral power of the original image but with different phase information. Based on pilot data from three participants who did not take part in the final experiment, it was decided that phase scrambling parameters of 80%, 65%, 55%, and 0% would be used.

Notably, with these parameters the fourth level of phase scrambling is identical to the original picture and will henceforth be referred to as intact pictures.

Procedure

The block and trial procedure is presented in Figure 2. The experiment was divided into two blocks, the order of which was counterbalanced across participants. In one block (mixed block), the order of pictures was pseudorandomized. In neighboring trials of the mixed block, a different picture would always be presented. In the other block (sequential block), all four versions of the same picture would be presented in a row, beginning with the most degraded one and progressively revealing picture content. Pictures which were presented in one block were not presented in the other one, and all pictures were equally often assigned to either block across all participants. In each block, all four versions of the pictures were presented. The two blocks were otherwise identical.

Each trial began with the presentation of a fixation cross, which remained onscreen for 500 ms (Figure 2). Then a picture was presented that remained onscreen for 1 s. After picture offset a question mark appeared, signaling that a response was required. Participants were asked to decide whether the picture they had just seen represented a person or an animal, and indicate their choice by pressing one of two keys (Z or M) on the computer keyboard. The 1-s exposure time and the assessment of responses following picture offset were chosen so that ERPs related to the offset of the visual stimulus or related to the preparation of motor response would not contaminate ERPs in the time interval of interest. Participants were required to respond to all trials, and the association between the response key and the category was counterbalanced across participants. After an intertrial interval of 2 s, the next trial began.

Prior to the first block, eight practice trials in the mixed or in the sequential condition were presented, in order to let participants familiarize themselves with the categorization task. Halfway though each block, and between the two blocks, participants were allowed to have a short break. Data from the practice trials were not analyzed.

EEG recording and processing

EEG was recorded at a sampling rate of 256 Hz from 256 active sites using an ActiveTwo Biosemi system. An additional sensor was placed below the participant's left eye, to allow for detection of blinks and eye movements. The EEG was referenced to an additional
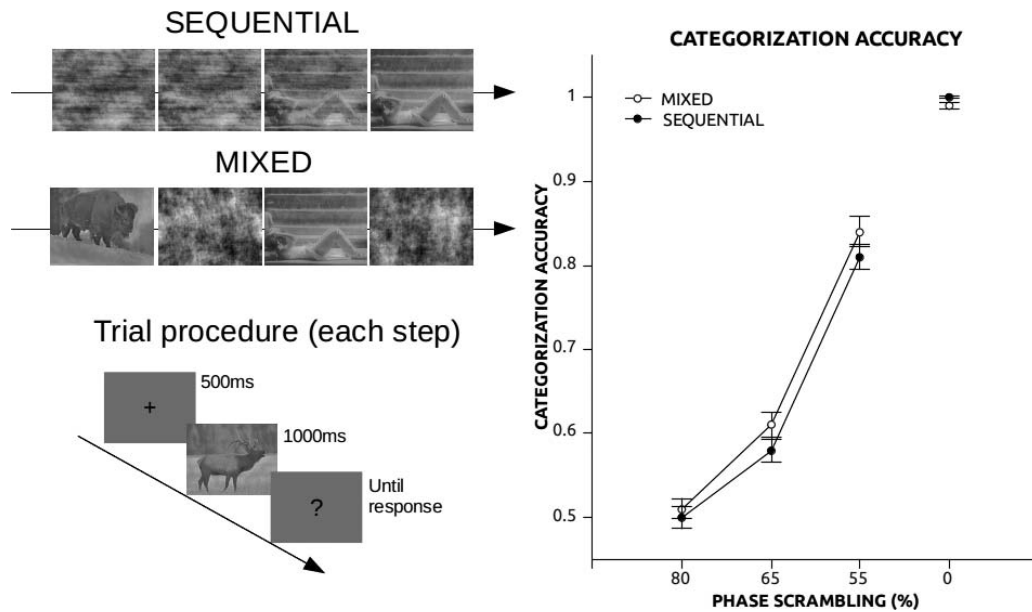
Figure 2. Procedure and accuracy. The top left panel represents the sequence of picture presentation in the mixed and sequential conditions. The bottom left panel represents the sequence of events for each picture in the sequence. The panel on the right represents categorization accuracy, for each of the three degradation conditions and for intact pictures. Error bars represent the standard error of the mean.

reference electrode located near Cz during recording. A hardware fifth-order low-pass filter with a -3-dB attenuation factor at 50 Hz was applied online. Off-line analysis was performed using Emegs (Peyk, De Cesarei, & Junghöfer, 2011). EEG data were initially filtered (0.1 Hz high-pass and 40 Hz low-pass), and eye movements were corrected by means of an automated regressive method (Schlögl et al., 2007). Trials and sensors containing artifactual data were detected through a statistical procedure specifically developed for dense-array EEG (Junghöfer, Elbert, Tucker, & Rockstroh, 2000). Trials containing a high number of neighboring bad sensors were discarded; for the rest of trials, sensors containing artifactual data were replaced by interpolating the nearest good sensors. Finally, data were re-referenced to the average of all sensors, and a baseline correction based on the 100 ms prior to stimulus onset was performed. For the image-based analysis, data were averaged over participants, whereas for the participant-based analysis, data were averaged across pictures. The distribution of good trials per condition was as follows: 80% phase scrambling, correct $M = 107.2$, $SD = 10.6$, incorrect $M = 103.45$, $SD = 12.9$; 65% phase scrambling, correct $M = 129.2$, $SD = 13.9$, incorrect $M = 87.65$, $SD = 12.21$; 55% phase scrambling, correct $M = 179.95$, $SD = 15.98$, incorrect $M = 38.05$, $SD = 14.91$; 0% phase scrambling, correct $M = 208.95$, $SD = 9.26$, incorrect $M = 1.83$, $SD = 2.29$.

In the reported analysis, ERPs in the three degraded conditions are scored and analyzed separately from those in the intact condition. This was decided based on

preliminary analyses, which indicated differences in ERP topography, number of accurate trials per condition, and sensitivity to picture content. Concerning ERP topography, a preliminary inspection of the ERP waveforms in the four degradation conditions indicated that ERPs in the intact condition differed from the other conditions in latency and amplitude and due to the presence of a pronounced N1 which was absent in the other conditions. Additionally, since for the present study it was critical to have enough trials in which an accurate categorization is not attained, the intact condition could not be included in the Accuracy $\times$ Degradation statistical design, as not enough incorrect trials were retained in the intact condition.

To collapse multidimensional ERP data for statistical analysis, the region and time interval of interest were selected based on previous studies.[1] Sensor selection was based on two previous studies that examined rapid categorization using averaged referenced data and reported the labels of the sensors used for analysis (Codispoti et al., 2006b; Rousselet et al., 2004). Based on those labels, we selected the sensors closest to T5, P5, PO5, PO7, O1, POz, Oz, T6, P6, PO6, PO8, and O2. All sensors in this scalp area were included in the analysis, for a total of 21 electrodes. From this sensor group, ERPs in the degraded conditions were scored and analyzed in the time interval 150–300 ms for consistency with previous studies (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Rousselet et al., 2004; Thorpe et al., 1996; VanRullen & Thorpe, 2001). In the intact condition,
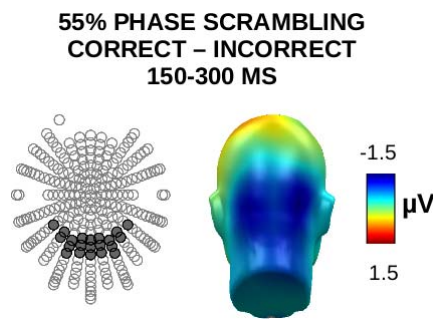
Figure 3. ERP modulation by categorization accuracy in the time interval 150–300 ms, in the 55% phase-scrambling condition. On the left, an overview of the 257-sensor net is presented (top view, looking upwards), with sensors used in the analysis highlighted in gray.

ERPs were scored as the average in the time interval 200–350 ms in the same scalp area. The time interval 200–350 ms was chosen based on visual inspection of ERP data; more specifically, in the intact condition, ERPs included a pronounced N1 which was absent in the other conditions, and the overall ERP waveform appeared delayed in the intact compared to the scrambled conditions. The sensors used in the analysis are reported in Figure 3, and the time intervals of interest are highlighted in Figures 4 and 5.

Data analysis

General strategy

The analysis was aimed primarily at investigating the effects of phase scrambling (80%, 65%, 55%, 0%) and procedure (mixed vs. sequential) on categorization and on electrocortical correlates of picture perception. Addi-
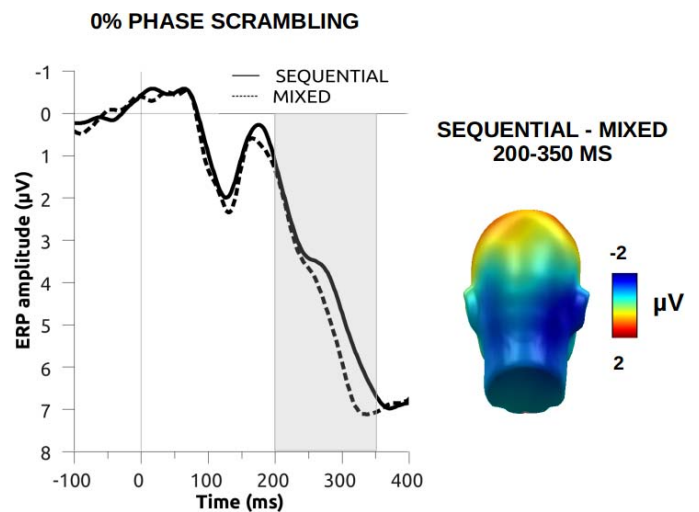

Figure 5. The effects of procedure on ERPs in the intact condition, in the examined occipitotemporal sensor group from 200 to 350 ms. The topography on the right reports a back view of the differential between sequential and mixed ERPs.

tionally, we asked whether differences between images in terms of spatial coherence might account for any behavioral or ERP effects. To this end, SC was calculated for the original images, and images were ranked in two groups according to the SC value (high vs. low). A high value indicates a locally fragmented scene, while a low value indicates a more homogeneous composition.

Behavioral and ERP data were analyzed through repeated-measures ANOVAs with factors of procedure, phase scrambling, and spatial coherence rank. In the analysis of ERP data of the scrambled scenes, an additional factor of accuracy (correct vs. incorrect) was included in the ANOVA design. For intact pictures, a one-way repeated-measures ANOVA was carried out with procedure and SC rank as factors.
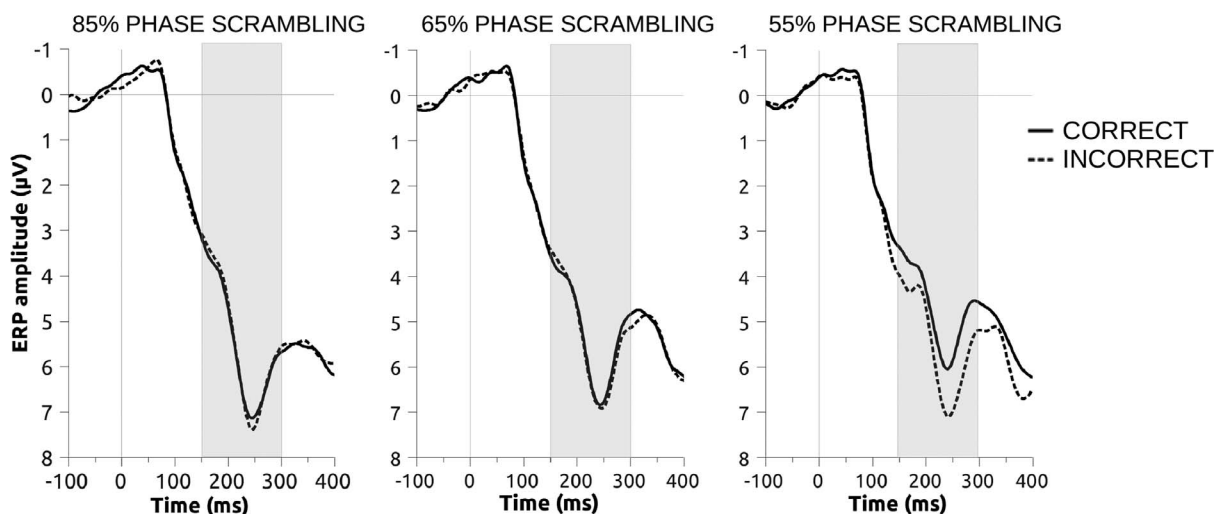

Figure 4. The effects of phase scrambling on early ERPs from the occipitotemporal sensor group analyzed, as modulated by phase scrambling and accuracy.

To deal with sphericity violations that increase the probability of type I error, a Huynh–Feldt correction was applied to the degrees of freedom. In all cases in which a significant main effect of a factor with more than two levels was observed, we proceeded with post hoc tests using a paired-sample t test. For all ANOVA effects, we calculated, and report, the partial eta squared, which reflects the proportion of variance that is accounted for by experimental manipulations.

Behavioral responses

Behavioral data were analyzed for accuracy and response times. Accuracy was averaged across participant, SC rank (high vs. low), phase scrambling (80%, 65%, 55%, 0%), and procedure (mixed vs. sequential). Accuracy in all phase-scrambling conditions was compared to the 0.50 chance level through a t test with $\alpha = 0.05$. Response times to trials which were correctly categorized were averaged across the same factors, using the mean as the index of central tendency and excluding trials that deviated more than ± 2 standard deviations from the mean of each participant.

ERPs

ERPs were first analyzed through an ANOVA with participants as a random factor and accuracy, phase scrambling, and SC rank as within-participant factors. In the remainder of this article, we will refer to this analysis as "participant-based." Additionally, an image-based analysis was carried out with the objective of investigating how categorization accuracy interacts with the effects of phase scrambling and procedure on early ERPs, while measuring and comparing differences between images in SC. To this end, an additional ANOVA was conducted using images as a random factor, i.e., by comparing ERPs between correct and incorrect categorizations of the exact same images. This analysis will be dubbed "image-based" in the following text. This allows the analysis to account for physical differences between stimuli. This analysis was conducted using accuracy (correct vs. incorrect) and phase scrambling (80%, 65%, 55%) as within-image factors and procedure (mixed vs. sequential) and SC rank (high vs. low) as between-images factors.

Results

Categorization performance: Accuracy and response times

Accuracy results are reported in Figure 2. Accuracy was initially at chance ($M = 0.51$, $SD = 0.06$). As pictures were revealed, accuracy increased and correct categorization was achieved for intact pictures ($M = 1.00$, $SD = 0.01$). The 80% phase scrambling did not differ significantly from chance, $t(19) = 0.861$, $p = 0.40$, while all other conditions significantly differed from chance, with t tests yielding significant results: 65%, $t(19) = 7.423$, $p < 0.001$; 55%, $t(19) = 21.269$, $p < 0.001$; 0%, $t(19) = 244.61$, $p < 0.001$.

A significant effect of phase scrambling was observed, $F(3, 57) = 734.91$, $p < 0.001$, $\eta^2_p = 0.98$, indicating that categorization accuracy increased as scenes were revealed. Post hoc tests between neighboring phase-scrambling levels indicated significant differences between all levels: 80% vs. 65%, $t(19) = -8.55$, $p < 0.001$; 65% vs. 55%, $t(19) = -33.341$, $p < 0.001$; 55% vs. 0%, $t(19) = -11.604$, $p < 0.001$.

A significant interaction of procedure and phase scrambling was observed, $F(1, 19) = 2.843$, $p = 0.046$, $\eta^2_p = 0.13$. Following this interaction, the effects of procedure were examined at each phase-scrambling level. Significant effects of procedure were observed for 65% and 55% phase scrambling—$F(1, 19) = 5.79$, $p = 0.026$, $\eta^2_p = 0.23$, and $F(1, 19) = 4.69$, $p = 0.043$, $\eta^2_p = 0.20$, respectively—indicating better performance in the mixed compared to the sequential condition. No effects of procedure were observed in the 80% and 0% phase-scrambling conditions: 80%, $F(1, 19) = 2.171$, $p = 0.157$, $\eta^2_p = .103$; 0%, $F(1, 19) = 1.021$, $p = 0.325$, $\eta^2_p = 0.051$. Overall, a significant effect of procedure was observed, $F(1, 19) = 5.43$, $p = 0.031$, $\eta^2_p = 0.22$, indicating better accuracy in the mixed compared to the sequential procedure.

Finally, a significant interaction of SC rank and phase scrambling was observed, $F(3, 57) = 14.40$, $p < 0.001$, $\eta^2_p = 0.43$. Examining the effects of SC rank at each phase-scrambling level, significant effects were observed for 65% and 55% phase scrambling—$F(1, 19) = 7.61$, $p = 0.012$, $\eta^2_p = 0.29$, and $F(1, 19) = 56.82$, $p < 0.001$, $\eta^2_p = 0.75$, respectively—with higher accuracy for images which were low in spatial coherence compared to high. SC rank failed to modulate accuracy in the 80% and 0% phase-scrambling conditions: 80%, $F(1, 19) = 1.370$, $p = 0.256$, $\eta^2_p = 0.067$; 0%, $F(1, 19) = 0.003$, $p = 0.959$, $\eta^2_p < 0.001$. An overall significant main effect of SC rank was observed, $F(1, 19) = 10.33$, $p = 0.005$, $\eta^2_p = 0.35$, indicating better categorization for stimuli which were low in spatial coherence as compared to high.

Response times were analyzed with the same ANOVA design as accuracy data, and are reported in Table 1. A highly significant effect of phase scrambling was observed, $F(3, 57) = 18.45$, $p < 0.001$, $\eta^2_p = 0.49$, indicating faster responses as phase scrambling was reduced. Post hoc tests indicated slower responses in the 80% phase-scrambling condition compared to all other conditions, $ts(19) > 4.488$, $ps < 0.001$, and in the

| Procedure | SC rank | Phase scrambling | | | | |
		80%	65%	55%	0%	Total
Mixed	Low	674.69 (235.65)	602.47 (164.01)	542.82 (155.51)	561.16 (141.08)	595.29 (174.0625)
	High	697.10 (238.28)	635.04 (180.96)	588.62 (152.92)	586.04 (164.52)	626.70 (184.17)
Sequential	Low	732.24 (194.64)	617.74 (185.08)	617.04 (191.82)	603.03 (177.54)	642.51 (187.27)
	High	743.62 (205.80)	651.65 (213.10)	647.93 (213.71)	632.37 (192.63)	668.89 (206.31)
	Total	711.91 (218.59)	626.73 (185.79)	599.10 (178.49)	595.65 (168.94)	633.35 (187.95)

Table 1. Response times (ms) in each of the conditions defined by procedure, phase scrambling, and spatial coherence (SC) rank. Values in parentheses represent standard deviations from the mean.

65% phase-scrambling condition compared to the 55% and 0% conditions, $t(19) = 2.325$, $p = 0.031$, and $t(19) = 2.330$, $p = 0.031$, respectively. No differences were observed between the 55% and 0% conditions, $p = 0.84$. Additionally, a significant main effect of SC rank was observed, $F(1, 19) = 16.73$, $p = 0.001$, $\eta^2_p = 0.47$, indicating faster responses to scenes which were low compared to high in spatial coherence.

In summary, scrambling the phase of the spatial frequency spectrum dampened performance in the categorization of objects in scenes; this effect was reflected both by lower accuracy and by slower responses as phase scrambling increased. Moreover, in the intermediate scrambling conditions (65% and 55%), more accurate performance was observed for trials in the mixed compared to the sequential condition, and for scenes which were low compared to high in spatial coherence.

ERPs: Degraded conditions

In the participant-based analysis, a main effect of phase scrambling was observed, $F(2, 38) = 4.948$, $p = 0.012$, $\eta^2_p = 0.207$, which indicated less positive ERPs in the 55% compared to the 65% and the 80% conditions: 80% vs. 55%, $t(19) = 4.227$, $p < 0.001$; 65% vs. 55%, $t(19) = 3.427$, $p = 0.003$. This effect was modulated by an interaction with categorization accuracy, $F(2, 38) = 5.103$, $p = 0.011$, $\eta^2_p = 0.212$. For correct trials, a main effect of phase scrambling indicated less positive ERPs as phase scrambling was reduced, $F(2, 38) = 13.81$, $p < 0.001$, $\eta^2_p = 0.421$, with significant differences between the 55% and all other degraded conditions: 80% vs. 55%, $t(19) = 4.557$, $p < 0.001$; 65% vs. 55%, $t(19) = 3.95$, $p = 0.001$. For incorrectly categorized trials, no effect of phase scrambling was observed, $F(2, 38) = 0.252$, $p = 0.778$ $\eta^2_p = 0.013$. Finally, after separate analysis of each phase-scrambling level, a significant effect of accuracy was observed only in the 55% phase-scrambling condition, $t(19) = 3.326$, $p = 0.004$, not for 80% and 65% phase scrambling, $t(19) = -1.433$, $p = 0.168$, and $t(19) = -0.277$, $p = 0.785$, respectively. Finally, a significant effect of SC rank was observed,

$F(1, 19) = 11.149$, $p = 0.003$, $\eta^2_p = 0.37$, indicating less positive ERPs for pictures which were low compared to high in spatial coherence.

The image-based analysis investigated the effects of phase scrambling and categorization accuracy on early ERP modulation, while controlling for the spatial coherence of the original scenes. ERP waveforms in the three degraded conditions are reported in Figure 4. Different from the participant-based analysis, the main effect of phase scrambling was not significant in the image-based analysis, $F(2, 348) = 0.688$, $p = 0.494$, $\eta^2_p = 0.004$. Similar to the participant-based analysis, a significant interaction of phase scrambling and accuracy was observed, $F(2, 348) = 4.451$, $p = 0.014$, $\eta^2_p = 0.025$. When participants correctly categorized pictures, a significant main effect of phase scrambling was observed, $F(2, 476) = 16.09$, $p < 0.001$, $\eta^2_p = 0.063$, indicating less positive ERPs in the 55% compared to the other degraded conditions: 80% vs. 55%, $t(239) = 5.426$, $p < 0.001$; 65% vs. 55%, $t(239) = 4.489$, $p < 0.001$. When participants could not categorize stimuli correctly, no significant effect of phase scrambling was observed, $F(2, 476) = 1.444$, $p = 0.238$, $\eta^2_p = 0.008$. On analyzing the Accuracy × Phase Scrambling interaction separately for each phase-scrambling level, significant effects of accuracy were observed for only the 55% phase-scrambling condition, $t(175) = 2.887$, $p = 0.004$ (see Figure 3)—with less positive ERPs for correct compared to incorrect categorizations—not for 80% and 65% phase scrambling, $t(239) = 0.31$, $p = 0.741$, and $t(239) = 0.542$, $p = 0.588$, respectively.

Different from the participant-based analysis, in which a significant effect of spatial coherence was observed, the main effect of SC rank did not reach standard significance in the image-based analysis, $F(1, 174) = 3.698$, $p = 0.056$, $\eta^2_p = 0.021$, nor was any interaction involving SC rank observed. Finally, the image-based analysis differed from the participant-based analysis in that a main effect of accuracy was observed, $F(1, 174) = 5.27$, $p = 0.02$, $\eta^2_p = 0.03$, indicating a less positive ERP amplitude for correctly categorized trials compared to incorrect trials.

In summary, both the participant-based and the image-based analyses revealed a significant interaction

of phase scrambling and categorization accuracy, with less positive ERPs for correct compared to incorrect trials in the least degraded level (55% phase scrambling). In the image-based analysis only, the main effect of accuracy reached significance. Finally, in the participant-based analysis, significant effects of spatial coherence and of phase scrambling were observed, but these effects were not observed in the image-based analysis, which used image as a random factor.

ERPs: Intact condition

Results in the intact condition are reported in Figure 5. In the intact condition, ERPs were scored and analyzed in the time interval 200–350 ms. In the participant-based analysis, a significant effect of procedure was observed, $F(1, 19) = 4.758$, $p = 0.042$, $\eta^2_p = 0.20$, indicating less pronounced positivity in the sequential compared to the mixed condition. No significant main effect or interaction involving SC rank was observed. Similarly, in the imaged-based analysis a significant procedure effect was observed, indicating that the ERP amplitude was less pronounced in the sequential compared to the mixed block, $F(1, 238) = 16.473$, $p < 0.001$, $\eta^2_p = 0.065$. No significant effect or interaction involving SC rank was observed.

Discussion

The present study examined the categorization of objects in natural scenes, aiming to functionally characterize an early ERP signature which was suggested to reflect the categorization of objects in scenes. The results support a scenario in which the early ERP differential reflects an activity that is related to later semantic categorization but does not build on low-level differences in color, spectral amplitude, or spatial coherence.

A number of previous studies have observed a reduction in ERP positivity over occipital sensors beginning 150 ms from stimulus onset when objects and scenes are selected or categorized (Codispoti et al., 2006a; Doniger et al., 2000; Johnson & Olshausen, 2003; Schendan & Kutas, 2007; Sehatapour, Molholm, Javitt, & Foxe, 2006). However, concerning the categorization of objects in scenes, there has been no previous assessment as to whether differences in diagnostic color, spatial frequency amplitude, and SC between categories are necessary to observe this early ERP differentiation. Building on a previous study (De Cesarei et al., 2013) which observed that the average spectral power was linearly related to the absolute amplitude of the P2, a positive ERP peak in the same

latency range as the early ERP differential, this study examined the extent to which the early ERP differential is associated with the correct categorization of objects in scenes when spectral differences in amplitude are ruled out.

The present results support the link between the early differential ERP negativity and correct categorization. In trials in which incorrect categorization was reported following picture offset, no increase in ERP amplitude was observed as pictures were revealed. A different pattern was observed in trials in which participants could, after picture offset, correctly categorize pictures as depicting people or animals. In these trials, ERP positive amplitude decreased compared to incorrect categorizations, until a significant difference was observed between scenes that would later be correctly or incorrectly categorized. This early ERP modulation is comparable in latency, direction, and topography to that observed in previous studies (Codispoti et al., 2006b; Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Johnson & Olshausen, 2003). Altogether, and consistent with previous results in a go/no-go task (Rousselet et al., 2004; VanRullen & Thorpe, 2001), these results suggest a link between this early ERP difference and categorization per se.

The present ERP results indicate that, as early as 150 ms, processing of cues which are predictive of accurate categorization is taking place. The possibility that accurate categorization of objects in scenes may begin early in time is supported by the results of a previous study, which used instructed saccades toward targets as an index of stimulus categorization (Kirchner & Thorpe, 2006). In that study, two intact monochromatic pictures were briefly presented in the left and right visual fields, and correct saccades toward the target image were first observed 120–130 ms after stimulus onset. As the preparation and execution of an instructed saccade requires substantially less time compared to manual responses (Schiller & Kendall, 2004), the authors concluded that at about 100 ms, categorization of this kind of image could begin. Taken together, electrophysiological and behavioral results point to the high efficiency of the visual system in using the available sensory information to modulate visual processing (ERPs) or guide eye behavior (eye tracking), ultimately supporting the categorization process beginning as early as 150 ms after picture onset.

In the present study, dissociations were observed between behavioral responses, which were executed about 1.5 s after picture onset (about 1630 ms), and early ERPs in the interval 150–300 ms after picture onset. Specifically, early ERPs were independently modulated by accuracy and spatial coherence, and no interaction was observed between these two factors; on the other hand, in behavioral performance, spatial coherence and accuracy were associated, and accuracy

was higher for pictures that were low in SC, compared to high, in the 65% and 55% conditions. Additionally, an ERP modulation by accuracy was observed in the 55% but not the 65% phase-scrambling condition, despite behavioral accuracy being above chance in both conditions. Finally, task context modulated behavioral accuracy but not ERPs in the 65% and 55% conditions, and vice versa (ERPs but not performance) in the intact condition. One possibility is that visual information, including that concerning task context and spatial coherence, continues to be processed later than 300 ms and contributes to a further increase in behavioral accuracy compared to what can be achieved based on early visual processing. Previous models of visual processing suggest that, as visual analysis proceeds, more complex processing is carried out, involving recurrent feedback from other structures such as the orbitofrontal cortex (Bar, 2004) and processing of visual input at a finer scale (Hegdé, 2008); as a result of this continued processing, a high accuracy observed in a late time interval can build on more information than that available shortly after picture onset. In the present paradigm, continuing analysis following the early ERP interval reported here may have been promoted by the chosen paradigm, in which participants had to respond after picture offset. Alternatively, it may be that the processes which subtend correct behavioral categorization are only partially reflected in the early occipital ERP modulation, and are rather evident in single-trial EEG or event-related oscillatory activity; in support of this possibility, earlier studies using intact images have indicated that correct categorization can be achieved at very short latencies following picture onset, with both manual and ocular responses (Kirchner & Thorpe, 2006; Thorpe et al., 1996).

Here, we examined whether differences in spatial coherence between scenes modulate categorization and electrocortical correlates of the processing of objects in scenes. The behavioral categorization of objects in scenes which, in the intact version, had a more coherent composition (low SC) was less hindered by degradation than that regarding scenes that were more locally fragmented (high SC). More specifically, performance varied in terms of both accuracy and response times. These effects were observed despite participants' responding after picture offset and not being required to respond as quickly as possible. Even under these unconstrained conditions, slower responses and higher error rates seem to indicate that stimuli which were high in SC required more time to be processed and categorized, compared to those low in SC. Taken together, these results suggest that the SC of intact images may predict overall difficulty resulting from phase scrambling.

Concerning the effects of SC on ERPs, a main effect of SC was observed in the participant-based analysis in the intact as well as the degraded conditions, with less positive ERPs for stimuli which were low compared to high in SC. The direction of this effect appears to be consistent with the findings of previous studies (Groen et al., 2013) and with the observation that scenes which are low in SC are easier to categorize compared to those which have high SC. However, it should be noted that the effect of SC on ERP amplitude did not interact with the effect of accuracy and phase scrambling. Moreover, when physical differences between the images were statistically controlled using image as a random factor, no significant effects of SC were observed. This pattern of results suggests that spatial coherence modulates early ERPs independently of later correct categorization. However, spatial coherence may be analyzed at a later time interval or in a less time-locked manner than is reflected by ERPs, and may contribute to accurate categorization of objects in scenes as reflected by behavioral accuracy.

It has been suggested that rapid categorization reflects the detection of one of more diagnostic features which define a target category (Evans, Horowitz, & Wolfe, 2011; Evans & Treisman, 2005) rather than a full semantic recognition of the scene. The understanding of visual scenes may be performed at very different levels, including perceptual decision (e.g., upwards/downwards), categorization of global scene content (e.g., indoor/outdoor), gist understanding, categorization of objects in scenes, and many other forms. Importantly, different perceptual cues, image statistics, and even processing modes (e.g., coarse to fine or vice versa) may support these different forms of scene understanding (De Cesarei & Loftus, 2011; Morrison & Schyns, 2001). For the extent to which the present task required the detection of humans and animals, it should suffice to detect a feature (e.g., glasses for humans or fangs for animals) which is unique to either category. The present results suggest that, as early as 150 ms, the diagnostic features which contribute to rapid categorization are being processed. Importantly, these features are distinct from those which were equalized or controlled here, namely color, global spatial frequency spectrum, and spatial coherence.

The present results, concerning the categorization of objects in natural scenes, complement previous findings in the domain of face perception. Those previous studies investigated a similar question, namely to what extent the well-known N170 component, a negative differential activity which is elicited by the perception of faces over occipitotemporal areas, merely reflects the presence of low-level perceptual cues or rather indicates that these cues are being processed into the percept of a face (Rossion & Jacques, 2008). Those studies indicated that, although cue-specific activity may be elicited in earlier intervals (less than 100 ms), true categorical

effects are only observed in the classic N170 window (Rossion & Caharel, 2011). Following the N170 time interval, other studies which examined a face/car categorization paradigm observed that a differential component around 200 ms reflects the accumulation of the available perceptual evidence, with the ultimate goal of performing a perceptual decision (Philiastides & Sajda, 2006, 2007).

The manipulation of task context was not effective here in modulating early ERP signatures of categorization in the degraded conditions. This result does not support the possibility that, in separate trials, coarse information interacts with fine-grained information, thus favoring categorization in the same way that has been suggested for the processing of a single scene (Bar, 2004). Rather, this result is consistent with several previous results which indicate that this early ERP modulation requires little or no attentional resources, is resistant to a concurrent selective-attention task (Fei-Fei, VanRullen, Koch, & Perona, 2005), can be similarly observed when two or more scenes have to be processed in parallel (Rousselet, Fabre-Thorpe, & Thorpe, 2002; Rousselet, et al., 2004), and is not facilitated by 3 weeks of training (Fabre-Thorpe et al., 2001). A significant effect of task context was observed on behavioral responses after picture offset. This behavioral effect may indicate that the effects of task context in the degraded conditions are exerted in a later time interval than the window of 150–300 ms that was examined here. However, the direction of the behavioral effect indicated worse performance in the sequential compared to the mixed condition, suggesting that the sequential context interfered with correct categorization rather than supported it (Bruner & Potter, 1964). Finally, an effect of task context was observed in the intact condition, with less pronounced ERPs in the sequential compared to the mixed condition. However, behavioral performance did not differ between the two contexts. Therefore, future studies could better explore the effects of task context on the behavioral categorization of objects in scenes and on its electrocortical correlates.

Conclusions

We examined the early ERP modulations associated with the categorization of objects in natural scenes which were equalized in color and spectral amplitude and controlled in terms of spatial coherence. An occipitotemporal ERP modulation was observed for correctly versus incorrectly categorized objects in scenes at an intermediate level of phase scrambling. These results suggest that, as early as 150 ms, an ERP modulation may reflect the processing of the available

diagnostic features, which is predictive of later accurate behavioral categorization.

Footnote

[1] In a preliminary analysis, the same pattern of results was also observed when selecting electrodes and time intervals based on the amplitude of the correct–incorrect differential. However, as this selection may lead to a statistical bias (as suggested by Kilner, 2013), we chose to report data from a region and time interval of interest based on independent data.

References

Antal, A., Keri, S., Kovacs, G., Janka, Z., & Benedek, G. (2000). Early and late components of visual categorization: An event related potential study. *Cognitive Brain Research, 9,* 117–119, doi:10.1016/S0926-6410(99)00053-1.

Arsenault, E., Yoonessi, A., & Baker, C. Jr. (2011). Higher order texture statistics impair contrast boundary segmentation. *Journal of Vision, 11*(10):14, 1–15, doi:10.1167/11.10.14. [PubMed] [Article]

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5,* 617–629, doi:10.1038/nrn1476.

Bieniek, M. M., Pernet, C. R., & Rousselet, G. A. (2012). Early ERPs to faces are driven by phase, not amplitude spectrum information: Evidence from parametric, test-retest, single-subject analyses. *Journal of Vision, 12*(13):12, 1–24, doi:10.1167/12.13.12. [PubMed] [Article]

Bruner, J. S., & Potter, M. C. (1964, April 24). Interference in visual recognition. *Science, 144*(3617), 424–425, doi:10.1126/science.144.3617.424.

Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review, 101,* 446–469, doi:10.1037/0033-295X.101.3.446.

Codispoti, M., Ferrari, V., De Cesarei, A., & Cardinale, R. (2006a). Implicit and explicit categorization of natural scenes. *Progress in Brain Research, 156,* 53–65, doi:10.1016/S0079-6123(06)56003-0.

Codispoti, M., Ferrari, V., Junghöfer, M., & Schupp, H. T. (2006b). The categorization of natural scenes: Brain attention networks revealed by dense sensor ERPs. *NeuroImage, 31,* 881–890, doi:10.1016/j.neuroimage.2006.04.180.

Crouzet, S. M., & Serre, T. (2011). What are the visual features underlying rapid object recognition? *Frontiers in Psychology, 2,* 326, doi:10.3389/fpsyg.2011.00326.

Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology, 12,* R476–477, doi:10.1016/S0960-9822(02)00960-0.

De Cesarei, A., Codispoti, M., Schupp, H. T., & Stegagno, L. (2006). Selectively attending to natural scenes after alcohol consumption: An ERP analysis. *Biological Psychology, 72,* 35–45, doi:10.1016/j.biopsycho.2005.06.009.

De Cesarei, A., & Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychonomic Bulletin & Review, 18*(5), 840–847, doi:10.3758/s13423-011-0133-6.

De Cesarei, A., Mastria, S., & Codispoti, M. (2013). Early spatial frequency processing of natural images: An ERP Study. *PLoS One, 8*(5), e65103, doi:10.1371/journal.pone.0065103.

Delorme, A., Rousselet, G. A., Macé, M. J., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research, 19,* 103–113, doi:10.1016/j.cogbrainres.2003.11.010.

Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B., Snodgrass, J. G., Schroeder, C. E., & Javitt, D. C. (2000). Activation time course of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure processes. *Journal of Cognitive Neuroscience, 12,* 615–621, doi:10.1162/089892900562372.

Evans, K. K., Horowitz, T., & Wolfe, J. M. (2011). When categories collide: Accumulation of information about multiple categories in rapid scene perception. *Psychological Science, 22,* 739–746, doi:10.1177/0956797611407930.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1476–1492, doi:10.1037/0096-1523.31.6.1476.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13,* 171–180, doi:10.1162/089892901564234.

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 1*(2), 893–924, doi:10.1080/13506280444000571.

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience, 8,* 1643–1646, doi:10.1038/nn1608.

Fize, D., Boulanouar, K., Chatel, Y., Ranjeva, J. P., Fabre-Thorpe, M., & Thorpe, S. J. (2000). Brain areas involved in rapid categorization of natural images: An event-related fMRI study. *NeuroImage, 11,* 634–643, doi:10.1006/nimg.2000.0585.

Gaspar, C. M., & Rousselet, G. A. (2009). How do amplitude spectra influence rapid animal detection? *Vision Research, 49*(24), 3001–3012, doi:10.1016/j.visres.2009.09.021.

Geusebroek, J. M., & Smeulders, A. W. M. (2005). A six-stimulus theory for stochastic texture. *International Journal of Computer Vision, 62*(1–2), 7–16, doi:10.1007/s11263-005-4632-7.

Ghebreab, S., Smeulders, A. W. M., Scholte, H. S., & Lamme, V. A. F. (2009). A biologically plausible model for rapid natural scene identification. *Advances in Neural Information Processing Systems, 22,* 629–637.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition, 12,* 878–892, doi:10.1080/13506280444000562.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20,* 464–472, doi:10.1111/j.1467-9280.2009.02316.x.

Greene, M. R., & Oliva, A. (2010). High-level aftereffects to global scene property. *Journal of*

Experimental Psychology: Human Perception and Performance, 36, 1430–1442, doi:10.1037/a0019058.

Groen, I. I. A., Ghebreab, S., Lamme, V. A. F., & Scholte, H. S. (2012). Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories. *PLoS Computational Biology, 8*(10), e1002726, doi:10.1371/journal.pcbi.1002726.

Groen, I. I. A., Ghebreab, S., Prins, H., Lamme, V.A.F., & Scholte, H. S. (2013). From image statistics to scene gist: Evoked neural activity reveals transition from low-level natural image structure to scene category. *The Journal of Neuroscience, 33*(48), 18814–18824, doi:10.1523/JNEUROSCI.3128-13.

Hansen, B. C., Jacques, T., Johnson, A. P., & Ellemberg, D. (2011). From spatial frequency contrast to edge preponderance: The differential modulation of early visual evoked potentials by natural scene stimuli. *Visual Neuroscience, 28*(3), 221–237, doi:10.1017/S095252381100006X.

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology, 84,* 405–439, doi:10.1016/j.pneurobio.2007.09.001.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 604–610, doi:10.1037/0096-1523.7.3.604.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision, 3*(7):4, 499–512, doi:10.1167/3.7.4. [PubMed] [Article]

Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision, 9*(1):2, 1–16, doi:10.1167/9.1.2. [PubMed] [Article]

Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology, 37,* 523–532, doi:10.1111/1469-8986.3740523.

Kayser, C., Körding, K. P., & König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Current Opinion in Neurobiology, 14,* 468–473, doi:10.1016/j.conb.2004.06.002.

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology, 124,* 2062–2063, doi:10.1016/j.clinph.2013.03.024.

Kirchner, H., & Thorpe, J. S. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776, doi:10.1016/j.visres.2005.10.002.

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology, 3,* 525–551, doi:10.1016/0010-0285(72)90021-7.

Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the Natural/Man-made distinction. *Journal of Vision, 8*(1):4, 1–9, doi:10.1167/8.1.4. [PubMed] [Article]

Luck, S. (2005). *An introduction to the event-related potential technique.* Cambridge, MA: MIT Press.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review, 8*(3), 454–469, doi:10.3758/BF03196180.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210, doi:10.1006/cogp.1999.0728.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175, doi:10.1023/A:1011139631724.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*(1), 23–36, doi:10.1016/S0079-6123(06)55002-2.

Peyk, P., De Cesarei, A., & Junghöfer, M. (2011). ElectroMagnetoEncephalography software: Overview and integration with other EEG/MEG toolboxes. *Computational Intelligence and Neuroscience, 2011,* 861705, doi:10.1155/2011/861705.

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex, 16*(4), 509–518, doi:10.1093/cercor/bhi130.

Philiastides, M. G., & Sajda, P. (2007). EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *Journal of Neuroscience, 27*(48), 13082–13091, doi:10.1523/JNEUROSCI.3540-07.2007.

Potter, M. C. (1975, March 14). Meaning in visual search. *Science, 187*(4180), 965–966, doi:10.1126/science.1145183.

Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research, 51*(12), 1297–1311, doi:10.1016/j.visres.2011.04.003.

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electro-

physiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage, 39*(4), 1959–1979, doi:10.10.16/j.neuroimage.2007.10.011.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629–630, doi:10.1038/nn866.

Rousselet, G. A., & Pernet, C.R. (2011). Quantifying the time course of visual object processing using ERPs: It's time to up the game. *Frontiers in Psychology, 23*(2), 107, doi:10.3389/fpsyg.2011.00107.

Rousselet, G. A., Pernet, C. R., Bennett, P. J., & Sekuler, A. B. (2008). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience, NN, 9,* 98, doi:10.1186/1471-2202-9-98.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research, 44*(9), 877–894.

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience, 19,* 734–749, doi:10.1162/jocn.2007.19.5.734.

Schiller, P. H., & Kendall, J. (2004). Temporal factors in target selection with saccadic eye movements. *Experimental Brain Research, 154*(2), 154–159, doi:10.1007/s00221-003-1653-8.

Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., & Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology, 118,* 98–104, doi:10.1016/j.clinph.2006.09.003.

Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W., & Lamme, V. A. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision, 9*(4):29, 1–15, doi:10.1167/9.4.29. [PubMed] [Article]

Sehatapour, P., Molholm, S., Javitt, D. C., & Foxe, J. J. (2006). Spatiotemporal dynamics of human object recognition processing: An integrated high-density electrical mapping and functional imaging study of "closure" processes. *NeuroImage, 29,* 605–618, doi:10.1016/j.neuroimage.2005.07.049.

Simoncelli, E. P., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24,* 1193–1216, doi:10.1146/annurev.neuro.24.1.1193.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522, doi:10.1038/381520a0.

Thorpe, S., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience, 14,* 869–876, doi:10.1046/j.0953-816x.2001.01717.x.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14*(3), 391–412, doi:10.1088/0954-898X/14/3/302.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5,* 682–687, doi:10.1038/nn870.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology, 2,* 365, doi:10.3389/fpsyg.2011.00365.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13,* 454–461, doi:10.1162/08989290152001880.

Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vision Research, 46*(8–9), 1520–1529, doi:10.1016/j.visres.2005.11.008.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, Instruments, & Computers, 42,* 671–684, doi:10.3758/BRM.42.3.671.

Yanulevskaya, V., & Geusebroek, J. M. (2009). Significance of the Weibull distribution and its submodels in natural image statistics. *Proceedings of the 4th International Conference on Computer Vision Theory and Applications 1,* 355–362.