Trouilloud, A., Kauffmann, L., Roux-Sibilon, A., Rossel, P., Boucart, M., Mermillod, M., & Peyrin, C. (2020). Rapid scene categorization: From coarse peripheral vision to fine central vision. Vision Research, 170, 60-72. https://doi.org/10.1016/j.visres.2020.02.008

Rapid scene categorization: From coarse peripheral vision to fine central vision

Audrey Trouilloud[1+], Louise Kauffmann[1,2+], Alexia Roux-Sibilon[1], Pauline Rossel[1],
Muriel Boucart[3], Martial Mermillod[1], Carole Peyrin[1*]

[+]These authors contributed equally

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000, Grenoble, France.
[2]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France.
[3]SCALab, University of Lille, Centre National de la Recherche Scientifique, Lille, France.

Short title: Peripheral vision advantage for scene categorization

***Corresponding Author:**
Carole Peyrin, PhD
Laboratoire de Psychologie et NeuroCognition (LPNC)
CNRS UMR 5105 - Université Grenoble Alpes
BSHM - 1251 Av Centrale CS40700
38058 Grenoble Cedex 9 - France
carole.peyrin@univ-grenoble-alpes.fr

Abstract

Studies on scene perception have shown that the rapid extraction of low spatial frequencies (LSF) allows a coarse parsing of the scene, prior to the analysis of high spatial frequencies (HSF) containing details. Many studies suggest that scene gist recognition can be achieved with only the low resolution of peripheral vision. Our study investigated the advantage of peripheral vision on central vision during a scene categorization task (indoor vs. outdoor). In Experiment 1, we used large scene photographs from which we built one central disk and four circular rings of different eccentricities. The central disk either contained or not an object semantically related to the scene category. Results showed better categorization performances for the peripheral rings, despite the presence of an object in central vision that was semantically related to the scene category that significantly improved categorization performances. In Experiment 2, the central disk and rings were assembled from Central to Peripheral vision (CtP sequence) or from Peripheral to Central vision (PtC sequence). Results revealed better performances for PtC than CtP sequences, except when no central object was present under rapid categorization constraints. As Experiment 3 suggested that the PtC advantage was not explained by a reduction of the visibility of the object in the central disk by the surrounding peripheral rings (CtP sequence), results are interpreted in the context of a predominant coarse-to-fine processing during scene categorization, with greater efficiency and utility of coarse peripheral vision relative to fine central vision during rapid scene categorization.

Keywords:

Scene perception; Peripheral vision; Central vision; Spatial frequencies; Coarse-to-Fine.

1. Introduction

We are able to categorize complex scenes very quickly despite their variations in luminance, contrast, colors, object content and arrangement, but also in terms of visual resolution. Indeed, visual resolution varies considerably across the visual field due to the non-homogeneous distribution of photoreceptors and ganglion cells throughout the retina. Cones and midget ganglion cells have small receptive fields and thus a high spatial resolution, while rods and parasol ganglion cells are characterized by larger receptive fields and thus lower spatial resolution. Critically, the density of the former is greatest in the fovea, while the density of the latter increases with retinal eccentricity (Curcio, Sloan, Kalina & Hendrickson 1990). Therefore, the central retina would mainly encode high spatial frequencies (HSF) while the peripheral retina would rather encode low spatial frequencies (LSF; Kauffmann, Ramanoël, & Peyrin, 2014).

Interestingly, mainy studies have shown a crucial role of spatial frequencies for scene perception (Bar, 2003, 2007; Bar et al., 2006; Hegdé, 2008; Kauffmann et al., 2014; Schyns & Oliva, 1994). According to these studies, the visual analysis of the scene begins with the parallel extraction of different basic features at different spatial frequencies and follows a predominant coarse-to-fine sequence of processing. LSF containing the coarse information (e.g., the global shape and spatial layout of the scene) are rapidly and predominantly sent through the fast magnocellular pathways to the occipital cortex. LSF then access high-order cortical areas of the dorsal stream (such as parietal and frontal areas) and of the ventral stream (occipito-temporal areas), in order to activate the more probable interpretations of the visual input. This rapid coarse analysis is then used to guide the later processing of HSF, containing the fine information (e.g., the edges and the object detail) and conveyed more slowly by the parvocellular pathways to the ventral stream (Bar, 2003; Kauffmann, Chauvin, Guyader & Peyrin, 2015; Kveraga, Avniel, Ghuman & Bar, 2007; Peyrin et al., 2010; Trapp & Bar, 2015). Recent studies directly examined whether a coarse-to-fine sequence was advantageous for a rapid scene categorization (Kauffmann, Chauvin, Guyader & Peyrin, 2015; Kauffmann et al., 2015). In these studies, the authors presented dynamic sequences composed of six filtered images of the same scene, assembled from LSF to HFS as in a coarse-to-fine (CtF) sequence, or from HSF to LSF as in a fine-to-coarse (FtC) sequence. Participants were faster to categorize CtF sequences than FtC sequences as indoor or outdoor scenes. This result confirms that the processing of LSF before HSF is advantageous for a rapid scene categorization.

Given that LSF signals would mainly come from the peripheral visual field, while HSF signals would mainly come from the central visual field, these past studies lead us to wonder about the relative contribution of peripheral vs. central processing for scene categorization. In agreement with a coarse-to-fine sequence of spatial frequency processing, does scene categorization also follow a peripheral-to-central processing? In support to this hypothesis, Boucart et al. (2013) have shown that scene categorization is still robust when scenes are presented at very large visual eccentricities (e.g., 70°). Importantly, psychophysical studies have highlighted the advantage of peripheral vision relative to central vision for scene categorization (Geuzebroek and van den Berg, 2018; Larson & Loschky, 2009; Loschky et al., 2019). Larson and Loschky (2009) originally used a "Window" and "Scotoma" paradigm. In the Window condition, participants viewed a scene through a circular window centered on the fovea. In the Scotoma condition, a circular area was hiding the central part of the scene and only the peripheral part was shown. In a control condition, the scene was presented entirely. Results showed that categorization performances were closer to the maximal performances (obtained in control condition) in the Scotoma condition (peripheral vision) than in the Window condition (central vision). These results therefore suggested that the low resolution of peripheral vision is more useful than the high resolution of central vision to categorize a scene (see however, Larson, Freeman, Ringer & Loschky, 2014). The importance of peripheral vision for natural scene perception has also been observed in pathologies. Roux-Sibilon et al. (2018) performed a study with open-angle glaucomatous patients in order to directly investigate the impact of a peripheral vision loss on scene categorization. This pathology is due to a progressive destruction of the ganglion cells and the optic nerve which mainly affect the peripheral retina. In this study, glaucomatous patients and healthy controls had to categorize scenes presented foveally. The results showed that patients had worse categorization performances than healthy participants, even though the stimuli appeared in their intact central visual field region. The integrity of peripheral retina therefore seems to be essential for scene categorization.

Overall, studies on spatial frequencies and central vs. peripheral visual field processing are consistent with the idea that the visual analysis of scenes follows a coarse/peripheral-to-fine/central sequence. The aim of the present study was to directly test this hypothesis using dynamic sequences to impose a Peripheral-to-Central (PtC) or Central-to-Peripheral (CtP) analysis of visual information. These sequences, inspired by the CtF and FtC sequences used by Kauffmann et al. (2015), were built from five versions of the same scene, revealing different parts of the visual field through circular rings at different eccentricities and assembled from the periphery to the center (PtC sequence) or from the center to the periphery (CtP sequence). In Experiment 1, we tested the categorization of each ring composing the dynamic scenes in the same way as the Larson and Loschky (2009) experiment. Participants had to categorize rings as indoor or outdoor scenes. We expected an advantage of peripheral on central vision on categorization performance. In this experiment, we also manipulated the content available in the central part of the scene. Indeed, it is possible that the advantage of peripheral vision observed by Larson and Loschky (2009) during scene categorization resulted from an absence of relevant/diagnostic information in the central portion of the scene. Manipulating the central semantic content of the scene thus allowed us to determine whether the contribution of central vision depends on the presence of relevant semantic content to categorize scenes but also whether the presence of such content is enough to suppress the advantage of peripheral vision. In our experiment, the central part of the scenes thus either contained an object semantically related to the scene category (e.g., kitchen utensils for an indoor scene, or a house for an outdoor scene; Figure 1) or no semantically relevant object (the object was removed from the picture; Figure 1). Critically, Larson et al. (2014) revealed that the relative contribution of the central and peripheral visual field changes over time. Actually, the authors observed that during the first 100 ms of processing, there was a superiority of central vision, the relative contribution of peripheral vision increasing thereafter. They interpreted their results according to a zoom-out hypothesis of covert attention. Attention is first focused on central vision and then rapidly expands outward. Experiment 1 was also designed to explore the effect of temporal constraints on the processing of central versus peripheral visual information by manipulating the presentation time of stimuli. Stimuli were presented either 100 ms or 33 ms. Based on the hypothesis of a predominant coarse-to-fine processing during scene categorization (Schyns & Oliva, 1994), we expected that the coarse information available in peripheral vision would be even more considered for very short presentation times. Alternatively, according to the zoom-out hypothesis (Larson et al., 2014), information in central vision should be prioritized for short presentation times.

In Experiment 2, we tested the categorization of PtC and CtP dynamic scenes. As in Experiment 1, stimuli composing the sequences activated approximately the same cortical surface on V1 and we manipulated the content available in the central part of the scene. We also manipulated the Exposure duration of the sequence. Sequences lasted either 165 ms (each ring within the sequence was presented for 33 ms as in the shortest exposure duration condition of Experiment 1) or 500 ms (each ring was presented for 100 ms as in the longest exposure duration condition of Experiment 1). We expected better categorization performance for PtC than CtP sequences. Experiment 3 was conducted as a control experiment in order to assess if a PtC advantage could be due to a metacontrast masking effect, i.e. a reduced visibility of stimuli presented in central vision by the following surrounding peripheral information.

2. Experiment 1

2.1. Method

2.1.1. Participants

Thirty-two undergraduate students of Psychology from University Grenoble Alpes (28 women, mean age: 20.63 ± 2.68) participated in the experiment. They were divided into two groups (N = 16 each) for which stimuli were presented for either 33 or 100 ms. The sample size for each group was chosen based on a power analysis with estimated effect size of 1.061 (effect size of the interaction between the Eccentricity and the Central framing of the scene in Experiment 1) to achieve power of .99 at an alpha level of .05. All were right handed, with normal or corrected-to-normal vision. The study was approved by the ethics committee of the University Grenoble Alpes (CER-Grenoble Alpes, COMUE University Grenoble Alpes, IRB00010290). All participants involved in the study gave their informed oral and written consent.

2.1.2. Stimuli

The stimuli were constructed from 40 color photographs from the Pixabay website (https://pixabay.com/fr/), a photo sharing site under CC0 (Creative Commons Zero) license. Half of the photographs represented indoor scenes (e.g., living room, kitchen, bathroom), and the other half represented outdoor scenes (e.g., cityscape, mountain, beach). In the context of rapid scene categorization, the choice of these categories was motivated by the fact that categorization at the superordinate-level has been found to precede basic-level distinctions during visual recognition (see for example Rousselet, Joubert & Fabre-Thorpe, 2005; Joubert, Rousselet, Fize and Fabre-Thorpe, 2007; Loschky & Larson, 2010; Kadar & Ben-Shahar, 2012).

For each scene, we constructed two square images of 1500 × 1500 pixels (Figure 1a) centered on an object[1] fitting in a square of 100 × 100 pixels. For example, for indoor scenes, the object could be a household appliance, dishes or kitchen utensils, a book, a clock, a lamp, a radio, a candle, a vase or a pair of spectacles. For outdoor scenes, the object could be a car, a boat, a house, a building, a tree, a plant, a mountain or a bridge. Scenes without a central object were built by removing objects in the central area of the image (~200 pixels wide) using Adobe Photoshop CS6 (Figure 1a). The tools used were Content Aware, Spot Healing Brush and Clone-Stamp. The Content Aware tool identifies similar details near the area selected and automatically replaces the removed object with details of the close environment. The Spot Healing Brush tool was used to complete the previous tool and clean more precisely the automatic correction defects. The Clone-Stamp tool is another technique used to remove an object. It allows to specifically select an area in a scene and to duplicate it on the object we want to remove. The size of the duplicated area and the area opacity can be defined in advance. In order to create a scene looking as natural as possible we have recreated shadows and light instead of the object with the high and low density tools. All images were converted to 256 gray levels by averaging the values from the three color channels at each pixel. We also equalized the stimuli luminance and contrast. All stimuli were normalized to obtain an average luminance of 0.5 and a RMS contrast (Root Mean Square) of 0.2 for pixel intensity values between 0 and 1 (i.e. an average luminance of 128 and an RMS contrast of 51 on a scale of 256 gray levels). Thus, we obtained 80 images (40 indoor scenes and 40 outdoor scenes) of size 1500 × 1500 pixels. Stimuli can be downloaded from https://osf.io/ke6np/.

We fixed the angular size of stimuli at ~30 ° of visual angle to cover both central vision and peripheral vision. Then, for each image, we built six stimuli that revealed different parts of the scene. The first stimulus (Whole disk) revealed the scene through a large disk with a radius of 15° of visual angle (i.e. 750 pixels) on a gray background of 0.5 average luminance. The other stimuli (CD, R1, R2, R3, R4; Figure 1b) were built using an equation empirically derived from retinotopic measurement (Wu et al., 2012; for a similar procedure, see Geuzebroek & van den Berg, 2018): . In this equation, *y* is the integrated V1 surface area in square millimeters, representing the activated cortical surface, and *r inner* and *r outer* are the inner and outer radii in degrees, respectively. This progression was preferred to better account for the properties of the visual system. Indeed, information from central and peripheral visual fields does not project as it is on the visual cortex, but undergoes a deformation so that the central information is over-represented at cortical level, relative to peripheral information (i.e. cortical magnification; Daniel & Whitteridge, 1961). In our study, the surface of the image revealed by the rings increased with their eccentricity in order to take into account these transformations at cortical level. Based on this equation, we created five stimuli of different eccentricities but with the same surface of activation on V1 (20%, i.e. ~ 136 mm^3). The CD stimuli revealed the scene through a small disk of only 1.66° radius of visual angle (i.e. 83 pixels). For R1, adjacent to the small central disk, the inner and outer edges were respectively fixed at 1.66° visual angle (83 pixels) and 4.10° visual angle (205 pixels). For R2, adjacent to the first ring, the inner and outer edges were respectively fixed at 4.10° visual angle (205 pixels) and 7.10° visual angle (355 pixels). For R3, adjacent to the second ring, the inner and outer edges were respectively fixed at 7.10° of visual angle (355 pixels) and 10.70° of visual angle (535 pixels). Finally, for R4, adjacent to the third ring, the inner and outer edges were respectively fixed at 10.70° visual angle (535 pixels) and 15° visual angle (750 pixels). Thus, the Whole disk was the sum or superposition of all other stimuli. Stimuli were presented on a gray background of 0.5 average luminance.

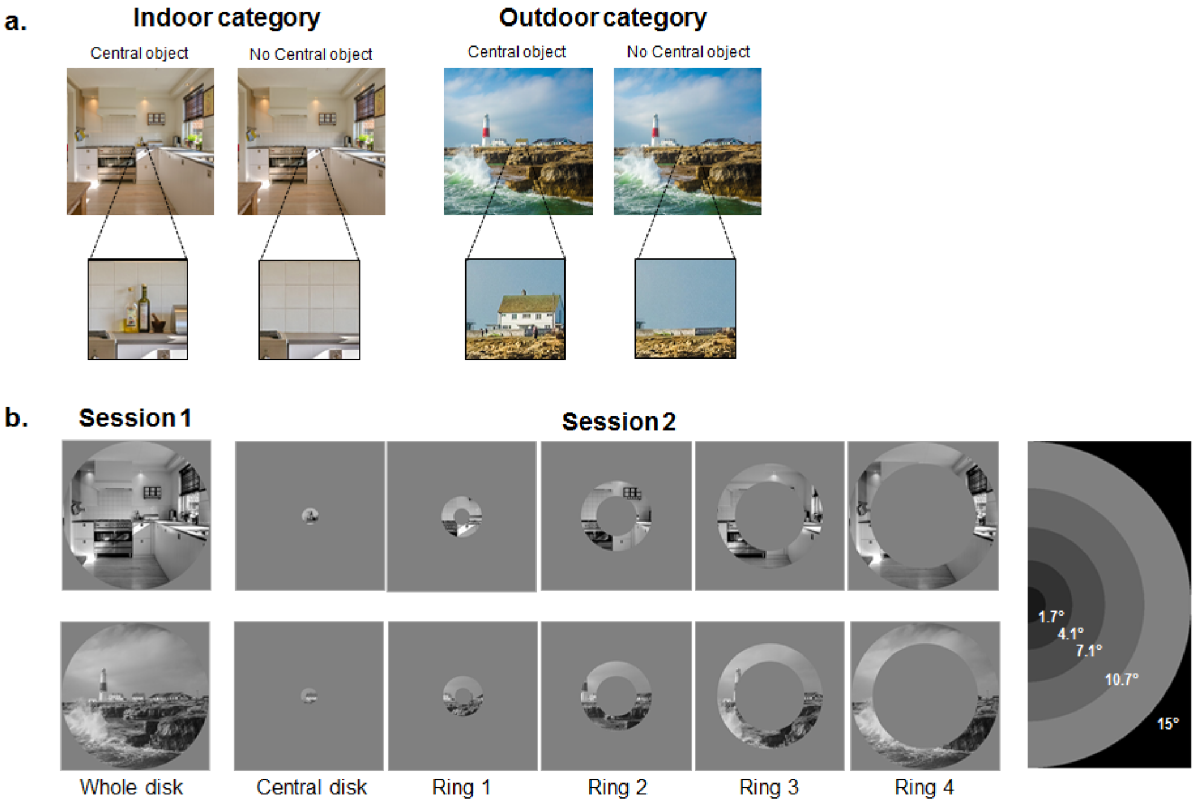

Figure 1. (a) Example of images used in Experiment 1: an indoor and an outdoor scene of 1500 × 1500 pixels presented with or without a relevant object in central vision. (b) Examples of stimuli used in Session 1 (Whole disk) and Session 2 (Central disk, Ring 1, Ring 2, Ring 3, and Ring 4). The right image illustrates the eccentricities (in degree of visual angle) of the rings' inner and outer edges.

2.1.3. Procedure

Stimuli were displayed using E-Prime software (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) on a 30' monitor (DELL ULTRASHARP), with a resolution of 2560 ×

1600 pixels and a refreshing rate of 60 Hz. The viewing distance was set at 70 cm in order to respect the stimuli angular size of 30° visual angle. Participants were divided into two groups (N = 16 in each group) for which stimuli were presented for either 33 ms or 100 ms. In each group, all participants performed two sessions. In the first control session, only the Whole disk stimuli were displayed. This session allowed us to assess if participants were able to accurately categorize the scenes, but also to test the effect of the presence of a relevant object in central vision when categorizing a large scene. In the second session, participants were presented with parts of the scenes (CD, R1, R2, R3, and R4).

For each session, a trial began with a central black fixation point presented for 500 ms on a 0.5 luminance background, followed by a stimulus (Session 1: Whole disk; Session 2: randomly CD, R1, R2, R3, or R4) for 33 or 100 ms on a 0.5 luminance background and then, by a gray screen (0.5 luminance) of 1900 ms during which participants could respond. Participants were asked to categorize the scene as either an indoor scene or an outdoor scene. Participants had to give their answer as rapidly and accurately as possible. They were instructed to to press the corresponding response key with the middle finger and the forefinger of the right hand. Keys were counterbalanced across participants. Session 1 included 80 trials (20 indoor scenes and 20 outdoor scenes with a central object, 20 indoor scenes, and 20 outdoor scenes without a central object) and lasted 3 min and 10 sec for the 33 ms Exposure duration condition, and 3 min 30 for the 100 ms Exposure duration condition. Session 2 included 400 trials (for each type of stimulus, i.e. CD, R1, R2, R3, R4, 20 indoor scenes and 20 outdoor scenes with a central object, 20 indoor scenes and 20 outdoor scenes without a central object) and lasted 16 min 20 sec for the 33 ms Exposure duration condition, and 16 min 60 sec for the 100 ms Exposure duration condition, including breaks every 80 trials. For each trial, response accuracy and response time (in ms) were recorded. Before each experimental session, participants performed a training session (10 trials) using stimuli that were not included in the main experiment.

2.2. Results

Results are shown in Figure 2. In each session, two repeated measures ANOVAs were performed on mean error rates (mER, in %) and mean correct response times (mRT, in ms). In order to reduce an effect of extreme RT values, RTs were trimmed for each participant's correct response of each condition. We removed responses less than and greater than three times the interquartile interval (0.15% of the trials for Session 1 and 0.49% of the trials for Session 2).

ANOVAs were performed using Statistica 13.3 software (Statsoft, Tulsa, USA). Effect sizes were estimated by calculating the partial eta-squared (η^2). The alpha level of tests was set at 0.05.

In Session 1 (Whole disk), mER was very low (under 5%). The ANOVAs included the Exposure Duration of each stimulus (100 ms and 33 ms) as between-subject factor, the presence of a Central object (Object and No object) and the Category of scenes (Outdoor and Indoor) as within subject factors. There was no effect of the Central object, neither for mER (Mean ± Standard deviation; Object: 4.53 ± 0.96%; No object: 4.53 ± 1.16%; $F(1,29) < 1$, $p = 1$), nor for mRT (Object: 615 ± 20 ms; No object: 622 ± 23 ms; $F(1,30) = 1.22$, $p = .277$). The main effect of the Exposure duration was not significant, neither for mER ($F(1,30) < 1$, $p = .903$), nor for mRT ($F(1,30) < 1$, $p = .486$) and this factor did not interact with the Central object (mER: $F(1,30) < 1$, $p = .844$; mRT: $F(1,30) < 1$, $p = .862$). Furthermore, the Category of the scene did not interact with the presence of a Central object, neither for mER ($F(1,30) < 1$, $p = .461$), nor mRT ($F(1,30) = 2.34$, $p = .136$). These results suggest that the presence of a central object semantically related to the scene category does not improve categorization performances, whatever the scenes exposure duration, although this interpretation can be limited by a potential ceiling effect.

In Session 2, the ANOVAs included the Exposure Duration of each stimulus (33 ms and 100 ms) as between-subject factor, the Eccentricity of visual information (CD, R1, R2, R3, R4), the presence of a Central object (Object and No object) and the Category of scenes (Outdoor and Indoor) as within subject factors. The ANOVA performed on mER revealed a main effect of Eccentricity ($F(4,120) = 65.79$, $p < .001$, ηp2 = .69). The linear and quadratic contrast tests performed between the different eccentricities were significant (Linear contrast: $F(1,30) = 84.09$, $p < .001$; Quadratic contrast: $F(1,30) = 114.24$, $p < .001$; CD: 15.74 ± 2.13%; R1: 5.51 ± 1.32%; R2: 5.00 ± 1.47%; R3: 5.00 ± 1,.44%; R4: 6.42 ± 1.89%). We subsequently performed planned comparisons between each stimulus in order to identify which ring leads to the best performance. These analysis showed a significant difference between CD and R1 ($F(1,30) = 146.13$, $p < .001$), but no significant difference between R1 and R2 ($F(1,30) = 0.51$, $p = .479$), neither for R2 and R3 ($F(1,30) < 1$, $p = .997$) or R3 and R4 ($F(1,30) = 3.268$, $p = .081$). These results suggest that errors exponentially decreased as eccentricity increased. Importantly, the main effect of the Central object was significant ($F(1,30) = 19.45$, $p < .001$, ηp2 = .39; Object: 6.49 ± 2.20%; No object: 8.59 ± 2.27%) and the presence of the Central object of the scene interacted with Eccentricity ($F(4,120) = 25.04$, $p < .001$, ηp2 = .45). The linear and quadratic contrast tests performed between the different eccentricities were significant both when there was a central object (Linear contrast: *F*(1,30) = 17.60, *p* < .001; Quadratic contrast: *F*(1,30) = 16.16, *p* < .001; CD: 10.39 ± 1.35%; R1: 5.78 ± 1.08%; R2: 5.39 ± 1.19%; R3: 4.61 ± 1.12%; R4: 6.62 ± 1.65%) and when there was no central object (Linear contrast: *F*(1,30) = 79.68, *p* < .001; Quadratic contrast: *F*(1,30) = 124.77, *p* < .001; CD: 21.09 ± 2.19%; R1: 5.23 ± 1.05%; R2: 4.61 ± 1.10%; R3: 5.39 ± 1,23%; R4: 6.60 ± 1.41%). We then performed planned comparisons between each stimulus in order to identify which ring leads to the best performance. When there was a central object, these analyses showed a significant difference between CD and R1 (*F*(1,30) = 23.28, *p* < .001), but no significant difference between R1 and R2 (*F*(1,30) < 1), neither for R2 and R3 (*F*(1,30) = 1.30, *p* = .264) or R3 and R4 (*F*(1,30) = 2.44, *p* = .129). When there was no central object, these analyses showed a significant difference between CD and R1 (*F*(1,30) = 163.05, *p* < .001), but no significant difference between R1 and R2 (*F*(1,30) < 1), neither for R2 and R3 (*F*(1,30) < 1) or R3 and R4 (*F*(1,30) = 1.21, p = .279). Thus, errors exponentially decreased as eccentricity increased irrespective of the presence of a central object in the scene. Critically, for CD, participants made less errors for categorizing a scene that contained an object in central vision (10.39 ± 1.35%; *F*(1,30) = 37.32, *p* < .001), compared to a scene that did not contain one (21.09 ± 2.19%). Yet, despite the presence of a central object that significantly improved the categorization performances, participants still made fewer errors when scene information was available in peripheral vision than in central vision. The main effect of the Exposure duration was not significant (*F*(1,30) = 1.33, *p* = .258) and this factor did not interact neither with the Eccentricity (*F*(4,120) = 1.42, *p* = .231), nor with the Eccentricity*Central object interaction (*F(*4,120) < 1, *p* = .542). Finally, the main effect of the Category was significant (*F*(1,30) = 18.85, *p* < .001, ηp2 = .39). Participants made less errors when categorizing outdoor scenes (5.93 ± 1.20%) than indoor scenes (9.14 ± 2.79%). But the category did not interact with the others manipulated factors (Eccentricity x Category: *F*(4,120) < 1; Eccentricity x Exposure duration x Category: *F*(4,120) < 1; Eccentricity x Central object x Category: *F*(4,120) = 1.27; Eccentricity x Central object x Exposure duration x Category F(1,24) < 1).

We also compared performances of each eccentricities of Session 2 (CD, R1, R2, R3, R4) to performances obtained in Session 1 (Whole disk). These comparisons showed that categorization performance of R4 and CD was significantly different to the categorization of the Whole disk (R4: *t*(32) = -2.41, *p* = .022; CD: *t*(32) = -12.60, *p* < .001). Performances on other eccentricities (R3, R2 and R1) were not significantly different (*ts*(32) < -1.62, *ps* > .114). These results suggest that scene categorization is disadvantaged when only the central visual information or the most peripheral visual information is available.

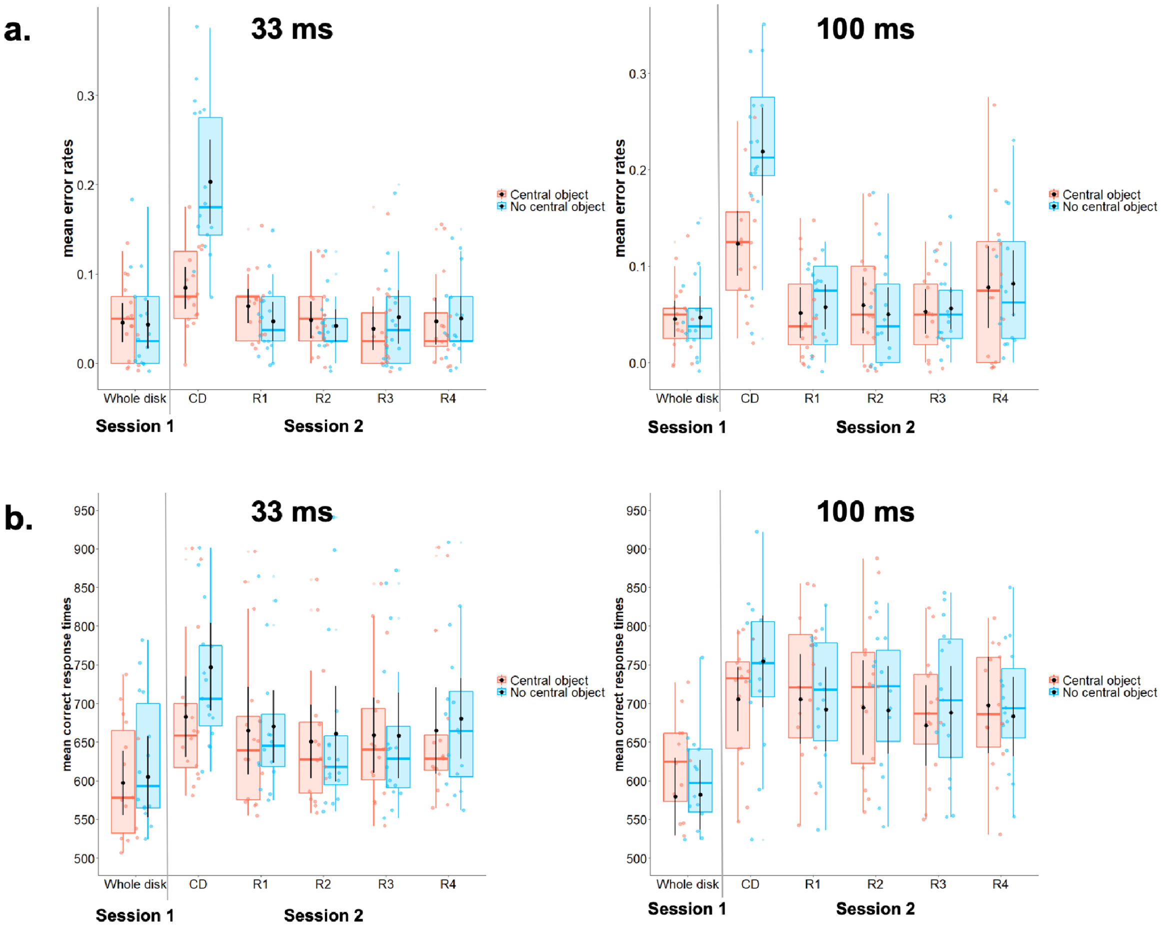

Figure 2. (a) Mean error rates in percentage and (b) mean correct response times in milliseconds in Experiment 1 during the categorization of the Whole Disk in Session 1 and each ring in Session 2 (CD, R1, R2, R3, R4), according to the exposure duration (33 ms and 100 ms) and the presence of a central object (Central object and No central object). Black dots and error bars indicate mean and standard error, respectively. Color dots are individual observations.

Concerning now the ANOVA performed on mRT, it revealed a main effect of Eccentricity ($F(4,120) = 29.32$, $p < .001$, ηp2 = .49). Similarly to the mER analysis, a linear and quadratic contrast test performed between the different eccentricities revealed both a linear and quadratic decrease of mRT as eccentricity increased (Linear contrast: $F(1,30) = 61.47$, $p < .001$: Quadratic contrast: $F(1,30) = 37.41$, $p < .001$; CD: 718 ± 32 ms; R1: 679 ± 33 ms; R2: 672 ± 35 ms; R3: 665 ± 33 ms; R4: 679 ± 34 ms). We performed planned comparisons between each stimulus in order to identify which ring leads to the best performance. These analysis showed a significant difference between CD and R1 (*F*(1,30) = 55.86, *p* < .001) and R3 and R4 (*F*(1,30) = 5.13, *p* = .031), but no significant difference between R1 and R2 (*F*(1,30) = 2.03, *p* = .164) or R2 and R3 (*F*(1,30) = 2.17, *p* = .151). These results suggest that mRT follows a U-shape curve. As for the mER analysis, the main effect of the Central object was significant (*F*(1,30) = 12.71, *p* = .001, ηp2 = .30). Participants categorized a scene that contained an object in central vision faster (677 ± 52 ms) than a scene that did not contain one (688 ± 52 ms). The presence of a Central object interacted with Eccentricity (*F*(4,120) = 10.32, *p* < .001, ηp2 = .26). The linearity and quadratic contrast tests performed between the different eccentricities were significant both when there was a central object (Linear contrast: *F*(1,30) = 13.71, *p* < .001; Quadratic contrast: *F*(1,30) = 11.05, *p* = .002; CD: 694 ± 21 ms; R1: 682 ± 26 ms; R2: 671 ± 25 ms; R3: 661 ± 22 ms; R4: 679 ± 26 ms) and when there was no central object (Linear contrast: *F*(1,30) = 71.37, *p* < .001; Quadratic contrast: *F*(1,30) = 43.01, *p* < .001; CD: 742 ± 25 ms; R1: 675 ± 22 ms; R2: 672 ± 26 ms; R3: 669 ± 25 ms; R4: 680 ± 23 ms). We then performed planned comparisons between each stimulus in order to identify which ring leads to the best performance. When there was a central object, these analysis showed a significant difference between R3 and R4 (*F*(1,30) = 4.32, *p* = .05), but no significant difference between CD and R1 (*F*(1,30) = 2.42, *p* = .130), neither for R1 and R2 (*F*(1,30) = 2.10, *p* = .157) or R2 and R3 (*F*(1,30) = 2.28, *p* = .141). When there was no central object, these analysis showed a significant difference between CD and R1 (*F*(1,30) = 60.80, *p* < .001), but no significant difference between R1 and R2 (*F*(1,30) < 1), neither for R2 and R3 (*F*(1,30) < 1) or R3 and R4 (*F*(1,30) = 1.87, *p* = .181). Critically, for CD, planned comparisons showed that participants categorized the scenes that contained an object in central vision (694 ± 21 ms; *F*(1,30) = 37.32, *p* < .001) faster than the scenes that did not contain one (742 ± 25 ms). The main effect of the Exposure duration was not significant (*F*(1,30) = 0.62, *p* = .438) and this factor did not interact neither with the Eccentricity (*F*(4,120) = 1.37, *p* = .248), nor with the Eccentricity*Central object interaction (*F*(4,120) = 2.16, *p* = .078).

Finally, the main effect of the Category was not significant (*F*(1,30) = 1.48, *p* = .232), but this factor interacted with the eccentricity (*F*(4,120) = 5.47, *p* < .001). Planned comparisons revealed that participants were faster for categorizing outdoor scenes than indoor scenes only for the R4 (*F*(1,30) = 8.53, *p* = .006; Indoor: 694 ± 26 ms; Outdoor: 664 ± 24 ms), but not for other eccentricities (CD: *F*(1,30) = 2.86, p = .101; R1: *F*(1,30) < 1; R2: *F*(1,30) < 1; R3: *F*(1,30) = 2.548, *p* = .121). Furthermore, the category did not interact with the others manipulated factors (Eccentricity x Exposure duration x Category: *F*(4,120) = 1.28, p = 0.28; Eccentricity

× Central object × Category: $F(4,120) < 1$; Eccentricity × Central object × Exposure duration × Category $F(1,24) < 1$).

As for the mER analysis, compared performances of each eccentricities of Session 2 (CD, R1, R2, R3, R4) to performances obtained in Session 1 (Whole disk). These comparisons showed that categorization performance of all eccentricities (R4, R3, R2, R1, CD) were significantly different from the categorization of the Whole disk both when there was no object and when there was one ($ts(32) > -4.17$, $ps < .001$). Thus, a partial presentation of a scene systematically lengthened response times.

2.3. Discussion

In Experiment 1, results showed that categorization performance improved as scene information was displayed in the peripheral vision (i.e. mean error rates and mean response times decreased). We also observed that error rates obtained in the near peripheral vision (Rings 1, 2, and 3 of Session 2) did not differ from those obtained for whole scenes (Whole disk in Session 1). These results, consistent with previous studies (Larson & Loschky, 2009, 2017; Larson et al., 2019), indicate that peripheral vision would play a more important role than central vision to quickly categorize a scene.

In Experiment 1, we also specifically tested whether the superiority of peripheral vision over central vision observed by Larson and Loschky (2009) could be explained by an absence of relevant information in central vision. Therefore, scenes contained either an object semantically related to the scene category in central vision (e.g., cup for indoor scenes and car for outdoor scenes), or no semantically relevant object. Results of the first Session (Whole scenes) revealed that the mere presence of a central object semantically related to the scene category does not improve scene categorization performance. Information in central vision − even relevant − would be less considered to categorize a scene when all information is available. In that case, the visual system would prioritize information from the periphery that would be enough to categorize scenes as indoor or outdoor. This hypothesis is consistent with the results obtained by Larson and Loschky (2009) showing that categorization performances are similar when the scene is entirely visible and when only peripheral information is available. Nevertheless, it can be noted that the error rates were extremely low in our experiment. This ceiling effect suggests that the categorization task was too easy in Session 1, increasing the difficulty to detect the expected effect. Indeed, in Session 2, the error rates obtained during the Central disk categorization were significantly greater than those obtained during the whole scene categorization (Whole disk in Session 1), and the presence of an object semantically related improved the categorization performance of the Central Disk. More importantly, in Session 2, we observed an interaction between the presence of a central object in the scene and the eccentricity of stimuli. Participants were better and faster for categorizing the Central Disk when an object was present. However, performances observed for the peripheral rings remained better than those observed for the Central Disk: Performance decreased as eccentricity increased irrespective of the presence of a central object in the scene. These results are consistent with our hypothesis, suggesting that reliance on central vision is partly based on the presence of specific visual contents, such as a relevant object to categorize. But they also support the results of Larson and Loschky (2009) by demonstrating superiority of peripheral vision over central vision regardless of their respective informational content.

Finally, contrary to what we expected, the exposure duration of the stimuli did not interact with their eccentricities. This result contrasts with the one observed by Larson et al. (2014). They rather suggest that visual information is automatically and very rapidly extracted from peripheral vision to scene gist recognition. The discrepancies between our results and Larson et al. (2014) may be due to differences in the experimental procedure. In this experiment, the authors used a “Window” and “Scotoma” paradigm. To control for confounding natural properties of central and peripheral vision (cortical magnification in favor of central vision vs. amount of viewable information in favor of peripheral vision), they defined stimuli using a critical radius, that is the radius that produces equal performance in both the window and scotoma conditions when stimuli were unmasked. In this experiment, the critical radius was 5.54°. Therefore, the window stimulus was a central disk of 5.54° radius of visual angle and the scotoma stimulus was a ring, with inner radius of 5.54° and an outer radius of 10.95°. Based on the equation empirically derived from retinotopic measurement (Wu et al., 2012), the cortical surface activated by the window and scotoma stimuli is estimated to 62% (344 mm^3) and 38% (213 mm^3), respectively. Therefore, the window stimulus is over-represented at the cortical level, relative to the scotoma stimulus. In addition, in Larson et al. (2014), the processing time of the stimulus was manipulated using visual masking. The stimulus was flashed (24 ms), followed by an interstimulus interval (0, 71, 165, 295, or 353 ms), and then by a mask (24 ms). Following the mask, there was a long duration blank screen (750 ms), and then a screen with a category label (e.g., beach) until the participants responded. More precisely, participants had to decide if the category label was congruent or not with the stimulus. Therefore, participants had to systematically wait for the category label screen before answering. Given the long time interval between the visual stimulation and the response, it might be possible that more attentional processes are involved, in contrast to our paradigm in which we privileged a response as soon as the stimulus appeared in order to assess more automatic processes.

Overall, results of Experiment 1 are consistent with a greater efficiency and utility of peripheral vision compared to central vision when categorizing visual scenes (as indoor or outdoor scenes), despite the low visual acuity inherent to peripheral vision. The scene categorization would not require a detailed analysis of various elements that compose it. Instead, the low resolution of peripheral vision would be sufficient for scene categorization (Boucart et al., 2013). We have connected these results to previous works indicating that visual information processing would follow a default coarse-to-fine (CtF) processing sequence (Kauffmann et al., 2015; Schyns & Oliva, 1994). In other words, we wanted to determine whether the rapid analysis of coarse information available in peripheral vision could allow a first categorization that would validate or not the slower details analysis in central vision. For this, we used dynamic scene sequences to impose a Peripheral-to-Central (PtC) or Central-to-Peripheral (CtP) analysis of visual information.

3. Experiment 2

In this experiment, we created dynamic sequences composed of five stimuli assembled from the periphery to the center (PtC) or from the center to the periphery (CtP), so that the whole scene was revealed successively through the five stimuli. As in Experiment 1, stimuli activated approximately the same cortical surface on V1 and we manipulated the content available in the central part of the scene. The sequence either contained - in the Central Disk - an object semantically related to the scene category (e.g., kitchen utensils for an indoor scene, or a house for an outdoor scene), or no semantically relevant object (the object was removed). We also manipulated the Exposure duration of the sequence. Sequences lasted either 165 ms (each ring within the sequence was presented for 33 ms as in the shortest exposure duration condition of Experiment 1) or 500 ms (each ring was presented for 100 ms as in the longest exposure duration condition of Experiment 1). We expected that PtC sequences would be categorized more rapidly and more accurately than CtP sequences in both exposure duration condition, and this, despite the presence of an object in central vision. In addition, based on the hypothesis of the rapid processing of LSF during scene categorization, we expected that the coarse information available in peripheral vision would be more considered when sequences lasted 165 ms, whereas the slower analysis of details in central vision would take over when sequences lasted 500 ms.

3.1. Method

3.1.1. Participants

Thirty-two undergraduate students of Psychology from University Grenoble Alpes (24 women, mean age: 19.66 ± 1.66) participated in the experiment. Inclusion criteria of participants were the same as in Experiment 1. The experiment was carried out within the same ethical framework as in Experiment 1.

3.1.2. Stimuli and Procedure

The stimuli that composed the sequences in Experiment 2 were exactly the same as those used in Session 2 of Experiment 1 (CD, R1, R2, R3, and R4, see Figure 3). Thus, we used 80 stimuli (40 indoor scenes and 40 outdoor scenes). All stimuli were normalized to obtain an average luminance of 0.5 and a RMS contrast (Root Mean Square) of 0.2 for pixel intensity values between 0 and 1 (i.e. an average luminance of 128 and an RMS contrast of 51 on a scale of 256 gray levels). For the PtC sequences, the five stimuli were assembled from the peripheral to the central ring: R4 was presented first, then R3, R2, R1, and finally CD. For the CtP sequences, the five stimuli were assembled from the center to the periphery: CD was presented first, then R1, R2, R3 and finally R4. Sequences lasted either 165 ms (each ring within the sequence was presented for 33 ms as in the shortest exposure duration condition of Experiment 2) or 500 ms (each ring was presented for 100 ms as in the longest exposure duration condition of Experiment 2). Figure 3 shows a schematic of a trial in each of the two sequence conditions (CtP vs. PtC).

Stimuli were displayed using E-Prime software (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) using the same 30’ monitor as in Experiment 1 with a viewing distance of 70 cm. All participants underwent one session and they were divided into two groups (N = 16 in each group) for which CtP and PtC sequences were presented for either 165 ms or 500 ms. The CtP and PtC sequences were randomly presented to the participants. A trial began with a central black fixation point presented for 500 ms on a 0.5 luminance background, followed by a sequence of five stimuli. Inter-sequences interval was a grey screen of 1900 ms during which participants could respond. Participants’ task was similar to Experiment 1 (i.e. indoor vs. outdoor categorization). The session included 160 trials (i.e. 20 stimuli x 2 Central framing conditions x 2 Categories x 2 Sequence orders). The experiment lasted about 7 minutes for the 165 ms Exposure duration condition, and about 8 min for the 500 ms Exposure duration condition, including breaks every 100 trials. For each trial, response accuracy and response time (in ms) were recorded. Before each experimental session, participants performed a training session (20 trials) using other stimuli.

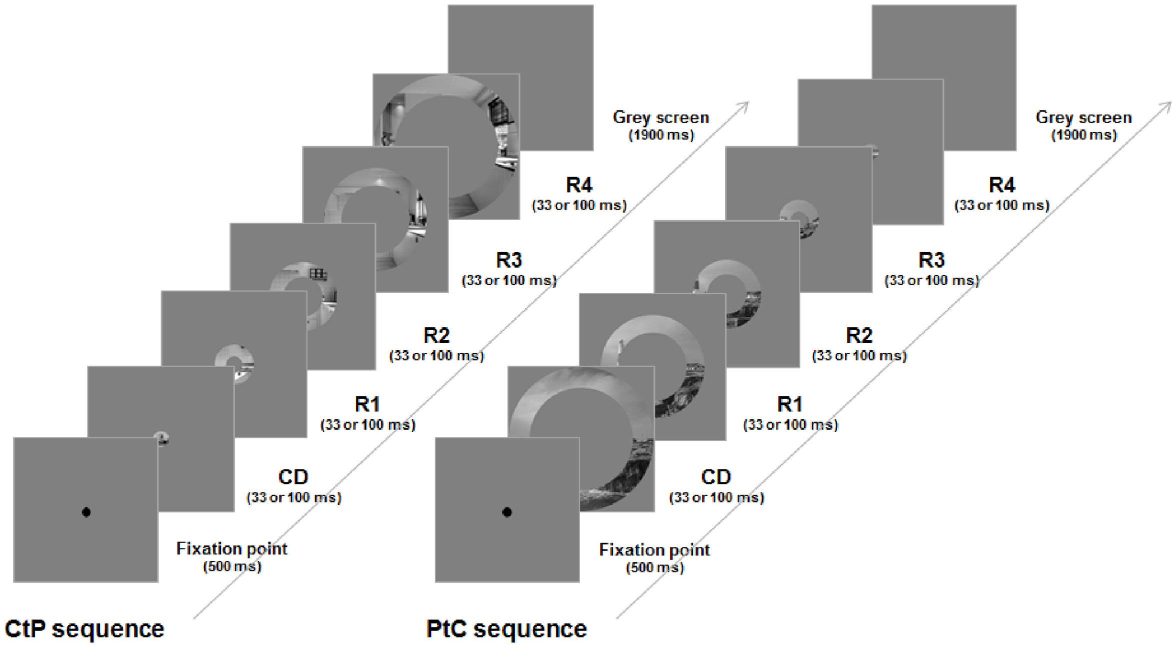

Figure 3. Trial schematic in Experiment 3. Central disk (CD), Ring 1 (R1), Ring 2 (R2), Ring 3 (R3) and Ring 4 (R4) are assembled from the center to the periphery (CtP sequences) or from the periphery to the center (PtC sequences). The sequences lasted either 160 ms (each stimulus was presented for 33 ms), or 500 ms (each stimulus was presented for 100 ms). Participants were requested to give their categorical answer as quickly and as accurately as possible, as soon as the first stimulus of the sequence appeared.

3.2. Results

Results are shown in Figure 4. Two repeated measures ANOVAs were performed on mean error rates (mER, in %) and mean correct response times (mRT, in ms). The ANOVAs included the Exposure Duration of the sequence (165 ms and 500 ms) as between-subject factor, and the Sequence (PtC and CtP), the presence of a Central object in the scene (Object and No object), and the Category of scenes (Outdoor and Indoor) as within subject factors. RTs were trimmed for each participant's correct response of each condition. We removed an average 0.25% of the trials. ANOVAs were performed using Statistica 13.3 software (Statsoft, Tulsa, USA). Effect sizes were estimated by calculating the partial eta-squared (η^2). The alpha level of tests was set at 0.05.

In this experiment, the mER was very low (under 6%). The ANOVA performed on mER did not show a significant effect of the Sequence (*F*(1,30) < 1; CtP: 5.27 ± 1.50%; PtC: 5.71 ± 1.64%). There was neither a main effect of the Exposure duration (*F*(1,30) = 3.16, *p* = .085), nor an interaction of this factor with the Sequence (Sequence × Exposure duration: *F*(1,30) < 1). Furthermore, there was neither a main effect of the Central object (*F*(1,30) = 1.56, *p* = .22), nor an interaction of this factor with the Sequence (*F*(1,30) = 3.09, *p* = .089), and these two factors did not interact with the Exposure duration (*F*(1,30) < 1). Finally, there was neither a main effect of Category (*F*(1,30) < 1), nor interaction between Category and the other manipulated factors (Sequence ´ Category: *F*(1,30) < 1; Sequence ´ Exposure duration ´ Category: *F*(1,30) < 1; Sequence ´ Central framing ´ Category: *F*(1,30) < 1; Sequence ´ Central framing ´ Exposure duration ´ Category: *F*(1,30) < 1).

The ANOVA performed on mRT revealed a main effect of the Sequence (*F*(1,30) = 65.88, *p* < .001, ηp2 = .69). Participants were faster for categorizing PtC sequences (613 ± 29 ms) than CtP sequences (654 ± 30 ms). The main effect of the Exposure duration was not significant (*F*(1,30) < 1), but this factor interacted with the Sequence (*F*(1,30) = 14.52, *p* < .001). Planned comparisons revealed that participants categorized PtC sequences faster than CtP sequences both when they were presented for 165 ms (PtC: 634 ± 41 ms; CtP: 656 ± 43 ms; *F*(1,30) = 9.27, *p* < .005) and for 500 ms (PtC: 592 ± 41 ms; CtP: 652 ± 43 ms; *F*(1,30) = 71.12, *p* < .001), the difference between the two sequences being greater for exposure duration of 500 ms than 165 ms. The main effect of the Central object was not significant (*F*(1,30) = 2.03, *p* < .164), and this factor did not interact with the Sequence (*F*(1,30) = 2.08, *p* = .162). However, the analysis revealed a significant interaction between the Sequence, the Exposure duration and the Central object (*F*(1,30) = 7.75, *p* = .009, ηp2 = .20). We thus tested the interaction between the Sequence and the Exposure duration for each condition of Central object separately. This interaction was significant when there was no central object (*F*(1,30) = 19.44, *p* < .001), but not when there was a central object (*F*(1,30) = 2.13, *p* =.154). More precisely, when there was a central object, participants categorized PtC sequences faster than CtP sequences irrespective of the exposure duration of sequences (for 500 ms: PtC: 597 ± 29 ms, CtP: 640 ± 33 ms; for 165 ms: PtC: 630 ± 29 ms, CtP: 657 ± 33 ms). However, when there was no central object participants categorized faster a PtC than a CtP sequence only when sequences lasted 500 ms (for 500 ms: PtC: 588 ± 30 ms, CtP: 663 ± 29 ms, *F*(1,30) = 63.37, *p*

< .001; for 165 ms: PtC: 638 ± 30 ms, CtP: 655 ± 29 ms, $F(1,30) = 2.98$, $p = .095$). These results suggest that the PtC advantage was reduced when there was no central object in the sequences lasting 165 ms.

Furthermore, there was no main effect of Category ($F(1,30) < 1$), but the interaction between the Category and the Sequence was significant ($F(1,30) = 4.85$, $p = .035$, ηp2 = .14). Planned comparisons showed that the difference between CtP and PtC sequences was significant both for Indoor category (CtP: 657 ± 23 ms; PtC: 610 ± 21 ms; $F(1,30) = 53.52$, $p < .001$) and Outdoor category (CtP: 651 ± 21 ms; PtC: 617 ± 21 ms; $F(1,30) = 46.84$, $p < .001$), the PtC advantage being greater for the outdoor scenes. Finally, the Category did not interact with the others manipulated factors (Exposure duration ´ Category: $F(1,30) < 1$; Sequence ´ Exposure duration ´ Category: $F(1,30) < 1$; Sequence ´ Central framing ´ Category: $F(1,30) < 1$; Sequence ´ Central framing ´ Exposure duration ´ Category: $F(1,30) = 1.77$, $p = .193$).

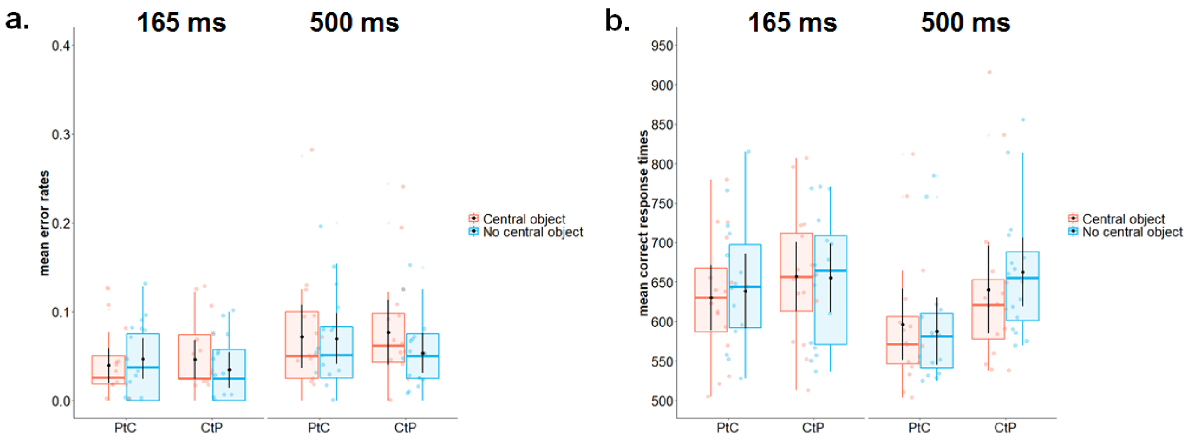

Figure 4. (a) Mean error rates in percentage and (b) mean correct response times in milliseconds during the categorization of the PtC and CtP sequences in each exposure duration condition (165 ms, 500 ms), according to the presence of a central object or not. Black dots and error bars indicate mean and standard error. Color dots are individual observations.

4.3. Discussion

Experiment 2 was specifically designed to test the advantage of a rapid analysis of coarse information available in peripheral vision, which could allow a first categorization, prior to its validation through the slower detailed analysis in central vision. For this, we used dynamic scene sequences composed of the central disk and the four rings assembled from the central disk to the more peripheral ring to impose a Central-to-Peripheral analysis (CtP sequences) or assembled from the more peripheral ring to the central disk to impose a Peripheral-to-Central analysis of visual information (PtC sequences). In this experiment, we also tested whether the superiority of peripheral vision was enhanced by a short exposure duration. Sequences were presented for either 165 ms or 500 ms, so that each stimulus that composed the sequence was presented for either 33 ms or 100 ms, respectively.

Results showed that the PtC advantage was greater for the longest than shortest exposure duration of the sequences. Importantly, this interaction was actually due to an unexpected effect when no object was present in central vision for the shortest exposure duration condition. This result will be further discussed in the General Discussion. However, a PtC advantage could be due to a methodological bias. Indeed, the surrounding peripheral rings presented after the central disk (CtP sequences) may have caused metacontrast masking of the central disk processing. Metacontrast masking is a particular masking effect observed when a visual mask does not overlap with a target stimulus location. Metacontrast masking is also called visual backward masking (Breitmeyer et al., 2000, 2006; Enns & Lollo, 2000). More precisely, it refers to a reduced visibility of a briefly presented stimulus (a target) by the presence of another brief stimulus (the mask), presented at the surround of the target. Importantly, this masking effect only appears under temporal constraints (Alperna, 1952). The target visibility is reduced when the mask is presented after the target. Therefore, the surrounding peripheral rings presented after the central disk (CtP sequences) may have reduced the visibility of the central disk.

Usually, metacontrast masking is characterized by the visibility reduction as a function of the duration of time between the target and the mask (stimulus onset asynchrony, SOA) can be plotted with a U-shaped function. More precisely, the metacontrast masking effect is greatest at relatively short SOAs (ranging from 30 to 80 ms; Breitmeyer & Ogmen, 2006). Therefore, it would operate more in the 165 ms condition, than the 500 ms condition. Our results showed the reverse pattern. In addition, if peripheral rings may have caused metacontrast masking of the central disk processing, it becomes difficult to explain why a metacontrast masking would manifest when a central object is present (condition for which we observed a PtC advantage), but not when a central object was absent (condition for which we did not observe a PtC advantage). However, given the existence of such methodological bias, we conducted a control experiment (Experiment 3) in order to quantify any contrast masking effect during the processing of CtP and PtC sequences and to assess if it is actually greater for a peripheral stimulus surrounding a central stimulus than the reverse.

In this control experiment, we manipulated the eccentricity of the stimuli to categorize and the presence of a mask sequence following the stimulus. In a non masking condition, participants had to categorize the central disk (CD) and the more peripheral ring (R4) of a scene. In a masking condition, they had to to categorize a sequence composed of five stimuli assembled from the center to the periphery in which the CD of a scene was presented first, then R1, R2, R3 and R4 of a pink noise mask (Central-to-Mask sequence) and a sequence composed of five stimuli assembled from the periphery to the center in which the R4 of a scene was presented first, then R3, R2, R1 and CD of a pink noise mask (Peripheral-to-Mask). If the processing of CtP and PtC sequences involved a visual backward masking effect, performances should be worse in the masking conditions. Importantly, if the visual backward masking is greater for a peripheral stimulus surrounding a central stimulus (CtP sequences) than the reverse (PtC sequences), the eccentricity of the stimulus to categorize should interact with the masking conditions, that is the difference between the masking and the no masking conditions should be greater for the categorization of the central disk than the peripheral ring. Thus, based on the results observed in Experiment 1, if a PtC advantage is due to a visual backward masking effect, we should observed a significant interaction between the eccentricity of the stimulus to categorize (central and peripheral stimuli) and the presence of a mask (mask and no mask) for the specific condition in which the central part of the scenes contained an object semantically related.

4. Experiment 3

4.1. Method

4.1.1. Participants

Thirty-two undergraduate students of Psychology from University Grenoble Alpes (28 women, mean age: 20.63 ± 2.68) participated in the experiment. They were divided into two groups (N = 16 each) for which individual stimuli were presented for either 33 or 100 ms, and sequences were presented for either 165 or 500 ms, respectively. The data of one participant of the group 500 ms were not used because he did not complete all the conditions. Inclusion criteria of participants were the same as in Experiments 1 and 2. The experiment was carried out within the same ethical framework as in Experiments 1 and 2.

4.1.2. Stimuli and Procedure

Central disk (CD) and Ring 4 (R4) stimuli were exactly the same as those used in Experiments 1 and 2. The mask was a large disk of 15° radius built with 1/f pink noise. We built stimuli that revealed different parts of the mask using the equation empirically derived from retinotopic measurement (Wu et al., 2012). As in Experiments 1 and 2, we created five stimuli of different eccentricities but with the same surface of activation on V1 (20%, i.e. ~ 136 mm^3). All stimuli were normalized to obtain an average luminance of 0.5 and a RMS contrast (Root Mean Square) of 0.2 for pixel intensity values between 0 and 1 (i.e. an average luminance of 128 and an RMS contrast of 51 on a scale of 256 gray levels). For the Central-to-Mask sequences, five stimuli were assembled from the center to the periphery: a CD of a scene was presented first, then R1, R2, R3, and R4 of the pink noise. For the Peripheral-to-Mask sequences, five stimuli were assembled from the periphery to the center: a R4 of a scene was presented first, then R3, R2, R1 and CS of the pink noise. Central and Peripheral individual stimuli were presented for either 33 or 100 ms. Sequences lasted either 165 ms (each stimulus within the sequence was presented for 33 ms) or 500 ms (each stimulus was presented for 100 ms). Figure 5 shows a schematic of a trial in each experimental condition.

Stimuli were displayed using on E-Prime software (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) using the same 30’ monitor as in Experiment 1 with a viewing distance of 70 cm. All participants performed one session and they were divided into two groups (N = 16 in each group) depending on the presentation of stimuli. Central stimuli, Peripheral stimuli, Central-to-Noise sequences and Peripheral-to-Noise sequences were randomly presented to the participants. A trial began with a central black fixation point presented for 500 ms on a 0.5 luminance background, followed an individual stimulus or a sequence. Inter-stimuli interval was a grey screen of 1900 ms during which participants could respond. Participants’ task was similar to Experiment 1 (i.e. indoor vs. outdoor categorization). The session included 320 trials (i.e. 20 stimuli x 2 Eccentricities x 2 Masking x 2 Categories). The experiment lasted about 14 minutes for the short Exposure duration condition, and about 15 min for the long Exposure duration condition, including breaks every 80 trials. For each trial, response accuracy and response time (in ms) were recorded. Before each experimental session, participants underwent a training session (20 trials) using other stimuli.

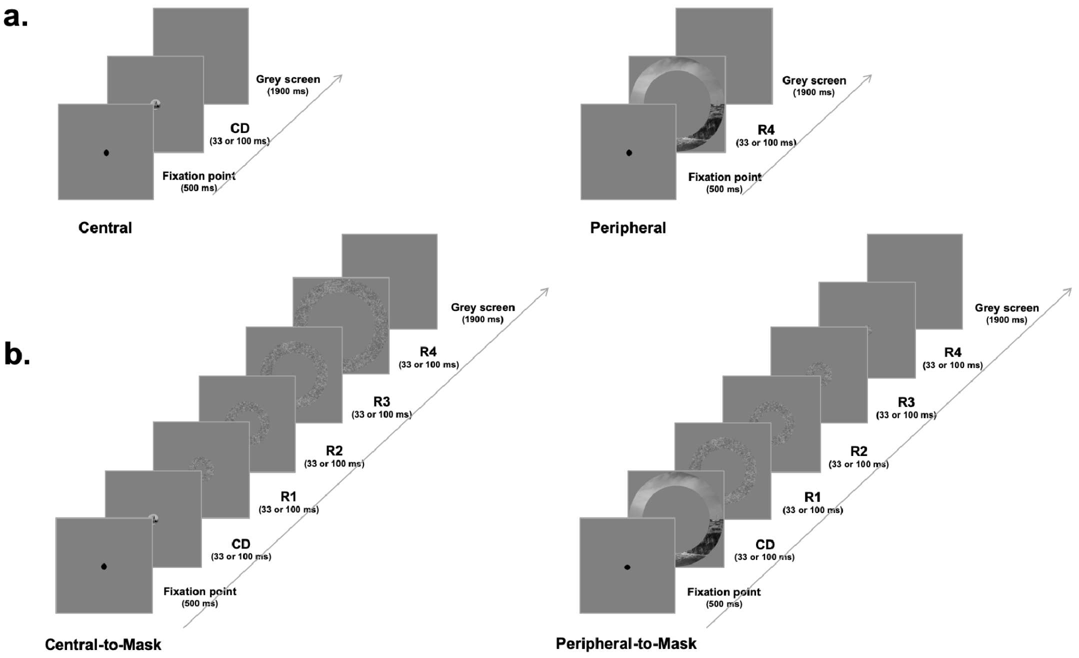

Figure 5. Trial schematic of Experiment 3. (a) Participant had to categorize a Central disk (CD) or a Ring 4 (R4) of a scene presented either for 33 or 100 ms. (b) Participant had to categorize a Central-to-Mask sequence composed of five stimuli assembled from the center to the periphery in which the CD of a scene was presented first, then R1, R2, R3 and R4 of a pink noise or a Peripheral-to-Mask sequence composed of five stimuli assembled from the periphery to the center in which the R4 of a scene was presented first, then R3, R2, R1 and CD of a pink noise. The sequences lasted either 165 ms (each stimulus was presented for 33 ms), or 500 ms (each stimulus was presented for 100 ms). Participants were requested to give their categorical answer as quickly and as accurately as possible, as soon as the first stimulus of the sequence appeared.

4.1.3. Results

Results are shown in Figure 6. Two repeated measures ANOVAs were performed on mean error rates (mER, in %) and mean correct response times (mRT, in ms). RTs were trimmed for each participant's correct response of each condition. We removed an average 0.80% of the trials. ANOVAs were performed using Statistica 13.3 software (Statsoft, Tulsa, USA). Effect sizes were estimated by calculating the partial eta-squared (η^2). The alpha level of tests was set at 0.05.

The ANOVAs included the Exposure Duration of the sequence (165 ms and 500 ms) as between-subject factor, and the Eccentricity of the stimulus to categorize (CD and R4), the presence of a Mask following the stimulus (Mask and No mask), the presence of a Central object in the scene (Object and No object), and the Category of scenes (Outdoor and Indoor) as within subject factors. As the mER was very high for the categorization of the Central disk without a central object (beyond 20%; Figure 6a), but the reaction times were very short for this condition (under 650 ms; Figure 6b), we first conducted ANOVAs on mean error rates (mER, in %) and mean correct response times (mRT, in ms) to ensure that there was no speed-accuracy trade-off for this specific condition. Critically, there actually was a speed-accuracy trade-off. Indeed, for the mER, the interaction between the Eccentricity and the presence of a Central object was significant (*F*(1,29) = 175.87, *p* < .001), due to more errors for the Central disk than the Ring 4 when no central object was present (*F*(1,29) = 162.69, *p* < .001; CD: 24 ± 2.10%, R4: 9.67 ± 1.65%), and no difference between the Central object and Ring 4 when a central object was present (*F*(1,29) < 1; CD: 9.01 ± 7.69%, R4: 8.55 ± 1.63%). However, for the mRT, performances were reversed. The interaction between the Eccentricity and the presence of a Central object was significant (*F*(1,29) = 75.67, *p* < .001), but this time it was due to faster reaction time for the Central disk than the Ring 4 when no central object was present (*F*(1,29) = 75.67, *p* < .001; CD: 621 ± 23 ms, R4: 668 ± 22 ms), and no difference between the Central object and Ring 4 when a central object was present (*F*(1,29) = 1.95, *p* = .173; CD: 662 ± 17 ms, R4: 675 ± 22 ms). Thus, for this specific condition, participants made more errors for categorizing the central disk than the peripheral disk, but they were faster. We conducted a second analysis without considering the No central object condition. It should be noted that in Experiment 2 we observed a PtC advantage irrespective of the exposure duration only when a central object was present. Thus, in order to test whether this PtC advantage is due to a visual backward masking effect, it would seem consistent to consider only the central object condition. The new ANOVAs included the Exposure Duration of the sequence (165 ms and 500 ms) as between-subject factor, and the Eccentricity of the stimulus to categorize (CD and R4), the presence of a Mask following the stimulus (Mask and No mask), and the Category of scenes (Outdoor and Indoor) as within subject factors.

The ANOVA performed on mER showed no significant main effect of the Eccentricity (*F*(1,29) < 1; R4: 8.55 ± 1.89%; CD: 9.01 ± 2.18%), but the main effect of the Mask was significant (*F*(1,29) = 12.67, *p* = .001, ηp2 = .30). Participants made less errors for categorizing a stimulus without a mask following it (7.22 ± 1.70%) than a stimulus with a mask (10.34 ± 2.33%). This result suggests a visual backward masking effect. However, the Mask did not interact with the Eccentricity (*F*(1,29) < 1), suggesting that the visual backward masking was not greater for a peripheral stimulus surrounding a central stimulus than the reverse. The Exposure duration did not interact with the Eccentricity and the Mask (*F*(1,29) < 1). Finally, there was neither a main effect of Category (*F*(1,29) < 1), nor interaction between the Category, the Eccentricity and the Mask (*F*(1,29) < 1).

The ANOVA performed on mRT showed no significant main effect of the Eccentricity (*F*(1,29) = 1.95, *p* = 173; R4: 675 ± 22 ms; CD: 662 ± 17 ms), but the main effect of the Mask was significant (*F*(1,29) = 45.68, *p* < .001, ηp2 = .61). Participants were faster for categorizing a stimulus without a mask following it (702 ± 32 ms) than a stimulus with a mask (744 ± 36 ms). Again, this result suggests a backward masking effect. As the mER analysis, the Mask did not interact with the Eccentricity (*F*(1,29) = 2.73, *p* = .109), suggesting that the visual backward masking was not greater for a peripheral stimulus surrounding a central stimulus than the reverse. The Exposure duration did not interact with the Eccentricity and the Mask (*F*(1,29) < 1). Furthermore, the main effect of Category was significant (*F*(1,29) = 4.30, *p* = .047, ηp2 = .13; Indoor: 730 ± 41 ms; Outdoor: 717 ± 33 ms), and this factor interacted with Eccentricity (*F*(1,29) = 5.88, *p* = .022). Planned comparisons showed that participants categorized faster the CD than the R4 for indoor scenes (R4: 743 ± 31 ms; CD: 717 ± 21 ms*;* *F*(1,29) = 4.83, *p* = .036) but not for outdoor scenes (*F*(1,29) < 1). Finally, the Category did not interact with the Eccentricity and the Mask (*F*(1,29) = 2.47, p = 127).

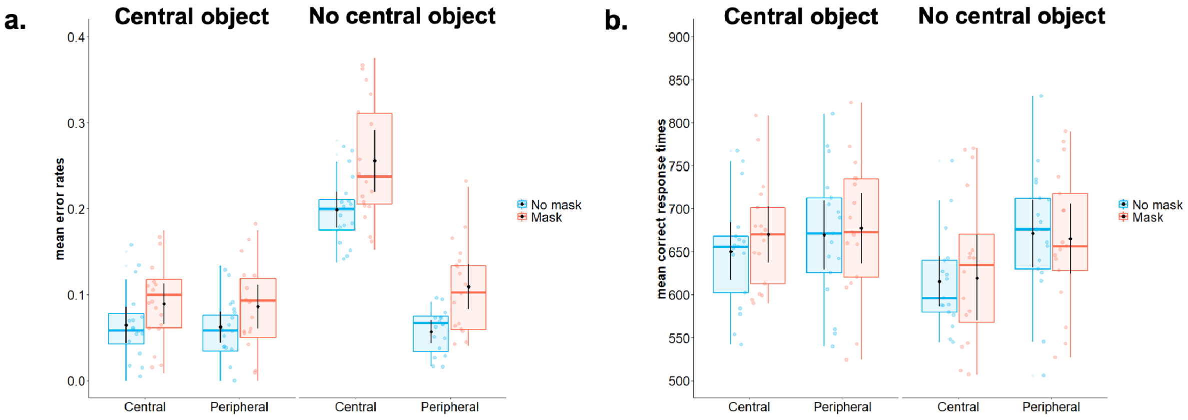

Figure 6. (a) Mean error rates in percentage and (b) mean correct response times in milliseconds during the categorization of the central (Central disk) and peripheral (Ring 4) stimuli followed or not by a sequence of masks, according to the presence of a central object or not. Black dots and error bars indicate mean and standard error. Color dots are individual observations.

5. General Discussion

The present study investigated the advantage of the peripheral vision on the central vision during rapid scene categorization. Given that spatial resolution decreases with retinal eccentricity, peripheral visual information would be mainly encoded in LSF while central visual information would be mainly encoded in HSF. Based on the hypothesis of a predominant coarse-to-fine processing of spatial frequencies, we wondered if scene analysis follows a coarse/peripheral-to-fine/central processing. Three experiments were conducted to address these issues. We used large scene from which we built one central disk and four circular rings of different eccentricities. In Experiment 1, participants had to categorize individual stimuli as indoor or outdoor scenes. In Experiment 2, they had to categorized dynamic sequences composed of the central disk and the four rings assembled from peripheral to central vision (PtC sequences) or assembled from central to peripheral vision (CtP sequences). Experiment 3 was designed to assess if our results could be biased by a metacontrast masking effect. In all experiments, we manipulated the semantic content available in the central part of the scene. The Central Disk either contained an object semantically related to the scene category (e.g., kitchen utensils for an indoor scene, or a house for an outdoor scene), or no semantically relevant object. These experiments yielded two key findings: (1) categorization performances improved as scene information was presented in peripheral vision (Experiment 1), (2) better performances were observed during categorization for PtC than CtP sequences when a relevant object was present in the Central disk (Experiment 2). However, the presence of a central object semantically related to the category of the scene significantly improved the categorization performances of the central disk, as well as of the CtP sequences presented for 500 ms.

In Experiment 1, we also observed that the absence of central object semantically related to the category of the scene was more detrimental to the categorization of indoor than outdoor scenes. Computational works (Quattoni & Torralba, 2009; Torralba, Murphy, Freeman & Rubin., 2003) have shown that computational models have better recognition performance for outdoor scenes than indoor scenes. Quattoni and Torralba (2009) suggest that outdoor scenes are usually well characterized by global spatial properties. Indoor scenes could also be characterized by global spatial properties (e.g., a scene with a corridor) but most of the time, they are better characterized by the objects they contain (e.g., a book in a bookstore, a sofa in a living room). Thus, the absence of an object in central vision could bias participants to consider an indoor scene as an outdoor scene. Importantly, the advantage of peripheral vision was observed even when the categorization could be facilitated by the presence of a central object semantically related to the category of the scene. In addition, in Experiment 1, accuracy of responses observed in near peripheral vision (categorization of Rings 1, 2 and 3 in Session 2) did not differ from performances obtained in central vision (Whole Disk categorization in Session 1), suggesting that this part of peripheral vision could be sufficient for scene categorization. Therefore, both experiments supports the greater efficiency and utility of peripheral vision relative to central vision during rapid scene categorization (Larson & Loschky, 2009), despite the low visual acuity inherent to peripheral vision and the informative value of central vision. Interestingly, our results suggest that there is a wide area of peripheral vision for which the categorization is optimum. Beyond this part, the recognition becomes less efficient.

It should be noted that we constructed our stimuli so that the surface of visual information revealed by each ring took into account the cortical magnification factor. Indeed, retinotopic projections to the primary visual cortex (V1) are deformed. A large number of V1 cells are dedicated to processing visual information from the central retina while a smaller number of areas are dedicated to more peripheral vision (Engel et al., 1994; Engel, Glover, & Wandell, 1997; Daniel & Whitteridge, 1961; Virsu, Näsänen & Osmoviita, 1987; Virsu & Rovamo, 1979). Therefore, visual information from the central visual field is over-represented at the cortical level. The foveal portion of the retina, which represents only 1% of the total surface of the retina, is represented on 50% of V1. This cortical magnification contributes to explain the decrease in visual acuity with retinal eccentricity (Duncan & Boynton, 2003). In order to account for this cortical magnification, the width and surface of rings in our experiments increased with eccentricity and the most peripheral ring (Ring 4) contained more quantity of visual information potentially more useful for categorization than the central ring. Importantly, Geuzebroek and van den Berg (2018) showed that peripheral vision would have an advantage over central vision in scene categorization precisely because it covers a larger surface of visual field allowing to reveal more scene content. In their experiment, authors presented scenes to categorize using a Window-Scotoma paradigm. Scenes were presented according five conditions. In a central viewing condition C1, a 5° radius window was placed on the scene and participants saw only the scene center. In a central viewing condition C2, the whole scene was reduced in a 5° radius window. Thus, stimuli C1 and C2 activated the same surface of V1, but stimulus C2 contained more visual information than stimulus C2. In a peripheral viewing condition P1, the scene was presented within a ring with inner and outer edges at 5° and 40° respectively. In a peripheral viewing condition P2, the scene was presented within a ring with inner and outer edges at 12° and 40° respectively. Finally, in a CO1 condition, stimulus P2 was reduced taking into account the cortical magnification factor, so that the scene was presented with the inner and outer edges at 3.6° and 12°. Thus, P2 and CO1 stimuli activated the same V1 surface and contained the same quantity of visual information. The results showed that performances were lower in the conditions where only central information was available (C1 and C2) compared to the conditions where only peripheral information was available (P1 and P2), thus confirming the previous results of Larson and Loschky (2009), as well the results of our Experiment 1. More interestingly, performances were worse in the condition where only the central information was available (C1) compared to the condition where all scene information was presented in central vision (C2). However, for conditions P2 and CO1, performances were similar despite the eccentricity change in scene presentation. These results suggests that scene categorization processes are independent of scene eccentricity, provided that central vision and peripheral vision both activate the same number of cells in V1 and that the same quantity of information is available in central and peripheral vision, although this is never the case under natural conditions. Overall, these results indicate that retinal eccentricity plays a minimal role in scene perception and peripheral vision has an advantage compared to central vision only when it covers a larger visual field, allowing presenting a larger scene content. In the present study, we chose to respect the natural functioning properties of the visual system by manipulating the width of stimuli according to the cortical magnification factor. Furthermore, we believe it would have been irrelevant to present the same width of visible scene at each eccentricity. For example, we could have applied the central disk width (radius of 1° of visual angle) to the peripheral rings. But this experimental control would result in exceeding the visual system capabilities in peripheral vision, which is tuned to extract LSF components, and may have strongly hindered categorization. In contrast, if we had applied the width of Ring 4 (7.74° of visual angle) to the Central disk, it would have included a portion of peripheral visual field. Therefore, controlling the width of rings for each eccentricity would not respect natural vision properties and may bias the strategies used by the visual system to categorize scenes.

Experiment 2 was specifically designed to test the advantage of a rapid analysis of coarse information available in peripheral vision, which could allow a first categorization, prior to its validation through the slower detailed analysis in central vision. For this, we used dynamic scene sequences composed of the central disk and the four rings assembled from the central disk to the more peripheral ring to impose a Central-to-Peripheral analysis (CtP sequences) or assembled from the more peripheral ring to the central disk to impose a Peripheral-to-Central analysis of visual information (PtC sequences). Sequences were presented for either 165 ms or 500 ms, so that each stimulus that composed the sequence was presented for either 33 ms or 100 ms, respectively. Participants categorized PtC sequences faster than CtP sequences, but the exposure duration of sequences significantly influenced this difference. The PtC advantage was greater for the longest than shortest exposure duration of the sequences. Importantly, this interaction was actually due to an unexpected effect when no object was present in central vision for the shortest exposure duration condition. Indeed, we observed a significant interaction between the type of sequences, their exposure duration and the presence of a central object. When the central part of the scenes contained an object semantically related to the scene category, participants categorized PtC sequences faster than CtP sequences irrespective of the exposure duration of sequences. Surprisingly, when there was no object, a PtC advantage was only observed for the longest exposure duration. In comparison to the condition in which an object was present, reaction times increased for categorizing the PtC sequences, as if the absence of a relevant information in central vision following the peripheral processing was detrimental to rapid scene categorization (sequences lasted 165 ms). This result could be interpreted in the context of a predominant coarse-to-fine processing during scene categorization (Kauffmann et al., 2014; Schyns & Oliva, 1994). According to this hypothesis, a rapid extraction of predominantly coarse information in the peripheral visual field should provide the global shape and structure of the scene used for a rapid initial perceptual categorization. This perceptual categorization should be refined, confirmed, or infirmed by the processing of finer information available in the central visual field. In the present experiment, this periphery-based rapid categorization would be then confirmed by the presence of a relevant object in the central part of the scene. When no relevant information was present in central vision, the visual system would need more visual information, delaying the categorization time. The stronger PtC advantage for the longest sequences irrespective of the presence of a central object suggests that the fine information from central vision would not even be used. It is possible that for sequences of longer duration (500 ms), the categorization could have been only based on the first ring, presented for 100 ms. Results of Experiment 1, as well as previous studies on scene recognition (Johnson & Olshausen, 2003; Thorpe, Fize, & Marlot., 1996; VanRullen & Thorpe, 2001), confirms the ability of the visual system to categorize a scene presented for just 100 ms.

Alternatively, a PtC advantage could be due to a methodological bias. Indeed, the surrounding peripheral rings presented after the central disk (CtP sequences) may have caused metacontrast masking of the central disk processing. We conducted a control experiment (Experiment 3) in order to quantify any contrast masking effect during the processing of CtP and PtC sequences and to assess if it is actually greater for a peripheral stimulus surrounding a central stimulus than the reverse. In this experiment, participants had to categorize a Central disk or a Ring 4, presented alone or followed by a sequence of masks of different eccentricities inducing a metacontrast masking similar to the one that could be induced by the CtP and PtC sequences, respectively. Results showed a main effect of masking suggesting that a sequence of masks presented after the Central disk or the Ring 4 may have reduced their visibility. In addition, there was no significant interaction between the eccentricity of the stimulus to categorize (Central disk or Ring 4) and the presence of a mask following the stimulus when there was a central object in the scene, suggesting that the visual backward masking observed for a peripheral stimulus surrounding a central stimulus (CtP sequences) was similar to the one observed for a central stimulus surrounding a peripheral stimulus (PtC sequences). Therefore, even if our paradigm intrinsically caused metacontrast masking, this methodological bias would be as strong for the CtP sequences as for the PtC sequence containing an object in central vision, and so it could hardly account for the PtC advantage in this condition.

Another alternative account for a stronger PtC advantage for longer exposure duration may be linked to eye-movements. Indeed, for the longest sequences, it is likely that participants fixated each stimulus sequentially. Unfortunately, we could not record eye movements during this experiment in order to control that participants maintained their gaze on the central fixation throughout the sequence presentation. Therefore, it cannot be totally excluded that participants made eye movement when sequence lasted 500 ms. We however believe it is unlikely that such behavior occurred and explains the PtC advantage. Given that participants did not know in advance the eccentricity of the first stimulus (either the central disk or Ring 4), it would have been inefficient without impairing scene categorization. Participants could also initiate a saccade from one stimulus to the next stimulus in the sequence, but this strategy would have been demanding. Indeed, saccadic eye movements are usually initiated within 100-150 ms (Fisher & Weber, 1993). If participants made a saccade during the presentation of one stimulus (presented during 100 ms), this puts severe constraints on the time left to process visual information in the next stimulus.

Further studies investigating the preferential use of central vs. peripheral vision at different stages of visual processing could help to refine these results. For example, it could be considered to adapt the hybrid scene paradigm (categorization of stimuli composed of an LSF scene superimposed with an HSF scene of another category and presented at short and long exposure durations) introduced by Schyns and Oliva (1994), by combining the central part of a scene belonging to a particular category with the peripheral part of another scene belonging to another category in order to examine which scene information is preferentially used at early and late stages of visual processing. In this context, a recent study by Vanmarcke et al. (2016) for example showed that the presentation of a background context interferes on the object categorization at very short exposure duration. This effect disappeared as the exposure duration increased suggesting an early contribution of peripheral vision. It should also be noted that in this study, we did not filter out the spatial frequency content from the scene, since peripheral vision acts as a natural low-pass filter. It seems actually unlikely that such manipulation would have impacted scene processing, given the relatively short exposure duration of stimuli used in our study. Indeed, using a gaze-contingent display during scene exploration, in which the periphery was occasionally filtered in LSF during fixations, Loschky et al. (2005) showed that participants did not detect such degradation when fixations were below 100 ms. Yet, to our knowledge, the spatial frequency content of central vs peripheral scenes has never been manipulated in the context of scene categorization and future studies might be relevant to further address this issue. Further studies may also be necessary to assess the extent to which a preferential peripheral-to-central processing strategy can depend on the task demands. As shown in the context of spatial frequency processing, it is likely that a flexible use of central vs. peripheral information can occur according to different factors such as the categorization level (e.g., more subordinate categorization tasks may require a greater use of fine information available at the center; Collin & McMullen, 2005, Collins, 2006, Mermillod et al., 2005), the constraints of the visual task (e.g., gender categorization facial expression categorization; Schyns & Oliva, 1999) or inter-individual differences (see for example Vanmarcke & Wagemans, 2016). As a matter of fact, results of Experiment 1 revealed that participants made less errors for categorizing outdoor than indoor scenes. For RTs, there was a significant interaction between the Category and the Eccentricity of stimuli. This interaction was due to faster reaction times for outdoor than indoor scenes only for the more peripheral stimulus (Ring 4), as if peripheral vision was specifically relevant to the recognition of outdoor scenes. In Experiment 2, there was a significant interaction between the Category and the Sequence for RTs, the PtC advantage being greater for the outdoor scenes. Even if responses keys were counterbalanced across participants, these results suggest a response bias in favor of outdoor scenes, maybe because an outdoor vs. indoor categorization task with large stimuli favors the categorization of outdoor scenes.

In conclusion, results of the present study support the greater efficiency and utility of peripheral vision than central vision when rapidly categorizing scenes, despite the low visual acuity inherent to peripheral vision and the informative value of central vision. In agreement with a coarse-to-fine processing of spatial frequencies (Kauffmann et al., 2014; Kauffmann et al., 2015; Schyns & Oliva, 1994), the rapid analysis of LSF information available in peripheral vision could allow a first categorization that would be validated or not by the slower HSF information in central vision

Footnote:

In order to verify the semantic value of the object contained in the Central Disk, we conducted a pilot experiment. A group of ten undergraduate students were shown all Central Disk stimuli containing an object semantically related to the category of the scenes (100 stimuli). They had to categorize the stimuli (indoor vs. outdoor) and then to name the objects (e.g., a lamp, a boat). Objects were named correctly (97% of correct responses) and all stimuli were categorized in correspondence to the category of the scene from which they were extracted (100% of correct responses) confirming that our stimuli were centered on an object that was semantically related to the category of the scene.

Acknowledgment

This work was supported by NeuroCoG IDEX Université Grenoble Alpes in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02). We warmly thank Lester Loschky and another anonymous Reviewer for their valuable and constructive comments on earlier versions of the manuscript.

References

Alpern, M. (1952). Metacontrast; historical introduction. *American Journal of Optometry and Archives of American Academy of Optometry*, 43, 648-657.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of cognitive neuroscience*, *15*(4), 600-609.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in cognitive sciences*, *11*(7), 280-289.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … others. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454.

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, *86*, 35-42.

Breitmeyer, B. G., Kafalıgönül, H., Öğmen, H., Mardon, L., Todd, S., & Ziegler, R. (2006). Meta-and paracontrast reveal differences between contour-and brightness-processing mechanisms. *Vision Research*, 46(17), 2645-2658.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, 62(8), 1572-1595.

Collin, C.A. (2006). Spatial-frequency thresholds for object categorisation at basic and subordinate levels. *Perception*, 35(1), 41–52.

Collin, C.A., & McMullen, P.A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Perception & Psychophysics*, 67(2), 354–364.

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of comparative neurology*, *292*(4), 497-523.

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology*, *159*(2), 203-221.

Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, *38*(4), 659-671.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral cortex*, *7*, 181–92.

Engel, S. A., Rumelhart, D., Wandell, B. A., Lee, A., Glover, G. H., Chichilnisky, E.-J., & Shadlen, M. (1994). fMRI of human visual cortex. *Nature*, *369*, 525.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in cognitive sciences*, 4(9), 345-352.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences, 16*(3), 553-567.

Geuzebroek, A. C., & van den Berg, A. V. (2018). Eccentricity scale independence for scene perception in the first tens of milliseconds. *Journal of Vision*, *18*(9), 9-9.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in neurobiology*, *84*(4), 405-439.

Johnson, J. S., & Olshausen, B. A. (2003). Time-course of neural signatures of object recognition. *Journal of Vision*, *3*(7), 4-4.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision research*, 47(26), 3286-3297.

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.

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, *107*, 49-57.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in integrative neuroscience*, *8*, 37.

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and cognition*, *65*(2), 145-168.

Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 471.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10), 6-6.

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.

Loschky, L. C., McConkie, G. W., Yang, J., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*, 12(6), 1057-1092.

Loschky, L. C., Szaffarczyk, S., Beugnet, C., Young, M. E., & Boucart, M. (2019). The contributions of central and peripheral vision to scene-gist recognition with a 180° visual field. *Journal of Vision*, *19*(5), 15-15.

Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: Evidence from neuro-computational modeling. *Brain Cognition,* 57, 151– 157.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., ... & Vuilleumier, P. (2010). The neural substrates and timing of top–down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Journal of cognitive neuroscience*, *22*(12), 2768-2780.

Quattoni, A., & Torralba, A. (2009). Recognizing indoor scenes. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 413-420). IEEE.

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes?. *Visual Cognition*, 12(6), 852-877.

Roux-Sibilon, A., Rutgé, F., Aptel, F., Attye, A., Guyader, N., Boucart, M., ... & Peyrin, C. (2018). Scene and human face recognition in the central vision of patients with glaucoma. *PloS One*, *13*(2), e0193465.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science*, *5*(4), 195-200.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69(3), 243-265.

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

Torralba, A., Murphy, K. P., Freeman, W. T., & Rubin, M. A. (2003). Context-based vision system for place and object recognition. *In Proc, ICCV.*

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*(1), 190-198.

Vanmarcke, S., Calders, F., & Wagemas, J. (2016). The time-course of ultrarapide categorization: the influence of scene congruency and top-down processing. *i-perception*, 7(5), 1-23.

Vanmarcke, S. & Wagemans, J. (2016). Individual differences in spatial frequency processing in scene perception: the influence of autism-related traits. *Visual Cognition*, 24(2), 115-131.

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.

Virsu, V., Näsänen, R., & Osmoviita, K. (1987). Cortical magnification and peripheral vision. *JOSA A*, *4*(8), 1568-1578.

Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research*, *37*(3), 475-494.

Wu, J., Yan, T., Zhang, Z., Jin, F., & Guo, Q. (2012). Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging. *Human brain mapping*, *33*(7), 1727-1740.