Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Discriminating scene categories from brain activity within 100 milliseconds

Matthew X. Lowe [a,*], Jason Rajsic [a], Susanne Ferber [a,b] and Dirk B. Walther [a,b]

[a] Psychology Department, University of Toronto, Canada
[b] Rotman Research Institute, Baycrest, Toronto, Canada

ARTICLE INFO

Article history:
Received 23 October 2017
Reviewed 10 January 2018
Revised 26 February 2018
Accepted 1 June 2018
Action editor Sam Schwarzkopf
Published online xxx

Keywords:
EEG
Time course
Perception
Line drawings
Global
Visual cortex

ABSTRACT

Humans have the ability to make sense of the world around them in only a single glance. This astonishing feat requires the visual system to extract information from our environment with remarkable speed. How quickly does this process unfold across time, and what visual information contributes to our understanding of the visual world? We address these questions by directly measuring the temporal dynamics of the perception of colour photographs and line drawings of scenes with electroencephalography (EEG) during a scene-memorization task. Within a fraction of a second, event-related potentials (ERPs) show dissociable response patterns for global scene properties of content (natural versus manmade) and layout (open versus closed). Subsequent detailed analyses of within-category versus between-category discriminations found significant dissociations of basic-level scene categories (e.g., forest; city) within the first 100 msec of perception. The similarity of this neural activity with feature-based discriminations suggests low-level image statistics may be foundational for this rapid categorization. Interestingly, our results also suggest that the structure preserved in line drawings may form a primary and necessary basis for visual processing, whereas surface information may further enhance category selectivity in later-stage processing. Critically, these findings provide evidence that the distinction of both basic-level categories and global properties of scenes from neural signals occurs within 100 msec.

1. Introduction

A fundamental property of human perception is the ability to efficiently perceive and understand the visual world, from dense forests to soaring cityscapes, in only a single glance (Greene & Oliva, 2009a, 2009b). Investigations have often explored the neural mechanisms underlying scene perception within the spatial domain. These investigations have revealed that in addition to early visual cortex, the scene-selective parahippocampal place area (PPA) (Epstein & Kanwisher, 1998) is part of a network of regions subserving the ability to

* *Corresponding author*. Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON, M5S 3G3, Canada.
E-mail address: matthew.lowe@utoronto.ca (M.X. Lowe).

https://doi.org/10.1016/j.cortex.2018.06.006

distinguish among natural scene categories (Walther, Caddigan, Fei-Fei, & Beck, 2009). Critically, category-specific neural activity patterns were found to be similar between colour photographs and line drawings (Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011), suggesting that the structure preserved in line drawings is sufficient for scene categorization. Yet recent investigations have revealed a role for the PPA in processing the surface information of a scene (Lowe, Rajsic, Gallivan, Ferber, & Cant, 2017, 2016; Park & Park, 2017). Both these features may therefore be integral to our experience of the visual world. To fully understand the mechanisms underlying these processes, however, we must not only seek to explore these dynamics across space, but also across time. These features may be closely interwoven spatially within the human brain, affording us limited precision when exploring their unique neural markers. Within the temporal domain, however, they may unfold on different timescales along a hierarchy of visual processing.

Pioneering work using electroencephalography (EEG) has shown event-related potentials (ERPs) reflect properties of a visual stimulus within the first 150 msec following an image (Thorpe, Fize, & Marlot, 1996; Vanrullen & Thorpe, 2001). Intracranial recordings have reported selectivity for object categories within only 100 msec in monkeys (Vogels, 1999) and humans (Liu, Agam, Madsen, & Kreiman, 2009). Multiple category levels may be extracted, including basic-level categories and global scene properties (Oliva & Torralba, 2001). Basic-level categories correspond to the most common categorical representation (e.g., forest, mountain), and members of these categories tend to share similar shapes and functions. In contrast, global scene properties correspond to various kinds of abstraction, which include descriptors of scene content (natural or manmade) as well as spatial boundary (open or closed scene boundaries) and scene affordance (navigability), among others. These global properties capture the holistic and diagnostic structure of an image and typically represent category resemblance at a low level rather than the more high-level meaning of a scene (Oliva & Torralba, 2006). An important question of concern involves defining the precise temporal ordering of these different levels of categorization (e.g., basic versus global). Neurophysiological evidence in humans has revealed sensitivity to global scene properties within ~250 msec (Cichy, Khosla, Pantazis, & Oliva, 2017; Groen, Ghebreab, Lamme, & Scholte, 2016, 2013; Harel, Groen, Kravitz, Deouell, & Baker, 2016), yet basic-level categories may differ from these global categorizations. There is some behavioural research which suggests that basic categories emerge prior to global scene properties (Rosch, Mervis, Grey, Johnson, & Boyes-Braem, 1976; Tversky & Hemenway, 1983), but there is also work which suggests the opposite is true (Greene & Oliva, 2009a, 2009b; Kadar & Ben-Shahar, 2012; Loschky & Larson, 2010; Sun, Ren, Zheng, Sun, & Zheng, 2016). In contrast, evidence has indicated that a basic or global category advantage may be flexible (Banno & Saiki, 2015), or require the same amount of information for scene recognition (Fei-Fei, Iyer, Koch, & Perona, 2007).

Visual properties may influence properties of the neural signal. For instance, there is evidence to suggest that edge-based information is processed with higher priority than surface information within the visual stream (Fu et al., 2016). The precise temporal dynamics of this relationship in scene perception, however, are unclear. The present study applied a multifaceted approach using EEG to examine the neural time course of natural scene processing for six scene categories (beach; city; forest; highway; mountain; office) of colour photographs and line drawings during an orthogonal memorization task. Our approach explored the temporal dynamics of scene perception across basic-level categories and global scene properties using traditional event-related potentials (ERPs), and a novel classification analysis which examined the correlations of within-category versus between-category discriminations at a precise temporal scale. The aims of the present study were therefore twofold: To investigate the relative neural timing for basic-versus global-scene categorization, and to explore how available scene information contained within line drawings and colour photographs influences this timing.

2. Materials and methods

2.1. Participants

Sixteen paid participants (13 females, mean age 19.1 ± 2.3 years; all right-handed) with normal or corrected-to-normal visual acuity and no history of neurological impairments were recruited from the University of Toronto community. This sample size is consistent with those used in previous research exploring image categorization and EEG (e.g., Groen, Ghebreab, Prins, Lamme, & Scholte, 2013; Harel et al., 2016; Thorpe, 1996; Vanrullen & Thorpe, 2001). All participants gave informed consent in accordance with the University of Toronto Ethics Review Board. No participants were excluded from the analysis.

2.2. Experimental design and statistical analysis

The experiment was programmed, displayed, and analysed using Matlab (MathWorks, Natick, MA, R2014a) software running on a desktop computer, with a ViewSonic 21-inch monitor (1280 × 1024 resolution, 85 Hz refresh rate). The viewing distance was 57 cm, and participants made responses with their right hand on the mouse, and their left hand on the "z", and "x" keys of a standard keyboard. To initiate the start of either the study or test phase between experimental blocks, participants pressed the "enter" key with either hand. Continuous, unreferenced EEG was recorded at a sampling rate of 512 Hz using a BioSemi ActiveTwo system with 64 Ag/AgCl scalp electrodes in standard 10–20 placement with additional electrodes at each mastoid for off-line re-referencing, below each eye for blink detection, and at the outer canthus of each eye for lateral eye movement detection. All re-referencing and filtering was done off-line using ERPLAB (v13.0.0) software.

432 colour photographs from six real-world scene categories (three natural: beaches, forests, and mountains; three man-made: city streets, highways, and offices; 72 per category; See Fig. 1 for stimuli examples) were used following previous research (Walther et al., 2011). These images were

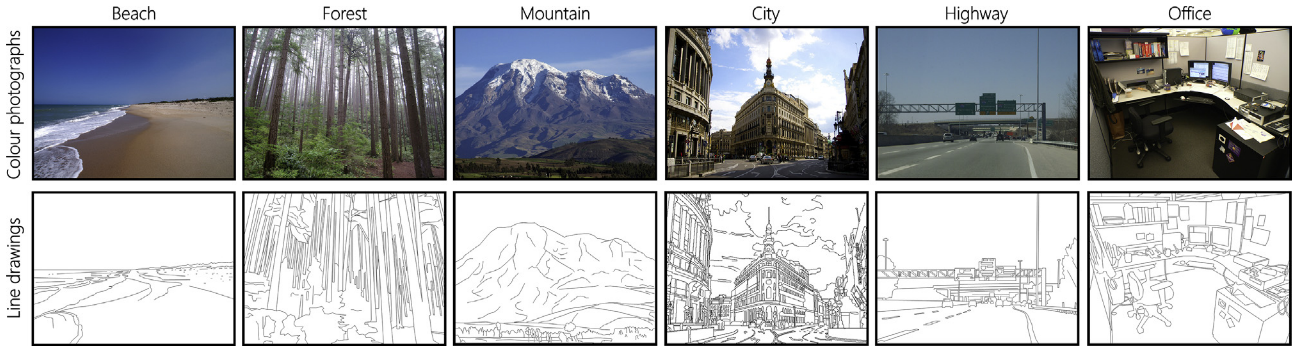

Fig. 1 – Stimuli examples from each of the six scene categories as colour photographs (above), and their corresponding line drawings (below).

chosen from a set of 4,025 images downloaded from the Internet as the best exemplars of their categories according to ratings by an average of 137 observers per image (Torralbo et al., 2013). Images were resized to 800 × 600 pixels. Line drawings were produced by trained artists at the Lotus Hill Research Institute (Wuhan, Hubei Province, People's Republic of China), who traced outlines in the colour photographs using a custom graphical user interface. Line drawings were rendered by connecting the anchor points with black straight lines on a 800- × 600-pixel white background. Note that line drawings, while largely preserving scene structure and content, alter the spatial frequency spectrum drastically. Photographic images of natural scenes typically contain most contrast energy at low spatial frequencies, with a drop-off in energy following an inverse power law for higher spatial frequencies (Field, 1987). By contrast, for line drawings, almost all energy is shifted to high spatial frequencies (Walther et al., 2011, Figure S5).

Participants were instructed to view and memorize a series of images, after which they would receive a recall test. During each trial, participants were asked to maintain central fixation and refrain from all eye movements, including blinking. In the study-phase, images were presented within twelve randomized blocks. Each block contained sixty-six randomly-presented images obtained from all of the six scene categories, either for colour photographs or line drawings. Images from each stimulus category were evenly distributed across blocks of images. After the initial instruction to view and memorize a series of images, participants initiated each block by pressing "enter" on the keyboard, and began each trial by clicking the left mouse button. Trials were self-initiated to allow for sufficient time for rest between trials which would therefore minimize artifacts caused by eye movements. The onset of each trial with respect to participants' button presses was randomly jittered across three temporal selections (300 msec, 500 msec, or 700 msec), and each image was displayed for 1500 msec. Following the presentation of sixty-six randomly presented study images, participants were asked to recall whether an image was new or old during the test-phase. The test-phase consisted of twelve self-initiated and randomly-presented trials containing six new and six old images, and participants were instructed to press the "x" key if the image was old (familiar), and the "z" key if the image was new (unfamiliar).

Each individual data set was filtered using a finite impulse response (FIR) filter high-passed at .01 Hz and low-passed at 70 Hz. The continuous data were then re-referenced to the average of the two mastoids. Lateral eye movements were detected using a step-like function (horizontal eye electrodes, threshold: 30 μv) and blinks were detected using a moving window peak-to-peak threshold (vertical eye electrodes, threshold: 80 μv). Data was segmented into 1200 msec analysis windows – composed of the 200 msec immediately preceding onset of scene stimuli as the baseline period, which was baseline corrected to zero, and 1000 msec post-stimulus onset as the critical window. Analyses focused on a single site of interest: centro-occipital (Oz) cortex, because we were interested in the earliest neural activity as visual information arrives in neocortex. Only study-phase trials were used in the analysis.

To identify the time-course of content and boundary, we calculated *t*-statistics comparing the amplitude of pairs of ERPs (e.g., natural versus man made) at each time point of the ERP, using an alpha value of .01, to determine the first time point at which they differed. Given that this entailed calculating a large number of *t*-values, the possibility of a false alarm is high. To account for this possibility, we sought to distinguish between spuriously significant comparisons at random points in the time series from meaningfully significant comparisons that would cluster in time (that is, consecutive time-points would also show differences in amplitude). To this end, we calculated the number of consecutive, significant time points that would be expected by chance in a series of paired comparisons where no differences exist using a Monte Carlo simulation. We generated 100,000 epochs of normally distributed noise data. The data were temporally filtered the same way as the experimental data, and thresholded at the value where the cumulative normal distribution was smaller than $\alpha = .01$ or exceeded $(1-\alpha) = .99$. The time points passing either threshold were deemed to be significant purely by chance. We analysed clusters of temporally contiguous significant time points in all 100,000 samples. Our simulation revealed that a run of length 11 or greater occurred in fewer than 5% of the simulated noise data sets, and thus used this as a criterion for considering a given pairwise comparison's statistical significance as meaningful (i.e., it needed to be followed by at least 10 statistically significant time-points).

3. Results

3.1. Global scene properties

Previous research has shown that the distinction between global scene properties, such as the content of a scene (i.e., natural versus manmade), and the spatial boundary, or layout, of a scene (i.e., open versus closed) is an important factor mediating scene perception and our ability to navigate through an environment (Kravitz, Peng, & Baker, 2011; Park, Brady, Greene, & Oliva, 2011). We therefore began our analysis by investigating the time course of these scene properties across scene categories and stimulus type (stimulus examples can be seen in Fig. 1). To do so, we examined scene content by comparing natural (beach, forest, mountain) and manmade (city, highway, office) scenes, and we examined boundary by comparing open (beach, highway) and closed (forest, city) scenes, following previous research (Oliva & Torralba, 2001; 2006). We then examined, in colour photographs and line drawings separately, whether conditions (e.g., natural versus manmade) could be dissociated from patterns of responses during early electrophysiological activity. By including both colour photographs and line drawings in our investigation, we could examine the extent to which surface information, such as colour and texture, influences the time course of scene categorization in the human brain. Site selection was based on the distribution of electrophysiological activity across the scalp for all study-phase images averaged across participants, irrespective of image type and scene category (Fig. 2). This activity revealed maximal response amplitudes over central occipital electrodes, and thus a centro-occipital site (Oz) was chosen as our region of interest. Note that we only used overall EEG power for site selection and not any of the discriminations between conditions. This site selection technique avoids statistical circularity similar to the "most active voxel" voxel selection technique (Pereira, Mitchell, & Botvinick, 2009).

An ERP analysis was performed for both colour photographs and line drawings, separately, for scene content (Fig. 3), and scene boundary (Fig. 4). After computing a difference wave for conditions in our site of interest, we compared each time point in the difference waves to zero using a paired-samples *t*-test ($\alpha = .01$) to find the earliest point at which the ERPs diverged. To ensure that these differences were not statistically spurious, we only considered timepoints where a significant comparison was followed by 10 additional statistically significant time points. We were therefore able to identify the earliest time point at which a cluster was significantly different from zero using a moving window with a minimum of 11 consecutive significant *t*-tests. The leading edge of said clusters occurred within the early visually-evoked P1 (50–130 msec) and P2 (150–275 msec) components (Busch, Debener, Kranczioch, Engel, & Herrmann, 2004; Calvo & Beltrán, 2014).

When examining scene content in line drawings, the earliest significant discrimination between natural and manmade scenes during the examined time windows occurred during the P1 component (84 msec). For colour photographs, the earliest significant discrimination for natural and manmade scenes occurred during the tail-end of the P1

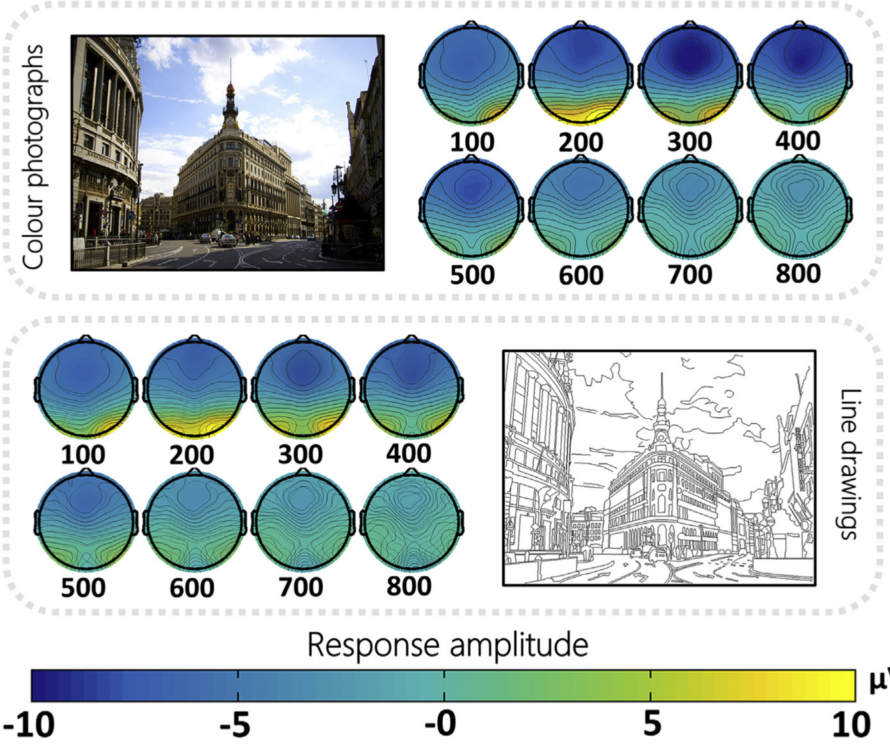

Fig. 2 – Site selection (above) was based on overall higher posterior response amplitudes using scalp maps averaged across participants for colour photographs and line drawings (below) from 100 to 800 msec. Site selection therefore included centro-occipital site (Oz) cortex.

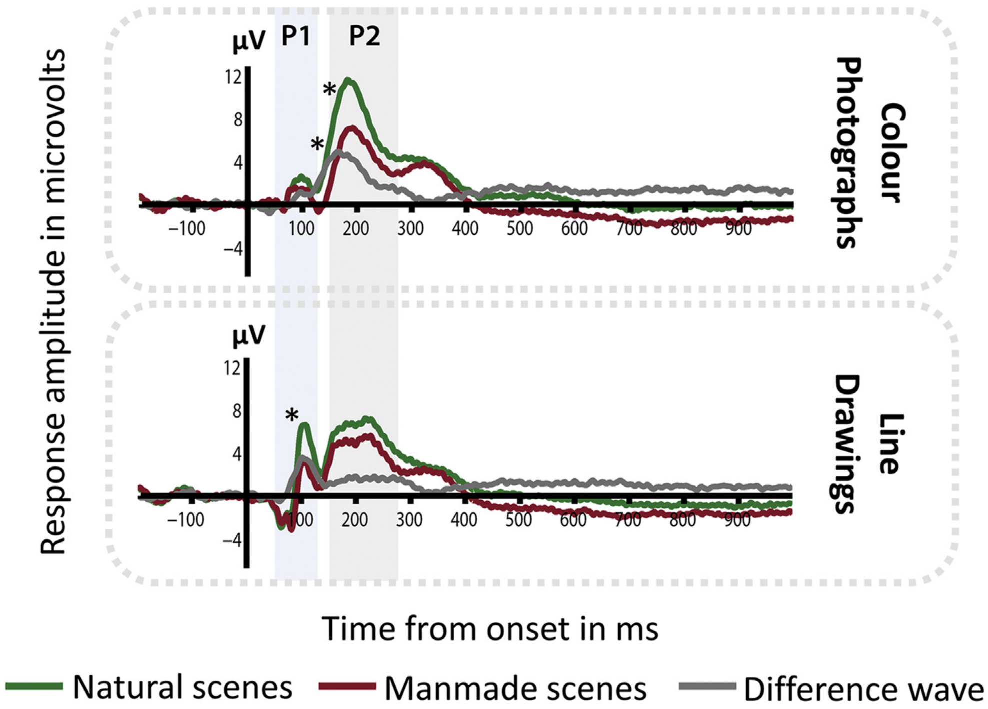

Fig. 3 – Event-related potentials (ERPs) and their difference waves for centro-occipital (Oz) cortex plotted in microvolts (µV) for natural (beach, forest, mountain) versus manmade (city, highway, office) scenes averaged across participants ($N = 16$). An asterisk indicates the earliest significant time point ($P < .01$) within a time component.

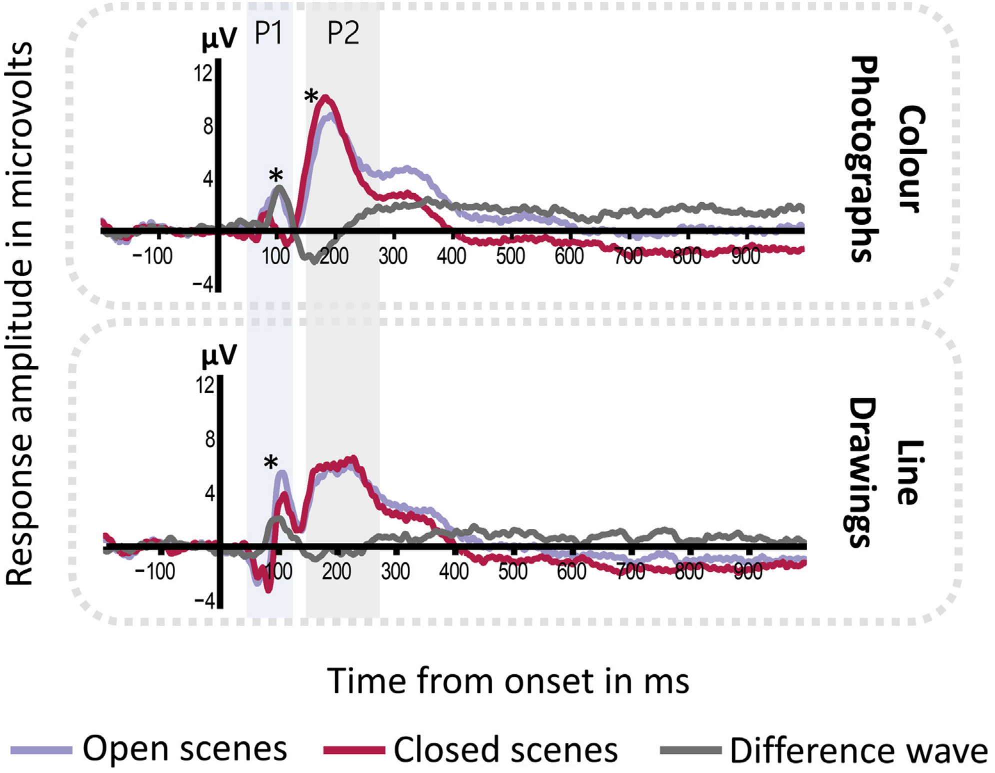

Fig. 4 – Event-related potentials (ERPs) and their difference waves for centro-occipital (Oz) cortex plotted in microvolts (µV) for open (beach, highway) versus closed (forest, city) scenes averaged across participants ($N = 16$). An asterisk indicates the earliest significant time point ($P < .01$) within a time component.

component (125 msec), with sustained discrimination into the P2 component. To determine whether these earliest time points are significantly different we performed a bootstrap analysis. We randomly resampled subject data with replacement and then performed the same ERP analyses to determine the earliest time points for scene content and boundary in line drawings and colour photographs. We repeated this bootstrap for 1000 resamples and assessed the difference in the earliest time points that significantly discriminated between image types using a paired-samples *t*-test (two-tailed) over these 1000 resamples. The difference between the earliest significant discrimination for scene content in line drawings (84 msec) and colour photographs (125 msec) was statistically significant ($t = 21.4$, $P < .001$).

When examining spatial boundary, early discrimination between open and closed scenes occurred during the P1 component for both line drawings (82 msec) and colour photographs (100 msec), with subsequent discrimination occurring during the P2 component for colour photographs. This difference was not statistically significant ($t = .6$, $P = .55$).

These results show that neural categorization of global scene properties occurs rapidly in both colour photographs and line drawings, even as early as the P1 component. Thus, the findings here provide evidence that neural processing as early as the P1 component (50–130 msec) with sustained activity extending through the P2 component (150–275 msec), may be used to extract and differentiate information from scene categories. Statistical comparisons of the first times at which scene content could be discriminated using colour photographs compared to line drawings showed that line drawings led to earlier discrimination than colour photographs. This was not true for boundary discrimination. These findings suggest that the structure preserved in line drawings forms a primary and necessary basis for the distinction of scene content (natural versus manmade), and that additional information from colour and texture is taken into account in a later-stage, or subsequent wave, of processing.

3.2. *Basic-level scene categories*

Our next investigation aimed to determine the earliest time point at which basic category levels of scenes can be extracted from neural activity. We first plotted averaged ERPs for each scene category in both colour photographs and line drawings (Fig. 5). ERP traces for the individual categories are very similar, and therefore investigating the differences between each pair of categories is impractical. We instead derive a measure of category selectivity using correlational similarity analysis of the ERP traces as a function of time. This analysis is inspired by the successful representational similarity analysis of fMRI data (Kriegeskorte, Mur, & Bandettini, 2008). Specifically, we analysed the similarity of voltage changes over time within and between scene categories. To do this, we computed two ERPs for each category. For each participant, scene category (beach, forest, mountains, city, highway, office), and stimulus type (colour photographs; line drawings), trials that remained after artefact rejection were randomly divided into two disjoint sets. 24 grand-average ERPs were then calculated by averaging across trials and participants for each scene category, stimulus type, and subset.

We next calculated similarity matrices by correlating the grand average ERPs at Oz for each scene category in set 1 with those in set 2, separately for photographs and line-drawings (Fig. 6A). This was done by computing the correlation of the ERP time courses between these independently computed average ERPs for every possible category pair (e.g., beach (half 1) with beach (half 2), beach (half 1) with city (half 2) etc.). For the correlation, the average amplitude at each time point at Oz was interpreted as a vector element, and the two vectors composed of the two corresponding ERP time courses were Pearson correlated. The diagonal entries of the resulting matrices contain correlations of ERPs for matching categories, and the off-diagonal entries for non-matching categories. We quantified scene selectivity by contrasting the average similarity (i.e., *r* value) of the diagonal entries with the average similarity of the off-diagonal entries (Fig. 6B), converting *r* values to a normal distribution using the inverse hyperbolic tangent (Fisher's *z* transform). In general, this approach computes a “categorization signal”, which measures the extent to which scene category can be recovered from temporal voltage patterns in the ERP being measured. This analysis is common for analysing *spatial* patterns of brain activity from fMRI data (Kriegeskorte et al., 2008). We apply it here to the *temporal* pattern of neural activity measured from each EEG electrode.

Average categorization signals over the entire time-course revealed that whole-ERP basic category similarity was strongest at the posterior electrode for both image types (Fig. 7). Given our interest in the time-course of this similarity and the general scalp distributions of the signal, we calculated similarity indices again, but using a moving window containing 30 samples from ERPs in each window (spanning 58.6 msec) for channel Oz. We chose to use 30 samples in this analysis to strike an optimal balance between temporal resolution (i.e., smaller time windows) and statistical power (i.e., reliable estimations of the correlation between time-series). These time courses of scene selectivity are shown in Fig. 7. Green traces depict the category signal over time for individual iterations of sub-dividing the data, and the mean category signal over all 1000 iterations is the solid, black trace. To quantify the time at which basic category selectivity emerges in the ERP, we isolated the earliest time point at which the diagonal and off-diagonal entries differed on each of the 1000 iterations, using an independent samples *t*-test, with an alpha threshold of .01.

For colour photographs, the median of the earliest time points that exhibited scene selectivity across the 1000 iterations was 92 msec, $SD = 11$ msec, with all 1000 iterations showing at least one time point with a significant categorization signal at $P < .01$. For line drawings, the median of the earliest scene category signals appeared at 86 msec, $SD = 17$ msec, with 997 iterations showing at least one time point with significant categorization. Comparing with Fig. 7, these values align with the time course of the average categorization signal. Note that, given the moving-window approach, these are conservative estimates: the 92 msec estimate of scene category discrimination at Oz for colour photographs, for example, is based on a correlation of the voltage in the 34 msec-92 msec range window of grand average ERPs, and thus may reflect differences in neural responses even earlier than 92 msec for different scene

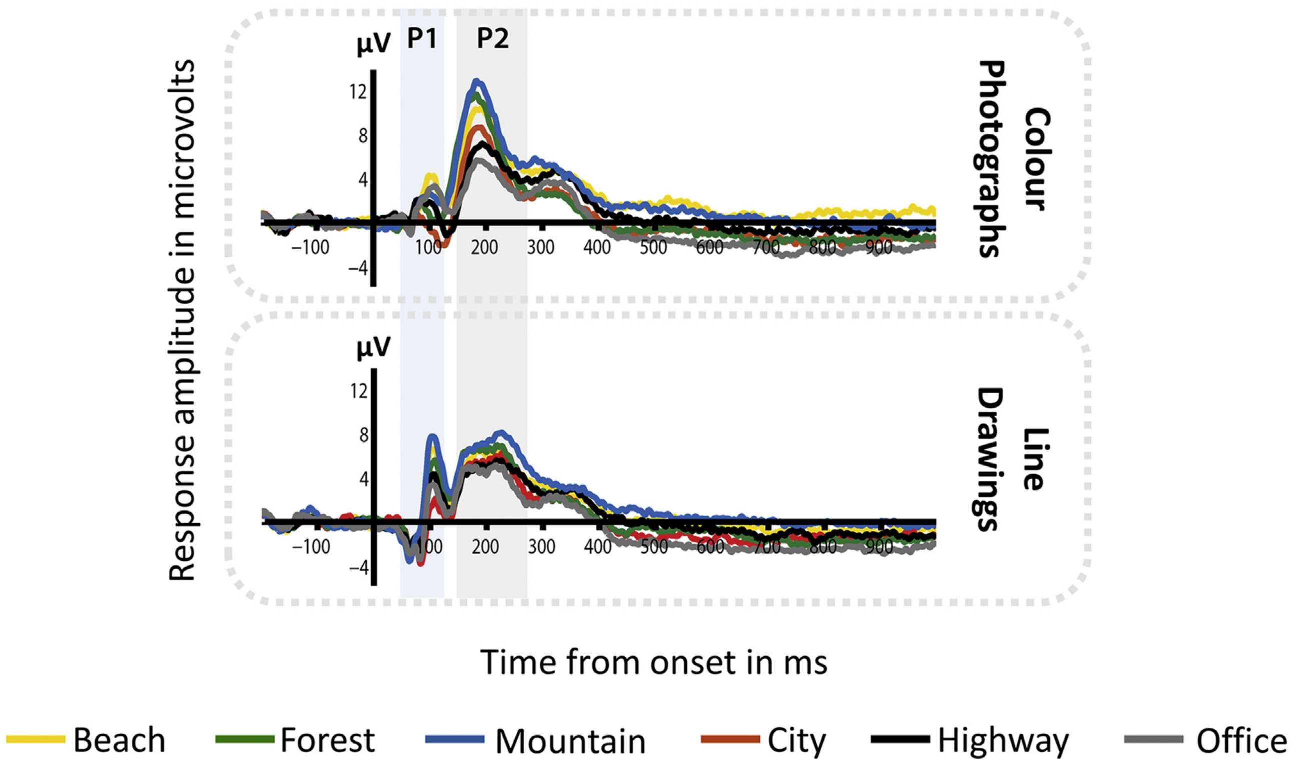

Fig. 5 – **Event-related potentials (ERPs) for centro-occipital (Oz) cortex plotted in microvolts (µV) for all scene categories (beach, forest, mountain, city, highway, office) averaged across participants (N = 16).**

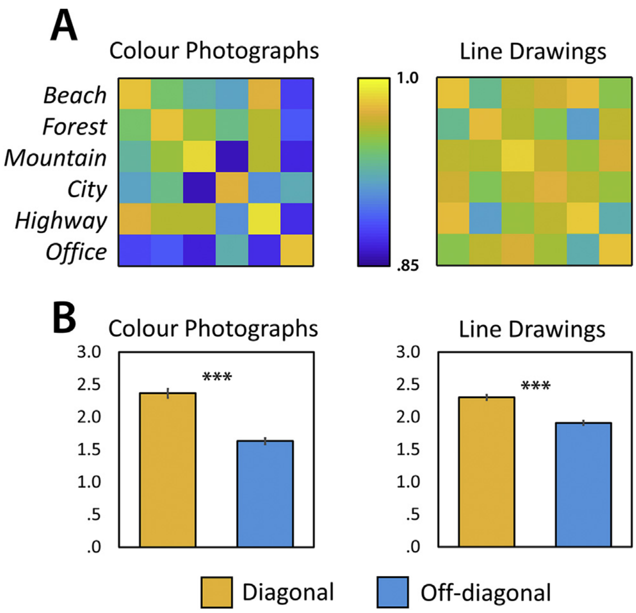

Fig. 6 – **(A) Similarity matrices showing the grand-averaged correlations of each scene category for all participants (N = 16) averaged across time (0–1000 msec) in centro-occipital (Oz) cortex (B) Independent samples t-tests comparing diagonal and off-diagonal values. In each of these cases, *r* values were transformed to a normal distribution using the inverse hyperbolic tangent (Fisher's z transform). ***P < .001.**

categories. For line drawings, we also observe a second wave of category-specific signals over the P300 component (~300 msec after stimulus onset). This later-stage wave of activity may reflect decision-making processing attributed to resolving task demands.

Our ERP results indicate that the structure preserved in line drawings is sufficient to discriminate neural signals for different categories of scenes. Differences in the discrimination of colour photographs and line drawings for global- and basic-level categorizations suggest that the colour and texture information present in colour photographs may influence category selectivity in the brain. One possible explanation for these results may relate to our understanding of the visual world: structure may form a primary basis for recognition, and surface information may contribute to recognition by filling in details not readily available from structure alone. To explore the time course of these contributions, we sought to determine to what extent and when in time neural activity patterns generalize between colour photographs and line drawings. When analysing the time course of the cross-image type analysis, we again rely on the 1000 iterations of the bootstrap in order to derive robust information about the earliest time point.

3.3. *Correlations across colour photographs and line drawings*

To answer this question, we examined similarities between scene categories in colour photographs and line drawings at

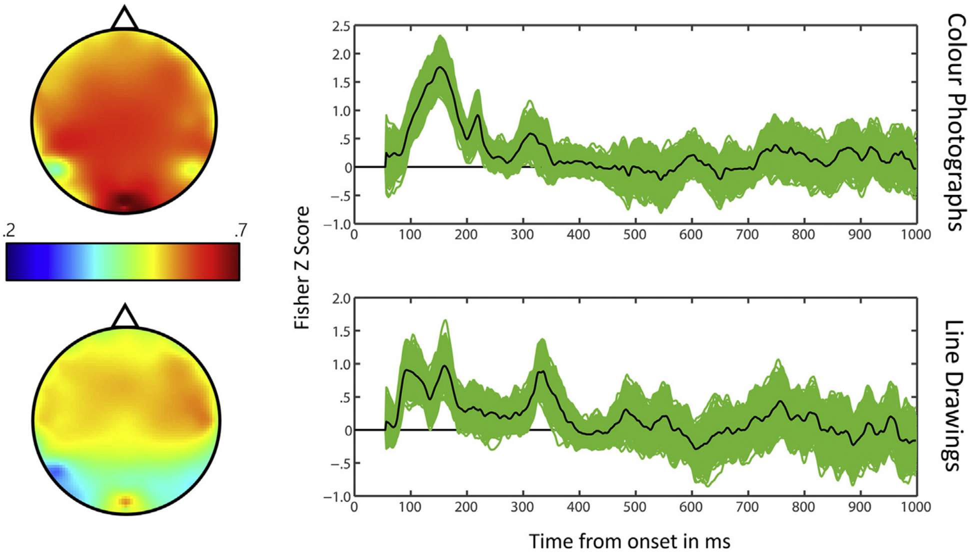

Fig. 7 – Grand-averaged discriminations (N = 16) of scene categories (diagonal vs. off-diagonal) for colour photographs (above) and line drawings (below) in centro-occipital (Oz) cortex. Green traces depict the category signal over time for individual iterations of subsampling, and the mean category signal over all 1000 iterations is the solid, black trace. Scalp plots (left) depict the topology of the category signal derived from the entire ERP at each electrode. These plots show the differences in within-category versus between-category Fisher-z corrected correlations. The colour bar presents the basic category signal; the difference between the average correlation on the diagonal and on the off-diagonal scene ERP correlation matrix.

Oz. We first created similarity matrices for all scene categories *across* colour photographs and line drawings (Fig. 8A) and examined whether diagonal entries (e.g., the similarity between colour photographs of beaches and line drawings of beaches, and vice-versa) would differentiate from off-diagonal entries (e.g., the similarity between colour photographs of beaches and line drawings of highways, and vice versa) across the full time course of scene perception, by converting *r* scores to Fisher Z values and subtracting the mean Fisher Z value of off-diagonal elements from on-diagonal elements. Even though data from line drawings and colour photographs are completely separate, we nevertheless performed a similar bootstrap analysis as before by randomly dividing the data into two halves for 1000 iterations. This approach provides us with a better estimate of the robustness of our findings than would be the case with a single correlation matrix.

507 of our 1000 iterations showed at least one time point with greater similarity within than between scene categories, across stimulus type. Examination of this time course showed a median significant category signal in Oz at 338 msec, *SD* = 168 msec. Given that the latency of this signal is considerably later than within-image type signals reported earlier, it is presumably related to feedback from higher-level brain regions following an initial feedforward perceptual encoding stage (Vanrullen & Thorpe, 2001). These higher-level, later-stage signals are likely to encode higher-order scene properties, which, unlike low-level features, transfer between line drawings and colour photographs (Fig. 8B). Later stages of scene processing may represent the abstract content of a scene and may be tolerant to low-level image changes (Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011). Given that this signal exhibited considerably more variability, however, this result should be interpreted with caution.

3.4. *Correlations with low-level image features*

To assess to what extent decoding from the EEG signal was driven by low-level image features, we computed a profile of orientation features for the stimulus images. To this end, we convolved the images with oriented Gabor filters at four orientations at four scales, as implemented in the orientation pyramids in the SaliencyToolbox (Walther & Koch, 2006). Filter responses were then averaged within a 3 × 3 grid, resulting in a feature vector of 144 elements for each image (4 orientations × 4 scales × 9 grid cells). We computed Pearson correlations of the feature vectors of all pairs of images, separately for photographs and line drawings, and averaged the Fisher z-transformed correlations for all pairings of categories, e.g., averaging all pairs of individual forest and highway images for the (forest, highway) entry. This procedure resulted in a symmetric 6 × 6 category similarity matrix.

For diagonal cells, we excluded the correlation of an image with itself, which is one by default and would artificially inflate the diagonal entries. We then calculated the correlation between off-diagonal cells in the image feature-based similarity matrices and the off-diagonal cells in the same ERP-based scene similarity matrices that were used to calculate the basic category signal, separately for colour photographs and line drawings. This allowed us to estimate the degree to which scene similarity in ERPs was related to

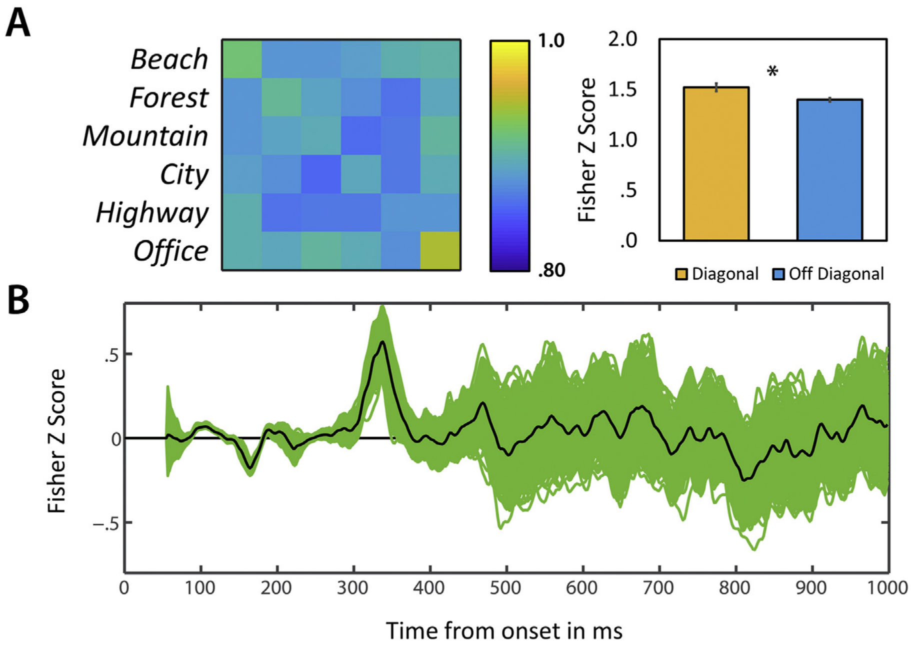

Fig. 8 – (A) Similarity matrix showing the grand-averaged cross-decoding correlations of each scene category for all participants (N = 16) averaged across time (0–1000 msec) in centro-occipital (Oz) cortex (left), and an independent samples t-test comparing diagonal and off-diagonal values where *r* values were transformed to a normal distribution using the inverse hyperbolic tangent (Fisher's z transform) (right) (B) Grand-averaged cross-decoding discriminations (N = 16) of scene categories (diagonal vs. off-diagonal) in centro-occipital (Oz) cortex. Green traces depict the category signal over time for individual iterations of subsampling, and the mean category signal over all 1000 iterations is the solid, black trace. *P < .05

scene similarity based on image features. The similarity matrices for photographs and line drawings are shown in Fig. 9. For Oz, both CP and LD had a median of 128.91 msec as the first point where off-diagonals in the EEG matrix correlated with the off diagonals in the image feature category matrix ($p < .05$).

3.5. *Exploratory analyses*

In addition to the hypothesis-driven analysis of the data recorded from the occipital site of interest (Oz), we also performed the same analysis with an exploratory-driven approach for all other electrode sites. The data are summarized in the Supplemental Information. Supplemental Figure 1 shows the earliest time points of first significance for scene categories in colour photographs, line drawings, and across colour photographs and line drawings for the basic-level category signal. Supplemental Table 1 shows the earliest time points for scene content (manmade versus natural) and boundary (open versus closed) discriminations for colour photographs and line drawings across the scalp. For colour photographs, we find discriminatory signals for both types of global scene properties even earlier in frontal than occipital electrodes. It is not uncommon to find early activity in frontal sites, and there is evidence to suggest that frontal cortex may be directly involved with stimulus processing (Imamoglu, Heinzle, Imfeld, & Haynes, 2014). Yet it should also be noted that, since the human head is a volume conductor, brain activity recorded across the scalp does not necessarily imply that this activity was generated at the electrode site. Nevertheless, this phenomenon will require further study in the future. Indeed, the scalp distribution of the time-course of ERP correlations with low-level features show a similarly widespread distribution, suggesting the signal is visually-driven.

The earliest time points for discriminating basic-level scene categories are listed in Supplemental Table 2. For both colour photographs and line drawings we find the earliest times at Oz and adjacent electrode sites. Correlations between colour photographs and line drawings appear to be successful earliest and most robustly at the centro-parietal electrodes (191 msec), which could potentially reflect later-stage processing in higher-level cortex. Importantly, the exploratory analysis of discriminatory signals across the scalp was not used for selecting the site of interest (Oz) for the present study, yet the results of the exploratory analysis provide a post-hoc validation of the selection of Oz as the recording site of interest. Supplemental Table 3 shows the earliest time points for significant correlations between ERP-based similarity matrices and feature-based similarity matrices for all electrode sites.

3.6. *Behavioural accuracy*

Participants performed well on the memorization test with an average accuracy of 68.5% ± 15.5%, confirming their attention to the task. To further analyse behavioural accuracy for recognition performance during the test phase, we conducted a two (stimulus type: colour photographs versus line

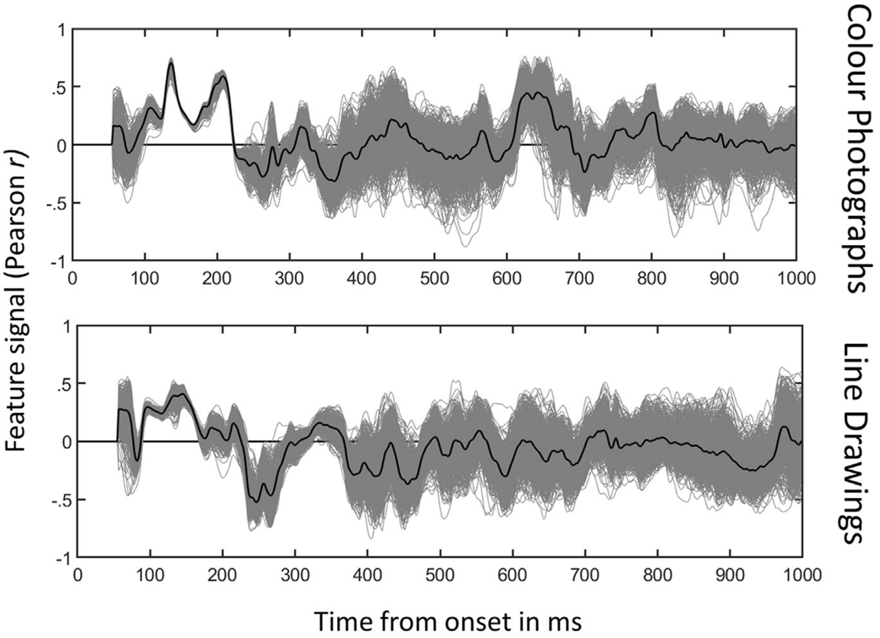

Fig. 9 – Scene image feature correlations (N = 16) between off-diagonal cells in the image feature-based similarity matrices and the off-diagonal cells in the ERP-based similarity matrices in centro-occipital (Oz) cortex. Grey traces depict the category signal over time for individual iterations of subsampling, and the mean category signal over all 1000 iterations is the solid, black trace.

drawings) by six (category: beach, city, forest, mountain, highway, office) repeated-measures ANOVA. This analysis revealed main effects of stimulus type ($F_{1,\ 15} = 43.27$, $P < .001$) and category ($F_{5,\ 75} = 7.82$, $P < .001$), but no interaction ($F_{5,\ 75} = 1.83$, $P < .118$). Overall, recognition rates for colour photographs ($M = 73.8\% \pm 14.2\%$) were higher than for line drawings ($M = 63.37\% \pm 15.0\%$). Response accuracy for all categories can be seen in Table 1. To explore these results in more detail, we conducted paired-samples *t*-tests (two-tailed) to examine differences across scene content (natural versus manmade) and boundary (open versus closed) collapsed across stimuli type (colour photographs and line drawings). These results revealed significantly higher recognition accuracy for manmade scenes compared with natural scenes ($t_{15} = 4.62$, $P < .001$), but no significant differences between open versus closed scenes ($t_{15} = .10$, $P = .925$). Given that closed scenes elicited overall larger response amplitudes compared with open scenes, despite no behavioural differences in recognition performance between these global properties, however, it is unlikely that task difficulty alone is driving neural differences across scene categories in the present study. While manmade scenes are defined by the *presence* of manmade artifacts, natural scenes are defined by their *absence*. For instance, adding a building to a forest turns the scene into a manmade scene. Adding a tree to a city, on the other hand, does *not* turn the city into a natural scene. It is therefore possible that manmade scenes are made intrinsically more memorable because they are defined by the presence of manmade object information within a scene. In fact, these results are consistent with previous work showing that natural scenes tend to be less memorable than manmade scenes (Isola, Xiao, Parikh, Torralba, & Oliva, 2014).

4. Discussion

Our findings indicate that scene categories can be discriminated from neural response amplitudes within a fraction of a second within the human visual system. Notably, both global scene properties (scene content, spatial boundary), and basic-level categories (beach, forest, mountain, city, highway, office) emerged from neural activity within the first 100 msec of perception in both line drawings and colour photographs. These findings provide evidence that information that can be used to make fundamental characterizations of natural scenes required for the perception of the visual environment is available even earlier than previously believed. Given the early onset of these categorizations, and the similarity of this neural activity with feature-based discriminations, it is likely that the categorizations observed here are driven by low-level

Table 1 – Behavioural recognition performance in the old/new memory task.

	Beaches	City	Forest	Highway	Mountain	Office
Colour Photographs	72.40 ± 2.82	79.69 ± 3.22	73.96 ± 1.84	73.44 ± 4.18	58.85 ± 2.69	85.38 ± 2.93
Line Drawings	65.15 ± 4.09	66.15 ± 3.99	57.81 ± 3.91	67.19 ± 3.69	56.25 ± 2.99	67.19 ± 3.18

All values represent mean (percent correct) ± SEM, chance = 50%.

image features. Low-level image statistics are highly correlated with scene scale and scene category (Torralba & Oliva, 2003), and may form a powerful basis for perception, as they capture the holistic and diagnostic structure of a scene (Oliva & Torralba, 2006). The findings presented here suggest that these statistics may form an efficient foundation for scene categorization in the human brain, and thus support evidence that scene information may be computed by extracting diagnostic image statistics from pooled responses in early visual cortex (Groen et al., 2013). Critically, these properties emerged from neural activity during an orthogonal memorization task, supporting previous suggestions that these properties form a concrete and fundamental basis to our understanding of the visual world and may be automatically extracted (Oliva & Torralba, 2006).

Overall, the time course of categorization over occipital cortex was similar for line drawings and colour photographs for early-stage recognition, which may highlight the role of scene structure, which is preserved in line drawings, as sufficient for both global scene properties and basic category levels. Notably, the structure preserved in line drawings, and the subsequent ability of the visual system to process both global scene properties and basic category levels in line drawings suggests that structure may be a fundamental property of scene perception. This suggestion is supported with the emergence of an earlier category signal for scene content in line drawings than for colour photographs, which may point to the importance of structural properties in the discrimination of natural versus manmade scenes. In further support of this conclusion, recent evidence shows that contour junctions underlie neural representations of scene categories (Choo & Walther, 2016), and evidence has linked deficits in scene perception from topographical disorientation to a primary reliance on structural properties (Robin et al., 2017). Together with the present findings, this highlights an integral role for structure in perception.

Yet texture and colour information also have a role in mediating scene recognition (Castelhano et al., 2008; Goffaux et al., 2005; Oliva & Schyns, 2000; Renninger & Malik, 2004; Steeves et al., 2004), and may interact across object and scene perception (Lowe, Ferber, & Cant, 2015). Our results reflect a role for surface information: For colour photographs, but not for line drawings, category discrimination for global scene properties extended over the later-stage P2 component. One explanation to account for these findings is that the processing of structure may be resolved earlier in time than surface information, and that surface information may provide additional context to further visual understanding. This explanation is consistent with previous ERP evidence suggesting edge-based information is processed with higher priority than surface information within the visual stream (Fu et al., 2016). Scene categorization is also influenced by diagnostically driven information (Lowe, Gallivan, Ferber, & Cant, 2016; Malcolm, Nuthmann, & Schyns, 2014), and these visual features may therefore shape scene categorization in distinct ways. We did not observe significant correlations across colour photographs and line drawings within the first 250 msec of activity over our occipital site of interest, which indicates that the human brain uses low-level image statistics that do not generalize well between line drawing and colour photographs for initial, first-wave scene processing.

The present findings also provide insight into a heated debate surrounding the temporal relationship between global scene properties, which represent the meaning of a scene, and basic-level categories, which represent the most common category descriptors. This debate concerns the hierarchical nature of scene processing, and which distinction (basic or global) emerges first in the visual processing stream. Behavioural evidence has suggested that global properties of a scene (e.g., scene content) may emerge prior to even basic-level distinctions (Greene & Oliva, 2009a, 2009b; Kadar & Ben-Shahar, 2012; Loschky & Larson, 2010; Sun et al., 2016). In contrast, some evidence suggests that basic-level distinctions emerge prior to global categorizations (Rosch et al., 1976; Tversky & Hemenway, 1983). Our results suggest these properties may be discriminated from neural activity within a similar time window within the visual processing stream: Both global scene properties and basic-level categories could be differentiated from brain activity within the first 100 msec of scene processing over early visual areas for both colour photographs and line drawings. Neural markers for these distinctions in the present study therefore support behavioural evidence suggesting that these properties may require the same amount of information in recognition (Fei-Fei et al., 2007). It is also important to note that our results provide evidence that the information contained within early neural signatures can be used to discriminate between scene categories, yet behavioural performance may differ from this temporal scale. For instance, while the information used to distinguish between scene categories may be present as early as 100 msec, human performance and recognition may be linked to task context and observer goals. Future research should therefore examine the extent to which task context influences how these different properties emerge during behavioural performance.

In summary, our results show that the separation of scene-related information occurs within the first 100 msec of activity in the human brain, suggesting that an efficient neural network is able to discern information from the environment in only an instant. Distinctions of global scene properties, such as content and layout, emerge within a similar time frame to basic-level distinctions, and these distinctions may be influenced by different visual information (e.g., line drawings versus colour photographs). These findings highlight the fundamental basis of categorization for the purposes of perceiving and understanding our visual environment, and the conjunctive roles of structure and surface information underlying neural representations of scene perception.

Acknowledgements

This study was supported by an Ontario Graduate Scholarship (OGS) to M.X.L, an NSERC grant to J.R., a Canadian Institutes of Health Research (CIHR) grant (#106436) to S.F, and an NSERC grant (#498390) to D.B.W.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.cortex.2018.06.006.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

Banno, H., & Saiki, J. (2015). The processing speed of scene categorization at multiple levels of description: The superordinate advantage revisited. *Perception, 44*(3), 269–288.

Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., & Herrmann, C. S. (2004). Size matters: Effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clinical Neurophysiology, 115*(8), 1810–1820.

Calvo, M. G., & Beltrán, D. (2014). Brain lateralization of holistic versus analytic processing of emotional facial expressions. *NeuroImage, 92*, 237–247.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of colour on the perception of scene gist. *Journal of Experimental Psychology. Human Perception and Performance, 34*(3), 660.

Choo, H., & Walther, D. B. (2016). Contour junctions underlie neural representations of scene categories in high-level human visual cortex. *Neuroimage, 135*, 32–44.

Cichy, R. M., Khosla, A., Pantazis, D., & Oliva, A. (2017). Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *NeuroImage, 153*, 346–358.

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *Journal of Neuroscience, 31*(31), 11305–11312.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598–601.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1), 10–10.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a, 4*(12), 2379–2394.

Fu, Q., Liu, Y. J., Dienes, Z., Wu, J., Chen, W., & Fu, X. (2016). The role of edge-based and surface-based information in natural scene categorization: Evidence from behavior and event-related potentials. *Consciousness and Cognition, 43*, 152–166.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition, 12*(6), 878–892.

Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive psychology, 58*(2), 137–176.

Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20*(4), 464–472.

Groen, I. I., Ghebreab, S., Lamme, V. A., & Scholte, H. S. (2016). The time course of natural scene perception with reduced attention. *Journal of neurophysiology, 115*(2), 931–946.

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

Harel, A., Groen, I. I., Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The temporal dynamics of scene processing: A multifaceted EEG investigation. *Eneuro, 3*(5). ENEURO-0139.

Imamoglu, F., Heinzle, J., Imfeld, A., & Haynes, J. D. (2014). Activity in high-level brain regions reflects visibility of low-level stimuli. *NeuroImage, 102*, 688–694.

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(7), 1469–1482.

Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of vision, 12*(13), 16–16.

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high- level visual cortex: it's the spaces more than the places. *Journal of Neuroscience, 31*(20), 7322–7333.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis- connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*, 4.

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron, 62*(2), 281–290.

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition, 18*(4), 513–536.

Lowe, M. X., Ferber, S., & Cant, J. S. (2015). Processing context: Asymmetric interference of visual form and texture in object and scene interactions. *Vision Research, 117*, 34–40.

Lowe, M. X., Gallivan, J. P., Ferber, S., & Cant, J. S. (2016). Feature diagnosticity and task context shape activity in human scene-selective cortex. *NeuroImage, 125*, 681–692.

Lowe, M. X., Rajsic, J., Gallivan, J. P., Ferber, S., & Cant, J. S. (2017). Neural representation of geometry and surface properties in object and scene perception. *NeuroImage, 157*, 586–597.

Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2014). Beyond gist: Strategic and incremental information accumulation for scene categorization. *Psychological Science, 25*, 1087–1097.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colours mediate scene recognition. *Cognitive Psychology, 41*(2), 176–210.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience, 31*(4), 1333–1340.

Park, J., & Park, S. (2017). Conjoint representation of texture ensemble and location in the parahippocampal place area. *Journal of Neurophysiology, 117*(4), 1595–1607.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage, 45*(1), S199–S209.

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision research, 44*(19), 2301–2311.

Robin, J., Lowe, M. X., Pishdadian, S., Rivest, J., Cant, J. S., & Moscovitch, M. (2017). Selective scene perception deficits in a case of topographical disorientation. *Cortex, 92*, 70–80.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.

Steeves, J. K., Humphrey, G. K., Culham, J. C., Menon, R. S., Milner, A. D., & Goodale, M. A. (2004). Behavioral and neuroimaging evidence for a contribution of colour and texture information to scene classification in a patient with visual form agnosia. *Journal of Cognitive Neuroscience, 16*(6), 955–965.

Sun, Q., Ren, Y., Zheng, Y., Sun, M., & Zheng, Y. (2016). Superordinate level processing has priority over basic-level processing in scene gist recognition. *i-Perception, 7*(6), 2041669516681307.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems, 14*(3), 391–412.

Torralbo, A., Walther, D. B., Chai, B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2013). Good exemplars of natural scene categories elicit clearer patterns than bad exemplars but not greater BOLD activity. *PloS One, 8*(3), e58594.

Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology, 15*(1), 121–149.

Vanrullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of cognitive neuroscience, 13*(4), 454–461.

Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2: Single- cell study. *European Journal of Neuroscience, 11*(4), 1239–1255.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience, 29*(34), 10573–10581.

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences, 108*(23), 9661–9666.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395–1407.