



**HAL**  
open science

## How does information from low and high spatial frequencies interact during scene categorization?

Louise Kauffmann, Alexia Roux-Sibilon, Brice Beffara, Martial Mermillod,  
Nathalie Guyader, Carole Peyrin

► **To cite this version:**

Louise Kauffmann, Alexia Roux-Sibilon, Brice Beffara, Martial Mermillod, Nathalie Guyader, et al..  
How does information from low and high spatial frequencies interact during scene categorization?.  
Visual Cognition, 2017, 25 (9-10), pp.853-867. 10.1080/13506285.2017.1347590 . hal-01599312

**HAL Id: hal-01599312**

**<https://hal.science/hal-01599312>**

Submitted on 23 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: <http://www.tandfonline.com/loi/pvis20>

## How does information from low and high spatial frequencies interact during scene categorization?

Louise Kauffmann, Alexia Roux-Sibilon, Brice Beffara, Martial Mermillod, Nathalie Guyader & Carole Peyrin

To cite this article: Louise Kauffmann, Alexia Roux-Sibilon, Brice Beffara, Martial Mermillod, Nathalie Guyader & Carole Peyrin (2017) How does information from low and high spatial frequencies interact during scene categorization?, *Visual Cognition*, 25:9-10, 853-867, DOI: [10.1080/13506285.2017.1347590](https://doi.org/10.1080/13506285.2017.1347590)

To link to this article: <https://doi.org/10.1080/13506285.2017.1347590>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 17 Aug 2017.



[Submit your article to this journal](#)



Article views: 507



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

## How does information from low and high spatial frequencies interact during scene categorization?

Louise Kauffmann<sup>a,b†</sup>, Alexia Roux-Sibilon<sup>a†</sup>, Brice Beffara<sup>a</sup>, Martial Mermillod<sup>a</sup>, Nathalie Guyader<sup>c</sup> and Carole Peyrin<sup>a</sup>

<sup>a</sup>Department of Psychology, University of Grenoble Alpes, CNRS, LPNC UMR 5105, Grenoble, France; <sup>b</sup>Neural Mechanisms of Human Communication Research Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; <sup>c</sup>Image and Signal Department, University of Grenoble Alpes, GIPSA-lab UMR5216, Grenoble, France

### ABSTRACT

Current models of visual perception suggest that, during scene categorization, low spatial frequencies (LSF) are rapidly processed and activate plausible interpretations of visual input. This coarse analysis would be used to guide subsequent processing of high spatial frequencies (HSF). The present study aimed to further examine how information from LSF and HSF interact and influence each other during scene categorization. In a first experimental session, participants had to categorize LSF and HSF filtered scenes belonging to two different semantic categories (artificial vs. natural). In a second experimental session, we used hybrid scenes as stimuli made by combining LSF and HSF from two different scenes which were semantically similar or dissimilar. Half of the participants categorized LSF scenes in hybrids, and the other half categorized HSF scenes in hybrids. Stimuli were presented for 30 or 100 ms. Session 1 results showed better performance for LSF than HSF scene categorization. Session 2 scene categorization was faster when participants attended and categorized LSF than HSF scene in hybrids. The semantic interference of a semantically dissimilar HSF scene on LSF scene categorization was greater than the semantic interference of a semantically dissimilar LSF scene on HSF scene categorization, irrespective of exposure duration. These results suggest a LSF advantage for scene categorization, and highlight the prominent role of HSF information when there is uncertainty about the visual stimulus, in order to disentangle between alternative interpretations.

### ARTICLE HISTORY

Received 19 December 2016  
Accepted 20 June 2017



### KEYWORDS

Spatial frequency;  
coarse-to-fine; hybrids;  
semantic interference


## Introduction

At a glance, the human visual system is able to robustly process and categorize complex stimuli such as natural scenes (e.g., Thorpe, Fize, & Marlot, 1996), despite their infinite variability in terms of objects they contained, or their spatial configuration. Converging data from neurophysiological recordings in primates (Bullier, 2001; De Valois, Albrecht, & Thorell, 1982; Poggio, 1972; Shams & von der Malsburg, 2002; Shapley & Lennie, 1985; Van Essen & Deyoe, 1995) and psychophysical studies in human (Hughes, Nozawa, & Kitterle, 1996; Parker, Lishman, & Hughes, 1996; Schyns & Oliva, 1994; for a review see Sowden & Schyns, 2006) indicate that the visual system

extract visual information through a set of channels/filters differently tuned to specific orientations and spatial frequency bands of the visual input. Based on these data, current models of visual perception have emphasized the role of spatial frequency information for visual categorization (Bar, 2003; Hegdé, 2008; Kauffmann, Ramanoël, & Peyrin, 2014; Peyrin et al., 2010). According to these models, visual analysis begins with the parallel extraction of different elementary features at different points of the spatial frequency spectrum that provide different information about the visual scene. Lower spatial frequencies (LSF) provide coarse information, such as the global shape and structure of a visual scene, and are

**CONTACT** Louise Kauffmann  [kauffmann@cbs.mpg.de](mailto:kauffmann@cbs.mpg.de)  Department of Psychology, University of Grenoble Alpes, CNRS, LPNC UMR 5105, F-38040 Grenoble, France; Neural Mechanisms of Human Communication Research Group, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1A, 04103 Leipzig, Germany

<sup>†</sup>These authors contributed equally.

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/13506285.2017.1347590>.

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

predominantly conveyed through fast magnocellular channels. Higher spatial frequencies (HSF) provide finer information about the scene, such as edges or object details, and are conveyed more slowly through parvocellular channels. On the basis of the neurophysiological properties of the magno- and parvocellular pathways (Bullier, 2001; Maunsell et al., 1999) and results of psychophysical studies in humans (Hughes et al., 1996; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012; Parker, Lishman, & Hughes, 1992; Schyns & Oliva, 1994), it has been suggested that visual scene analysis follows a predominantly coarse-to-fine processing sequence. LSF information would be extracted first, allowing a coarse parsing of the visual input, prior to the analysis of fine information contained in HSF. In this theoretical framework, it was also hypothesized that rapid LSF information may guide the subsequent processing of HSF (Bar, 2003; Kauffmann et al., 2014; Peyrin et al., 2010; Trapp & Bar, 2015).

A predominant coarse-to-fine processing of scenes was first evidenced in a behavioural study by Schyns and Oliva (1994). The stimuli used by these authors were hybrid scenes made by adding a scene filtered in LSF (e.g., a highway) to another scene filtered in HSF (e.g., a city) and presented for 30 or 150 ms. They showed that, for short exposure (30 ms), categorical perception of hybrids was based on the category of the LSF scene which illustrates a perception dominated by the scene in LSF. However, for longer exposure (150 ms), the perception of hybrids was dominated by the scene in HSF. Importantly, participants were not aware that the hybrids contained two different scene categories. These results therefore suggested that, in the absence of awareness of any apparent conflict between two scenes in the stimulus, LSF information is predominantly used at early stages of visual processing whereas HSF are used at later stages in accordance with a coarse-to-fine visual processing sequence. Further studies supported these results by showing for example that scenes filtered in LSF are categorized more quickly than scenes filtered in HSF (e.g., Kauffmann, Ramanoël, Guyader, Chauvin, & Peyrin, 2015) and that the processing of LSF before HSF information is more advantageous for scene categorization than the reverse processing sequence (i.e., HSF before LSF; e.g., Kauffmann, Chauvin, et al., 2015; Musel et al., 2012). However, little is known about the integration of LSF and HSF

before the full recognition of scenes and the potential influence of LSF information on HSF processing.

Recent neuroimaging studies investigated that issue using hybrid scenes as stimuli and examining directly how the categorization of the HSF component in hybrids is influenced by the LSF component (Kauffmann, Bourgin, Guyader, & Peyrin, 2015; Mu & Li, 2013). In these studies, participants were aware of the two different scenes in hybrids and were instructed to ignore the LSF scene in hybrids, attend the HSF scene, and categorize the latter. The two scenes composing the hybrids were either semantically similar (e.g., a natural scene in LSF superimposed with a natural scene in HSF) or dissimilar (e.g., a natural scene in LSF superimposed with a man-made scene in HSF). The categorization of HSF scenes was impaired when the two scenes composing the hybrids were semantically dissimilar. Although it was not relevant to perform the task, the semantic information contained in the LSF scene was still processed and interfered with the categorization of the HSF scene when the two scenes were semantically dissimilar. This semantic interference effect was even greater when the two scenes composing the two semantically dissimilar scenes were physically similar (Mu & Li, 2013), suggesting that the spatial overlap between LSF and HSF caused greater impairment of HSF categorization in the absence of any semantic congruence between the two scenes. The semantic interference effect was thus interpreted as the signature of the integration of semantic information from both LSF and HSF and the influence of LSF over HSF scene categorization.

On a neurobiological level, results from ERP recordings (Mu & Li, 2013) revealed that the semantic interference effect was associated with a negative frontal component (N1) 122 ms after stimulus onset, suggesting that this short amount of time is sufficient for the integration of semantic information from both LSF and HSF in frontal areas. Results from fMRI (Kauffmann, Bourgin, et al., 2015) further showed that the semantic interference effect involved the inferior frontal gyrus (including the orbitofrontal cortex) and the inferotemporal cortex, and was associated with greater functional connectivity between these two regions. Overall, results from these two studies supported the proactive model of visual recognition (Bar, 2003, 2007; Trapp & Bar, 2015). In that theoretical context, the semantic interference effect in frontal and inferotemporal areas was interpreted as reflecting

erroneous predictions generated in the inferior frontal cortex based on LSF information of hybrids in the semantically dissimilar condition that led to impaired HSF scene categorization in the inferotemporal cortex, resulting in greater connectivity between these areas.

These studies provided supplementary argument in favour of a coarse-to-fine categorization of scenes, by showing that fast processing of LSF information is automatic and cannot be voluntarily inhibited. Furthermore, they suggested that the extraction of semantic information contained in LSF strongly influences the processing of HSF information during scene categorization. However, whether HSF processing can also influence LSF categorization was not the main focus of these studies. Although the prominent role of LSF information for rapid scene categorization has been extensively investigated and documented (Hughes et al., 1996; Schyns & Oliva, 1994; for a review see Hegdé, 2008; Kauffmann et al., 2014), there is also considerable experimental evidence of a predominant processing of HSF information even for very short stimuli exposure duration (Campagne et al., 2016; Harel & Bentin, 2009; Morrison & Schyns, 2001; Rotshtein, Schofield, Funes, & Humphreys, 2010; Schyns, 1998; Schyns & Oliva, 1999). Furthermore, many studies have shown that the use of spatial frequency information during the processing of complex stimuli such as scenes is highly flexible and depends on many factors such as stimulus exposure duration (Schyns & Oliva, 1994), category (e.g., Awasthi, Sowman, Friedman, & Williams, 2013; Collin & McMullen, 2005; Rotshtein et al., 2010; Vannucci, Viggiano, & Argenti, 2001), or task constraints (Abrams, Barbot, & Carrasco, 2010; Campagne et al., 2016; Caplette, West, Gomot, Gosselin, & Wicker, 2014; Fradcourt, Peyrin, Baciú, & Campagne, 2013; Morrison & Schyns, 2001; Ozgen, Payne, Sowden, & Schyns, 2006; Schyns & Oliva, 1999; Sowden, Özgen, Schyns, & Daoutis, 2003). For example, different spatial frequency bands would be used according to their diagnosticity to categorize a specific visual stimulus. It has been shown that spatial frequencies below 2 cpd are diagnostic to perform basic-level categorization of scenes (e.g., forest, highway, mountain; McCotter, Gosselin, Sowden, & Schyns, 2005). However, intermediate spatial frequencies of 2.3–4 cpd would be required for basic-level categorization of objects (Caplette et al., 2014). Importantly, most of

these studies focused on the preferential use, the relevance, or the diagnosticity of a specific spatial frequency band for a given timescale or task constraint. However, how information from a spatial frequency band that is not directly relevant for the task at stake can contribute to categorization has received little interest. For example, in the context of a predominant coarse-to-fine sequence of spatial frequency processing, little is known about the relative role of HSF information and its impact on LSF processing. In their study, Mu and Li (2013) reported the results of an independent control experiment in which participants were instructed to attend and categorize the LSF scene in hybrid and ignore the HSF scene. In this experiment, they also observed a semantic interference effect suggesting that information from HSF also influenced LSF categorization. However, participants performance in the semantically dissimilar condition was not significantly different than chance, possibly indicating that they were not able to perform the task. Additionally, the authors did not address whether the semantic interference effect was greater in one attentional condition than the other (i.e., Attention to the LSF or HSF scene), that is, whether LSF interfered/influenced more strongly HSF categorization than the other way around. It is also likely that the relative weight of information from LSF and HSF varies over the time-course of scene categorization. In that sense, the influence of LSF information might be prominent at early stages of scene categorization, whereas information from HSF would take over at later stages.

The aim of the present behavioural study was to further examine how information from LSF and HSF interact and influence each other during visual scene categorization. To this end, we used a semantic interference paradigm (Kauffmann, Bourgin, et al., 2015; Mu & Li, 2013). We presented hybrid stimuli made by combining the LSF and HSF components of two different scene categories, and we manipulated the semantic and physical similarity between the two scenes. Critically, we also manipulated attentional constraints. One group of participants had to attend and categorize the HSF scene in hybrids (HSF attentional condition), whereas the other group had to attend and categorize the LSF scene (LSF attentional condition). Additionally, to examine whether the relative weight of LSF and HSF varies over time, we manipulated the exposure duration of stimuli which were

presented for 30 or 100 ms. Prior to the attentional task, all participants had to categorize the individual LSF and HSF scenes used in hybrids in order to examine the processing of filtered scenes independently of attentional constraints. This also allowed us to ensure that filtered scenes composing the hybrids could be accurately categorized at both exposure durations. We expected the semantic interference effect to be modulated by the attentional task and the exposure duration of stimuli. In the context of a predominant coarse-to-fine processing of scenes, we hypothesized that at short exposure duration the semantic interference effect would be greater in the HSF than LSF Attentional condition, suggesting greater interference of LSF over HSF categorization than of HSF over LSF categorization. However, we expected this effect to be weaker at longer exposure duration, or perhaps reversed in such a way that HSF would interfere more with LSF categorization than the other way around.

## Method

### Participants

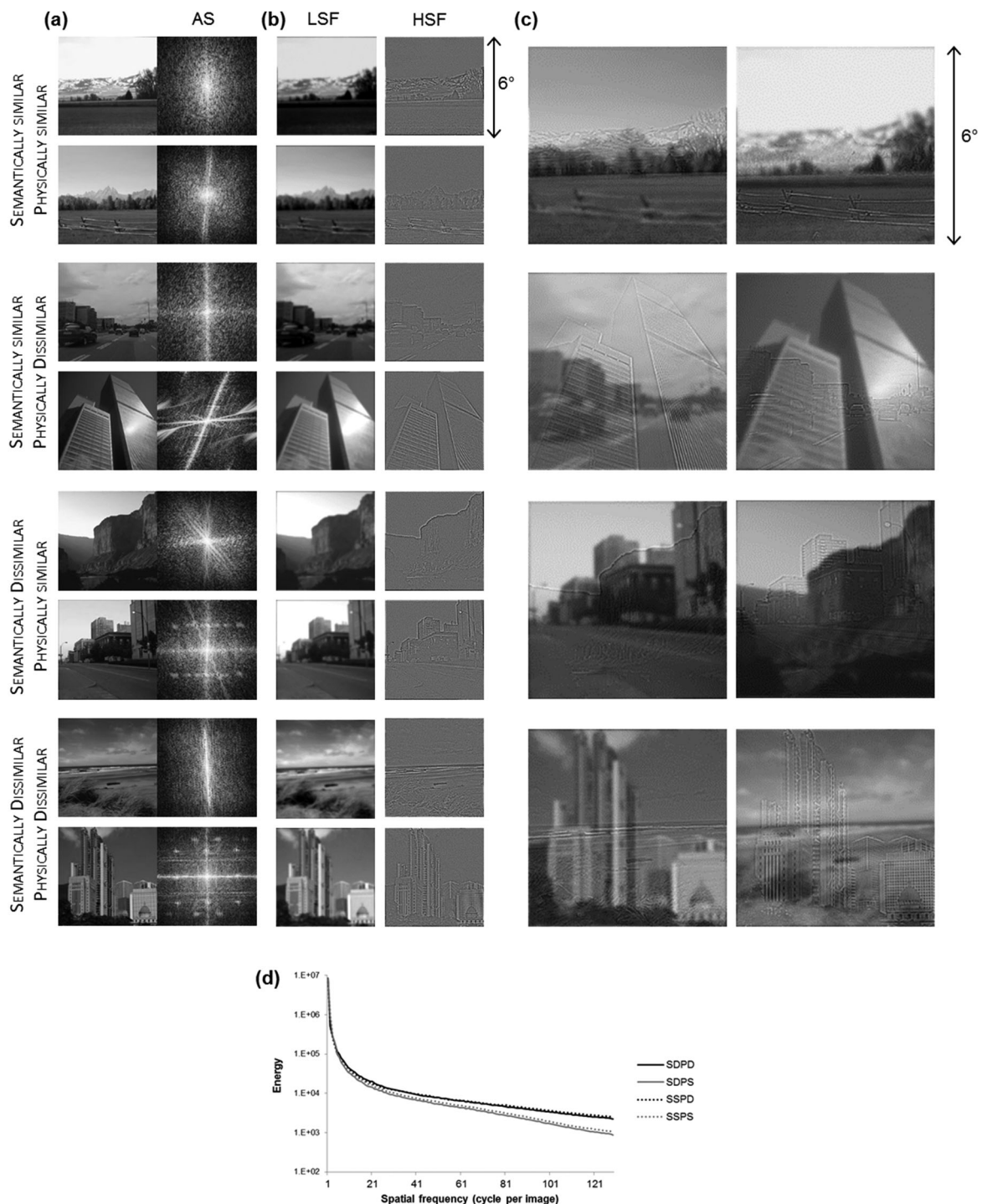
A total of 52 right-handed participants (15 male; Mean age  $\pm$  SD:  $22 \pm 3$  years, range: 18–30 years) with normal or corrected-to-normal vision and no history of neurological disorders were included in this experiment. All participants gave their informed written consent before participating in the study in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

### Stimuli and procedure

The stimuli used in the present study were black and white photographs (256-level grey-scales,  $256 \times 256$  pixels) taken among the scenes of the Labelme database (Oliva & Torralba, 2001) and subtended  $6 \times 6$  degrees of visual angle. In order to select our stimuli, we first classified the 2687 scenes from this database into two distinct categories: natural (e.g., forests, coasts, open countryside) and man-made (e.g., highways, streets, buildings). We then assessed the physical similarity of scenes in the database using (1) the correlation between images based on their pixel intensity values (PI correlation) and (2) the correlation

between their amplitude spectra (AS correlation, i.e., correlation between pixel intensity values of amplitude spectra; see Kauffmann, Bourgin, et al., 2015, for a similar procedure). We considered two scenes as physically similar if their PI and AS correlation coefficients were superior to 0.6 (only 1.59% of images pairs matched both criteria). We considered two scenes as physically dissimilar if their PI correlation coefficient was inferior to 0.01 and their AS correlation coefficient was inferior to 0.6 (only 1.37% of image pairs matched both criteria). Based on these computations, we selected among the pairs of images with similar physical properties 40 pairs of scenes semantically similar (20 for each category; PI:  $0.75 \pm 0.08$ ; AS:  $0.67 \pm 0.03$ ) and 40 pairs of scenes semantically dissimilar (PI:  $0.73 \pm 0.06$ ; AS:  $0.66 \pm 0.03$ ), and among the pairs of images with dissimilar physical properties 40 pairs of scenes semantically similar (20 for each category; PI:  $0.01 \pm 0.00$ ; AS:  $0.43 \pm 0.06$ ) and 40 pairs of scenes semantically dissimilar (PI:  $0.01 \pm 0.00$ ; AS:  $0.44 \pm 0.05$ ). This resulted in a total of 160 scenes that were used as stimuli. With respect to the physical similarity conditions, scenes were selected to have similar PI and AS correlation values for the semantically similar and dissimilar pairs of images. Furthermore, mean energy was similarly distributed across spatial frequencies for the semantically similar and dissimilar image groups (see Figure 1(d)).

Each scene was filtered to preserve either LSF or HSF information, using the MATLAB image processing toolbox (Mathworks Inc., Sherborn, MA, USA). Filtered scenes were obtained by multiplying the Fourier transform of the original images with Gaussian filters. The standard deviation of Gaussian filters was a function of the spatial frequency cut-off for a standard attenuation of 3 dB. We removed spatial frequencies above 4 cycles per degree of visual angle (cpd, i.e., 24 cycles per image) in LSF scenes<sup>1</sup> and below 6 cpd (i.e., 36 cycles per image) in HSF scenes. The resulting filtered scenes were then normalized to obtain a mean luminance of 0.5, for pixel intensity values between 0 and 1 (i.e., mean luminance of 128 on a grey-level scale). It should be noted that, for some images, this luminance normalization procedure resulted in obtaining luminance values below 0 or above 1. These values were reset to 0 and 1, respectively, after conversion of values into integer luminance values ranging from 0 to 255 (i.e., conversion into 8-bit data type). However, this did not affect the mean



**Figure 1.** (a) Examples of black and white of photographs of scenes used with their amplitude spectrum (AS). On the amplitude spectrum images, low spatial frequencies (LSF) are close to the centre and high spatial frequencies (HSF) are on the periphery. (b) In the Individual scene experimental session (Session 1), we presented artificial and natural scenes filtered in LSF or in HSF. All participants had to categorize LSF and HSF scenes. (c) Examples of the stimuli used in the Hybrid experimental session (Session 2). Hybrid scenes were made by adding a LSF scene and a HSF scene. The two scenes in hybrids were either (1) semantically and physically similar (SSPS; e.g., LSF of a natural scene superimposed with HSF of another natural scene with similar physical properties), (2) semantically similar and physically dissimilar (SSPD; e.g., LSF of a man-made scene superimposed with HSF of another man-made scene with dissimilar physical properties), (3) semantically dissimilar and physically similar (SDPS; e.g., LSF of a natural scene superimposed with HSF of a man-made scene with similar physical properties), or (4) semantically and physically dissimilar (SDPD; e.g., LSF of a natural scene superimposed with HSF of a man-made scene with dissimilar physical properties). Half of the participants had to attend and categorize the LSF scene in hybrids (LSF-Attention) whereas the other half had to attend and categorize the HSF scene (HSF-Attention). (d) Mean energy distribution across spatial frequencies of images in the four experimental conditions (SDPD, SDPS, SSPD, SSPS).

luminance of filtered scenes which stayed at 0.5 (i.e., 128 on a grey-level scale) after this conversion. Root-mean square (RMS) contrast of filtered scenes was not modified (see note 1) and was higher for LSF than HSF scenes (LSF natural: mean RMS  $\pm$  SD:  $0.24 \pm 0.06$ ; LSF man-made:  $0.24 \pm 0.05$ ; HSF natural:  $0.06 \pm 0.02$ ; HSF man-made:  $0.07 \pm 0.02$ ). The displayed luminance of filtered scenes was slightly higher for LSF (mean luminance  $\pm$  SD:  $14.89 \pm 2.26$  candela/m<sup>2</sup>) than HSF scenes (mean luminance  $\pm$  SD:  $11.75 \pm 1.33$  candela/m<sup>2</sup>). RMS contrast of filtered scenes was higher for LSF than HSF scenes (mean RMS  $\pm$  SD: LSF:  $13.29 \pm 3.95$  candela/m<sup>2</sup>; HSF:  $3.23 \pm 0.95$  candela/m<sup>2</sup>). Hybrid stimuli were created by combining the LSF component in one scene with the HSF component in another scene. The two scenes which made up the hybrids could be either semantically similar (SS; e.g., LSF component of a natural scene superimposed with HSF component of another natural scene) or semantically dissimilar (SD; e.g., LSF component of a natural scene superimposed with HSF component of a man-made scene). Furthermore, the two scenes composing the hybrid could be either physically similar (PS) or physically dissimilar (PD). There were, therefore, four hybrid conditions (SSPS, SSPD, SDPS, and SDPD, see Figure 1) and 160 hybrid stimuli (40 per hybrid condition, see Supplemental data 1–4 for illustration of all stimuli). The 160 filtered scenes used to create the hybrids were also used in a separate experimental session where they were presented individually (either in LSF or HSF, see Figure 1(a)). Stimuli were displayed using E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, USA) on a 17-inch monitor, with a resolution of 1024  $\times$  768 pixels, at a refresh rate of 80 Hz and with a viewing distance of 73 cm. In order to maintain the distance and the central position, participants' heads were supported by a chinrest.

All participants underwent two successive experimental sessions. In Session 1, they were presented with the individual LSF and HSF filtered scenes that were used to build up the hybrids. They were divided into two groups ( $N=26$  each) for which scenes were presented for either 30 or 100 ms. LSF and HSF filtered scenes were randomly presented to the participants. Participants had to categorize the filtered scenes as natural or man-made. In Session 2, participants were presented with hybrid scenes. Each group of exposure duration was divided into two

groups ( $N=13$  each) for which the attention was either directed on LSF or HSF. In the LSF Attentional group, they were instructed to ignore the HSF scene in hybrids, attend the LSF scene, and categorize it as natural or man-made. In the HSF Attentional group, they were instructed to ignore the LSF scene in hybrids, attend the HSF scene, and categorize it as natural or man-made. In total, there were four different groups according to the exposure duration of stimuli and the attentional task on hybrids (LSF-Attention/30 ms, LSF-Attention/100 ms, HSF-Attention/30 ms, and HSF-Attention/100 ms).

In both experimental sessions, each trial began with a central fixation point that was presented for 500 ms (in order to control the gaze direction to the centre of the screen) immediately followed by the stimulus (individual scene or hybrid) and a backward mask (built with 1/f noise) presented for 30 ms to prevent retinal persistence. All participants were requested to give their categorical answer as quickly and as accurately as possible, by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. Response keys were counterbalanced across participants. The experiment lasted about 20 min and response accuracy and reaction times (in ms) were recorded. Before the experiment, participants underwent a training session to get familiarized with tasks and stimuli.

## Results

### Data analysis

For both sessions we performed two ANOVAs on mean error rates (mER, in %) and mean correct response times (mRT, in ms). The ANOVAs in Session 1 included Attentional group in Session 2 (LSF-Attention, HSF-Attention) and Exposure duration (30 ms, 100 ms) as between-subject factors and Spatial frequency (LSF, HSF) of scenes as within-subject factor. It should be noted that the Attentional group factor was included in the analyses in order to assess whether there were any differences between the two attentional groups in Session 2 (LSF-Attention vs. HSF-Attention) during the processing of LSF and HSF scenes. For Session 2 the ANOVAs included Attentional condition (LSF-Attention, HSF-Attention) and Exposure duration of hybrids (30 ms, 100 ms) as between-subject factors and Semantic similarity



(semantically similar [SS], semantically dissimilar [SD]), and Physical similarity (physically similar [PS], physically dissimilar [PD]) of scenes in hybrids as within-subject factors. Because in some conditions error rates were close to ceiling, the ANOVA on mER for both sessions was performed after an Arcsine square root transformation, in order to reduce ceiling effects and ensure variance homogeneity. For both sessions, RT for each subject's correct response in each condition was trimmed in order to reduce the effect of extreme values, by removing responses inferior and superior to three times the interquartile interval. This resulted in removing an average 1.09% of the trials for Session 1 and 0.46% of the trials for Session 2. Analyses were performed using the Statistica 10.0 software (Statsoft, Tulsa, USA). Further pairwise comparisons were corrected using Bonferroni correction for the number of performed comparisons and corrected  $p$ -values are reported. Effect sizes were estimated by calculating the partial eta-squared ( $\eta^2$ ). The significant level of tests was set at 0.05.

### **Session 1: LSF and HSF filtered scenes**

The ANOVA performed on mER (see Figure 2(a)) revealed a main effect of Exposure duration ( $F(1,48) = 54.24, p < .001, \eta^2 = 0.53$ ). Participants made more errors when scenes were presented for 30 ms (mean  $\pm$  SE:  $16.30 \pm 8.69\%$ ) than during 100 ms ( $4.47 \pm 4.39\%$ ). We also observed a main effect of Spatial Frequency ( $F(1,48) = 11.04, p = .002, \eta^2 = 0.19$ ). Participants made more errors when categorizing HSF ( $12.40 \pm 6.96\%$ ) than LSF scenes ( $8.36 \pm 6.12\%$ ). Importantly, there was also a significant interaction between Exposure duration and Spatial frequency ( $F(1,48) = 6.35, p < .05, \eta^2 = 0.12$ ). Participants made more errors when categorizing HSF than LSF scenes when presented for 30 ms (LSF:  $12.69 \pm 8.18\%$ ; HSF:  $19.90 \pm 9.20\%$ ,  $p < .001$ ), but there was no difference in categorizing LSF and HSF scenes when presented for 100 ms (LSF:  $4.04 \pm 4.07\%$ ; HSF:  $4.90 \pm 4.72\%$ ,  $p = 1$ ). There was no main effect of Attentional group ( $F(1,48) < 1, p = .37$ ) and this factor did not interact with Exposure duration ( $F(1,48) < 1, p = .40$ ), Spatial frequency ( $F(1,48) < 1, p = .81$ ), or their interaction ( $F(1,48) < 1, p = .86$ ). This ensures that differences observed between the categorization of LSF scenes in hybrids by the LSF-Attention group and the categorization of HSF scenes in hybrids in the HSF-Attention

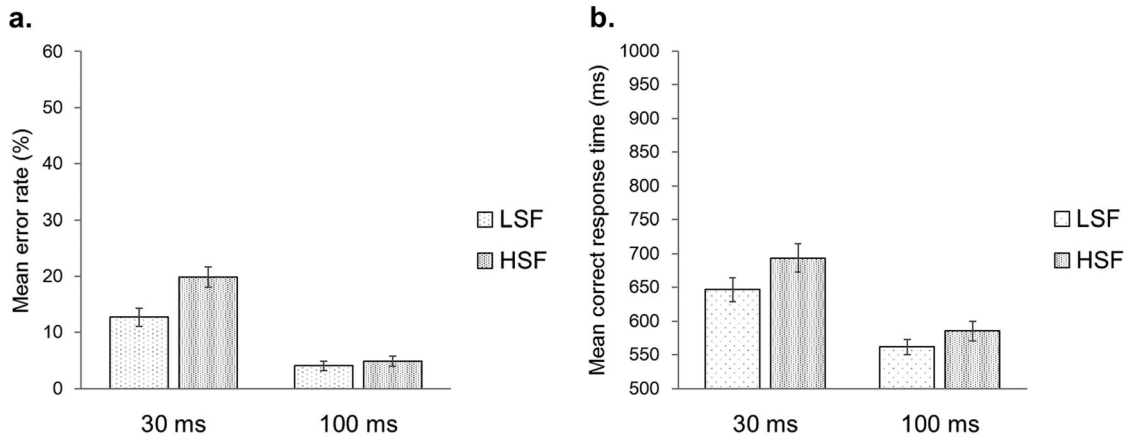
group in Session 2 are not biased by differences between groups in spatial frequency processing in Session 1.

The ANOVA performed on mRT (see Figure 2(b)) revealed a main effect of Exposure duration ( $F(1,48) = 19.56, p < .001, \eta^2 = 0.29$ ) and of Spatial Frequency ( $F(1,48) = 25.41, p < .001, \eta^2 = 0.35$ ). Participants categorized the scenes more quickly when presented for 100 than 30 ms (30 ms:  $670 \pm 97$  ms; 100 ms:  $573 \pm 65$  ms) and when they were filtered in LSF than in HSF (LSF:  $604 \pm 72$  ms; HSF:  $639 \pm 90$  ms). There was no significant interaction between these two factors ( $F(1,48) = 2.72, p = .11$ ). Furthermore, there was no main effect of Attentional group ( $F(1,48) = 2.10, p = .15$ ), this factor did not interact with either Exposure duration ( $F(1,48) = 2.72, p = .37$ ) or with Spatial frequency ( $F(1,48) < 1, p = .93$ ), and the three-way interaction between these three factors was not significant ( $F(1,48) = 1.68, p = .20$ ).

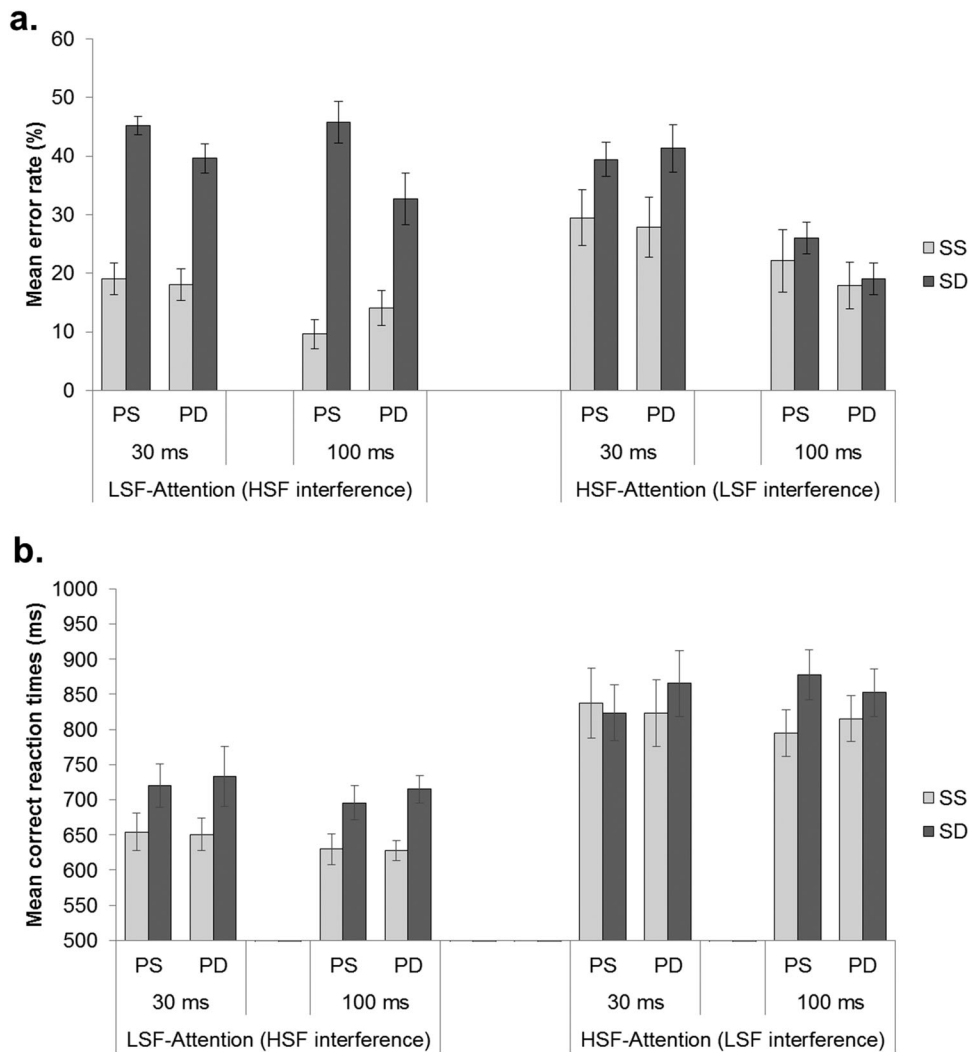
In summary, results of Session 1 showed that participants were able to accurately categorize the filtered scenes, even at short exposure duration. Results also indicated that participants performed better when categorizing LSF scenes than HSF scenes at short exposure duration of 30 ms and were overall faster in categorizing LSF than HSF scenes irrespective of their exposure duration. This suggests a LSF advantage (in terms of accuracy and reaction time) during scene categorization.

### **Session 2: hybrid scenes**

The ANOVA performed on mER (see Figure 3(a)) revealed a main effect of the Semantic similarity between the scenes composing the hybrids ( $F(1,48) = 61.28, p < .001, \eta^2 = 0.56$ ). Participants made more errors when the scenes in hybrids were semantically dissimilar ( $36.13 \pm 13.84\%$ ) than similar ( $19.76 \pm 15.02\%$ ). This indicated that when the two scenes were semantically dissimilar, the unattended scene was however processed at the semantic level and interfered with the categorization of the attended one (semantic interference effect). The main effect of Attentional condition was not significant ( $F(1,48) < 1, p = .77$ ). However, this factor significantly interacted with the Semantic similarity of the two scenes composing the hybrids ( $F(1,48) = 19.57, p < .001, \eta^2 = 0.29$ ). Pairwise comparisons revealed that participants made more errors when the two scenes in hybrids



**Figure 2.** (a) Mean error rates (in %) and (b) mean correct response times (in ms) for the categorization of the individual scenes filtered in LSF and in HSF, for exposure duration of 30 and 100 ms. Error bars indicate standard error.



**Figure 3.** (a) Mean error rates (in %) and (b) mean correct reaction times (in ms) for the categorization of hybrids according to the Semantic similarity of scenes in hybrids (semantically similar-SS, semantically dissimilar-SD), their Physical similarity (physically similar-PS, physically dissimilar-PD), their Exposure duration (30 or 100 ms), and the Attentional condition (LSF-Attention, HSF-Attention). Error bars indicate standard error.

were semantically dissimilar than similar in both the LSF-Attention group (SS:  $15.19 \pm 10.35\%$ ; SD:  $40.82 \pm 11.37\%$ ;  $p < .001$ ) and the HSF-Attention group (SS:  $24.33 \pm 17.54\%$ ; SD:  $31.44 \pm 14.31\%$ ;  $p < .05$ ). Importantly, this semantic interference effect was greater in the LSF (25.63%) than HSF (7.12%) attentional condition group.

We also observed a significant main effect of Exposure duration of hybrid scenes ( $F(1,48) = 12.98$ ,  $p < .001$ ,  $\eta^2 = 0.21$ ). Participants made more errors in categorizing hybrids when they were presented for 30 ms ( $32.50 \pm 12.63\%$ ) than during 100 ms ( $23.39 \pm 14.52\%$ ). However, the expected Attentional condition  $\times$  Exposure duration  $\times$  Semantic similarity interaction was not significant ( $F(1,48) = 2.46$ ,  $p = .12$ ).

Results analysis additionally revealed a significant interaction between the Physical similarity of scenes in hybrids, their Semantic similarity, and the Attentional condition ( $F(1,48) = 9.77$ ,  $p < .005$ ,  $\eta^2 = 0.17$ ), due to a significant interaction between Physical and Semantic similarity in the LSF-attention condition ( $p < .001$ ), but not in the HSF-attention condition ( $p = 1$ ). Pairwise comparisons revealed that when participants had to attend the LSF scene in hybrids, the semantic interference effect was greater when the two scenes in hybrids were physically similar than dissimilar (SSPS =  $14.33 \pm 10.45\%$ ; SDPS =  $45.48 \pm 9.72\%$ ;  $p < .001$ ; SSPD =  $16.06 \pm 10.25\%$ ; SDPD =  $36.15 \pm 13.02\%$ ;  $p < .001$ ).

It should be noted that there was also a significant interaction between the Exposure duration of scenes in hybrids, their Semantic similarity, and their Physical similarity ( $F(1,48) = 9.67$ ,  $p < .05$ ,  $\eta^2 = 0.17$ ). Pairwise comparisons showed that, for exposure duration of 100 ms, semantic interference effect was greater when the two scenes were physically similar than dissimilar ( $p < .001$ ), whereas there was no interaction between semantic and physical similarity of hybrids for exposure duration of 30 ms ( $p = 1$ ). Finally, the four-way Attentional group  $\times$  Exposure duration  $\times$  Semantic similarity  $\times$  Physical similarity interaction was not significant ( $F(1,48) < 1$ ,  $p = .36$ ).

The ANOVA performed on mRT (see Figure 3(b)) revealed a main effect of the Semantic similarity between the two scenes composing the hybrids ( $F(1,48) = 56.65$ ,  $p < .001$ ,  $\eta^2 = 0.54$ ). Response times were slower when the two scenes composing hybrids were semantically dissimilar ( $785 \pm 142$  ms)

than similar ( $729 \pm 147$  ms), suggesting a semantic interference effect on response times. There was also a main effect of the Attentional condition ( $F(1,48) = 24.10$ ,  $p < .001$ ,  $\eta^2 = 0.33$ ). Participants categorized the attended scene faster when it was in LSF ( $678 \pm 93$  ms) than in HSF ( $836 \pm 143$  ms) attentional condition. Importantly, there was a significant interaction between these two factors ( $F(1,48) = 6.65$ ,  $p < .05$ ,  $\eta^2 = 0.12$ ). Pairwise comparisons revealed that participants categorized the scenes more slowly in the semantically dissimilar than similar condition, but this semantic interference effect was greater in the LSF-Attention condition (SS:  $641 \pm 78$  ms; SD:  $716 \pm 108$  ms,  $p < .001$ ) than in the HSF-Attention condition (SS:  $818 \pm 147$  ms; SD:  $855 \pm 140$  ms,  $p < .005$ ).

The expected Semantic similarity  $\times$  Exposure duration  $\times$  Attentional group was not significant ( $F(1,48) = 2.11$ ,  $p = .15$ ). However, we observed a significant four-way interaction between the Semantic similarity of scenes, their Physical similarity, their Exposure duration, and the Attention condition ( $F(1,48) = 5.99$ ,  $p < .05$ ). Pairwise comparisons revealed that when participants had to categorize the LSF scene in hybrids (LSF-Attention), scenes were categorized faster in the semantically similar than dissimilar condition for each Physical similarity condition (Physically similar and Physically dissimilar) of each Exposure duration condition (30 ms and 100 ms; all  $ps < .05$ ). However, when participants had to categorize the HSF scene in hybrids (HSF-Attention), they categorized the scenes faster when they were semantically similar than dissimilar only for exposure duration of 100 ms and when the two scenes were physically similar (100 ms PSSS:  $695 \pm 119$  ms; 100 ms PSSD:  $877 \pm 129$  ms;  $p < .001$ ; all other  $ps > .05$ ).

Finally, to assess whether performances in categorizing LSF and HSF filtered scenes in Session 1 are linked to performances during categorization of hybrids (Session 2), we performed Pearson correlation analyses for mER and RTs between (1) categorization of LSF scenes in Session 1 and categorization of hybrids in the LSF-Attention condition in Session 2 and (2) categorization of HSF scenes in Session 1 and categorization of hybrids in the HSF-Attention condition in Session 2. Accuracy and reaction times for categorizing LSF scenes in Session 1 correlated positively with the ones for categorizing LSF scenes in hybrids in Session 2, irrespective of semantic similarity, physical similarity, and duration (mER:  $r = 0.64$ ,

$p < .001$ ; mRTs:  $r = 0.45$ ,  $p = .021$ ). Similarly, accuracy and reaction times for categorizing HSF scenes in Session 1 correlated positively with the ones for categorizing HSF scenes in hybrids in Session 2, irrespective of semantic similarity, physical similarity, and duration (mER:  $r = 0.53$ ,  $p = .006$ ; mRTs:  $r = 0.49$ ,  $p = .011$ ).

## Discussion

The present behavioural study aims at examining how information from low and high spatial frequencies interact and influence each other during scene categorization. To this end, we presented individual scenes and hybrids to participants who had to attend and categorize their LSF or HSF content. We manipulated the exposure duration of the stimuli which were presented for 30 or 100 ms. For hybrids stimuli, we also manipulated the semantic and physical similarity between the two scenes composing the hybrids. We used the semantic interference effect (i.e., poorer performance when the two scenes in hybrids are semantically dissimilar than similar) as the signature of the integration of semantic information from both spatial frequencies and as an estimate of their relative weight for scene categorization. Critically, this experimental paradigm allowed us to examine how information from a spatial frequency band that is not supposed to be explicitly processed implicitly influences processing of another spatial frequency content. Based on the hypothesis of a predominant coarse-to-fine processing sequence during scene categorization, we expected that information from LSF would weight more at short exposure duration (i.e., stronger interference of LSF over HSF scene categorization than the other way around), whereas information from HSF would take over with increasing exposure duration (i.e., stronger interference of HSF over LSF scene categorization than the other way around).

### *Low spatial frequency processing advantage*

Results on individual LSF and HSF scenes (Session 1) revealed that participants were able to accurately categorize the LSF and HSF scenes (average error rate below 20%) when presented alone at all exposure durations. This allowed us to ensure that the semantic information contained in the LSF and HSF scenes

used in hybrids could be accurately extracted even at short exposure duration of 30 ms. Furthermore, we observed that LSF scenes were categorized more quickly than HSF scenes irrespective of the exposure duration of scenes, and more accurately than HSF scenes for the 30 ms condition only. These results are consistent with previous findings of a temporal precedence of LSF over HSF scene categorization observed in numerous studies using a simple categorization task of LSF and HSF stimuli (De Cesarei & Loftus, 2011; Kauffmann, Ramanoël, et al., 2015; Loftus & Harley, 2005; Parker et al., 1996). A similar pattern of results was found in the Hybrid experimental session (Session 2). Participants were also faster to categorize the attended scene when it was in LSF (LSF-Attention condition) than in HSF (HSF-Attention condition), irrespective of the semantic similarity of the unattended scene. Furthermore, performance for categorizing LSF and HSF individual scenes in Session 1 was positively correlated with the one for categorizing respectively LSF and HSF scenes in hybrids in Session 2. In other words, performances were consistent between Session 1 and Session 2 for the spatial frequency band that was explicitly processed and the LSF advantage observed in Session 1 remained in Session 2. Taken together, these results suggest that LSF information allows for accurate and rapid scene categorization. Furthermore, they support the view that, within a short amount of time, categorization is more efficient when based on LSF than HSF information (Schyns & Oliva, 1994), in accordance with a coarse-to-fine processing strategy. It is important to note that the cut-off frequency used to filter LSF scenes in the present study (4 cpd, i.e., 24 cpi) was relatively high as compared to the ones used in previous studies (usually around 2 cpd; see Kauffmann, Bourgin, et al., 2015; Mu & Li, 2013; Schyns & Oliva, 1994). LSF scenes in the present study thus included a rather large part of the scene spatial frequency spectrum including low to intermediate spatial frequencies and therefore contained the most diagnostic features for scene and object categorization, which have been found to lie at 0–4 cpd (see Caplette et al., 2014; McCotter et al., 2005). Therefore, it is likely that the advantage for LSF processing observed in the present study also reflects the fact that the LSF scenes contained the most relevant information for categorization. It should also be noted that, in the present study, contrast was not equalized between

LSF and HSF filtered scenes. Because in natural scenes, luminance contrast typically decreases with increasing spatial frequencies (Field, 1987), LSF scenes were characterized by a higher contrast than HSF scenes. As it has been shown that reaction times decrease as contrast increases (Harwerth & Levi, 1978), it is therefore possible that the LSF advantage observed in the present study also results in part from a difference in contrast between LSF (mean displayed RMS contrast: 13.29 cd/m<sup>2</sup>) and HSF (mean displayed RMS contrast: 3.23 cd/m<sup>2</sup>) filtered scenes. The relative contribution of differences in contrast between spatial frequencies for a coarse-to-fine advantage during rapid scene categorization was the focus of a recent study conducted by our group (Kauffmann, Chauvin, et al., 2015). We used a design allowing us to examine separately the effect of spatial frequencies, contrast, and their combination, while participants performed a categorization task on sequences depicting a coarse-to-fine or a fine-to-coarse processing. Our results revealed an advantage for categorizing coarse-to-fine sequences relative to fine-to-coarse sequences. Importantly, we observed that, although this advantage was in part driven by differences in contrast between spatial frequencies, it predominantly relied on differences in terms of spatial frequencies.

### ***Spatial frequency semantic interference***

Results in the Hybrid experimental session (Session 2) additionally revealed that the categorization of scenes was impaired when the two scenes in hybrids were semantically dissimilar than similar. This result indicates that, although it was irrelevant to the task, the semantic information contained in the unattended scene in hybrids was processed and interfered with the categorization of the attended one. This semantic interference effect thus suggests the integration of semantic information contained in both scenes in hybrids. Furthermore, as expected, the semantic interference effect was modulated by the attentional constraints, i.e., by the spatial frequency content of the attended/unattended scenes in hybrids. Consistent with previous studies (Kauffmann, Bourgin, et al., 2015; Mu & Li, 2013), participants exhibited a semantic interference of LSF information on the categorization of attended HSF scenes in hybrids (HSF-Attention condition). In an original way, the present study also

revealed a semantic interference of HSF information on the categorization of attended LSF scenes (LSF-Attention condition), and that the semantic interference effect was greater in the LSF than HSF attentional condition, suggesting that information from HSF interfered more strongly with LSF scene categorization than the other way around. These results indicate that, when there is a semantic conflict within the stimulus, information from HSF weights more than LSF information for achieving scene categorization. We did not find the expected modulation of the semantic interference effect by the exposure duration of stimuli. Results rather indicated that the semantic interference effect was always stronger in the LSF than HSF attentional condition, irrespective of the exposure duration of stimuli.

This last result thus partially contradicts our hypotheses. Indeed, we expected that LSF would weight more than HSF at short exposure duration, and that the weight of HSF would increase with exposure duration. Although our results were not consistent with these predictions, we however believe that they do not necessarily speak against a predominant coarse-to-fine processing strategy during scene categorization. First, numerous studies have previously shown that, although LSF information is preferentially used at short exposure duration to enable efficient and rapid scene categorization, it does not follow that it is always used first to support visual recognition. Indeed, and as suggested by our results, HSF can also be extracted very early in the visual processing time-course, and their preferential use would be determined by the demands of the visual task (Campagne et al., 2016; Harel & Bentin, 2009; Morrison & Schyns, 2001; Oliva & Schyns, 1997; Ozgen et al., 2006; Rotshstein et al., 2010; Schyns, 1998; Schyns & Oliva, 1994, 1999). For example, in Schyns and Oliva (1994), the perception of hybrids at short exposure duration of 30 ms was dominated by the HSF scene in 28% of the trials. Further studies from the same authors (Schyns & Oliva, 1999) showed that, during the categorization of hybrid faces presented for 50 ms, HSF were preferentially used to determine whether the face was expressive or not whereas LSF were preferentially used to determine the nature of the emotion. Overall, these results suggested that all spatial frequencies are available very early for categorization,

and that their selection may depend on interactions between the perceptual information available and the demands of a given visual task, thus speaking against a strictly coarse-to-fine processing sequence. Interestingly, it is very likely that the relatively greater interference of HSF information observed in the present study reflects such flexible use of spatial frequency information according to the task constraints – in our case the presence of an apparent conflict within the stimulus. Indeed, when there was no ambiguity within the stimulus (LSF and HSF scenes in Session 1), participants performed better when categorizing LSF than HSF scenes, especially at short exposure duration. As previously mentioned, this indicates that, by default, processing of LSF might be more efficient than that of HSF to rapidly access the scene category. However, when semantic information contained in LSF and HSF was incongruent (semantically dissimilar hybrids in Session 2), participants were interfered more strongly by the unattended scene when it was in HSF (LSF-Attention) than in LSF (HSF-Attention). As previously mentioned, the LSF scenes in hybrids had a higher contrast than the HSF scenes and included intermediate spatial frequencies diagnostic for categorization. The greater semantic interference in the LSF-Attention condition thus suggests a critical and irrepressible influence of HSF information for categorization, despite a priori more salient and relevant information contained in the LSF scene. Overall, results suggest that when there is uncertainty or ambiguity about the category of the visual stimulus, the processing of fine details contained in HSF might be the most relevant and reliable to disentangle between alternative interpretations and would thus strongly weight for scene categorization. It should be noted as a limitation of the present study that due to the hybrid methodology employed, potentially important spatial frequencies were not considered in the stimuli (i.e., spatial frequencies of 4–6 cpd). It is thus possible that these spatial frequencies also weight for scene categorization. Further studies explicitly manipulating the cut-off frequencies of filtered scenes in hybrids would allow us to provide further insight on that question and refine the present results. It is also noteworthy that overall, participants took longer to categorize the attended scene in hybrids when it was in HSF (HSF-Attention) than in LSF (LSF-Attention), irrespective of the semantic similarity of the

unattended scene. This suggests that fast categorization is more efficient when based on LSF and that more time is needed to process and categorize HSF. It is thus possible that the semantic interference of LSF over HSF categorization in the HSF-Attention condition might have been less evident due to initially poorer performances in that condition.

### *Effect of physical similarity*

Our results additionally revealed that the semantic interference effect was modulated by the physical similarity between the two scenes composing the hybrids. In particular, the semantic interference effect was greater when the two scenes in hybrids were physically similar than dissimilar, especially for exposure durations of 100 ms. In their study, Mu and Li (2013) also reported such a modulation of the semantic interference effect by the physical similarity of scenes in hybrids, suggesting that the spatial overlap between LSF and HSF caused even greater impairment of scene categorization in the absence of any semantic congruence between the two scenes. Indeed, it is likely that, in the physically similar condition, the spatial overlap between the LSF and HSF scenes in hybrids might have hindered perceptual categorization performed on the basis of the attended scene physical properties. This could have resulted in enhanced semantic processing of the scenes in order to perform the task, especially when more time was available, resulting in even longer reaction times and error rates when the two scenes were semantically dissimilar. This would indicate that the processing of physical properties of LSF and HSF in scenes plays an important role and actively interacts with the processing of their semantic content over the time-course of scene categorization.

### *Impact for neurobiological models of scene perception*

The semantic interference of LSF over HSF scene categorization was previously interpreted in the context of the proactive model of visual recognition proposed by Bar and colleagues (Bar, 2003, 2007; Bar et al., 2006; Kveraga, Boshyan, & Bar, 2007; Trapp & Bar, 2015). According to this model, fast processing of LSF information allows us to generate coarse predictions about the nature of the visual input in the

orbitofrontal cortex. These predictions would enable us to reduce the number of alternative interpretations of the stimulus and thereby tune the bottom-up processing of HSF via top-down influence to the inferotemporal cortex. Critically, this model thus assumes that it is the bottom-up processing of HSF information that is ultimately used to refine the coarse predictions and achieve recognition. Results of the present study, however, revealed that LSF information interfered less strongly with HSF categorization (HSF-Attention condition) than did HSF information with LSF categorization (LSF-Attention condition). In line with the above-mentioned interpretation, this may suggest that, in case of incongruence between LSF and HSF semantic information, it might be difficult and counterproductive to ignore the evidence accumulated in HSF against LSF-based coarse predictions. In that sense, the semantic interference of LSF over HSF processing would reflect the effect of erroneous top-down LSF-based predictions (Kauffmann, Bourgin, et al., 2015), whereas the semantic interference of HSF over LSF would rather reflect persistent and overtaking bottom-up integration of HSF information against these rapid LSF-based predictions.

## Conclusions

To conclude, results of the present study support the current model of visual perception (Bar, 2003; Bullier, 2001; Kauffmann et al., 2014; Peyrin et al., 2010) suggesting that, during scene categorization, fast processing of coarse LSF information allows us to efficiently access the scene category and influence further processing of fine HSF information. Our results further allow us to refine them, by showing that semantic processing of HSF information can also occur very early and override processing of LSF information when there is a semantic conflict within the stimulus, thereby highlighting their critical role in ultimately mediating scene categorization. Overall, results of the present study indicate that the relative weight of LSF and HSF information during scene categorization varies according to the task constraints and the properties of the visual stimulus.

## Note

1. We conducted a pilot study on 29 participants (1) to replicate the results of Mu and Li (2013) and Kauffmann,

Bourgin, et al. (2015) when participants attend and categorize the HSF scene in hybrids, but also (2) to consider the reverse attentional condition when participants attend and categorize the LSF scene in hybrids. We removed spatial frequencies content above 2 cpd (vs. 4 cpd in the present study). This value was chosen based on Schyns and Oliva's (1994) study on spatial frequency processing during scene categorization. Contrast was either not modified (LUM filtered scenes) or equalized between LSF and HSF to obtain a RMS contrast of 0.1, i.e., 25.6 on a grey-level scale (RMS filtered scenes) using the MATLAB image processing toolbox. Hybrid stimuli were then created by combining a LSF and a HSF scene for which the contrast was not modified (LUM hybrids) and by combining a LSF and a HSF scene for which the contrast was equalized (RMS hybrids). LUM and RMS hybrids were presented in distinct experimental blocks (counterbalanced between participants). Fourteen participants had to attend and categorize the HSF scene in hybrids and 15 participants had to attend and categorize the LSF scene in hybrids for both the LUM and RMS conditions. Results showed that the mean error rate of the SDPS condition of the LSF Attentional group was equal to  $50\% \pm 10$  for LUM hybrids (not significantly different from the chance level;  $t(14) = 0$ ,  $p = 1$ ) and to  $64\% \pm 20$  for RMS hybrids (significantly below chance;  $t(14) = 2.77$ ,  $p < .05$ ). These results suggest that participants were not able to categorize the LSF scene in hybrids and when the contrast of the HSF scene was increased in RMS hybrids. It should be noted that contrast normalization induces severe modifications in the amplitude spectrum properties of scenes that may bias behavioural responses. For example, the enhancement of contrast in HSF scenes using RMS normalization induces an additional unnatural LSF content that could directly mask the LSF scene of hybrids. Therefore, based on these pilot study's results, the present study was conducted without RMS contrast normalization between LSF and HSF scenes, and we chose to increase the spatial frequency cut-off of LSF scenes at 4 cpd. Using these values, participants had performances above chance in the SDPS condition of the LSF Attentional group ( $45.48\% \pm 9.72$ ,  $t(25) = -2.37$ ,  $p < .05$ ). We also included a first experimental session to ensure that participants were able to categorize filtered scenes when they were not embedded in hybrids and independently of attentional constraints.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Agence Nationale de la Recherche grant (ANR-12-JHS2-0002-01 RECOR).

## References

- Abrams, J., Barbot, A., & Carrasco, M. (2010). Voluntary attention increases perceived spatial frequency. *Attention, Perception & Psychophysics*, 72(6), 1510–1521. doi:10.3758/APP.72.6.1510
- Awasthi, B., Sowman, P. F., Friedman, J., & Williams, M. A. (2013). Distinct spatial scale sensitivities for early categorization of faces and places: Neuromagnetic and behavioural findings. *Frontiers in Human Neuroscience*, 7, 91. doi:10.3389/fnhum.2013.00091
- 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. doi:10.1016/j.tics.2007.05.005
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2), 449–454. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1326160&tool=pmcentrez&rendertype=abstract>
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36(2-3), 96–107. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11690606>
- Campagne, A., Fradcourt, B., Pichat, C., Baciú, M., Kauffmann, L., & Peyrin, C. (2016). Cerebral correlates of emotional and action appraisals during visual processing of emotional scenes depending on spatial frequency: A pilot study. *Plos One*, 11(1), e0144393. doi:10.1371/journal.pone.0144393
- Caplette, L., West, G., Gomot, M., Gosselin, F., & Wicker, B. (2014). Affective and contextual values modulate spatial frequency use in object recognition. *Frontiers in Psychology*. doi:10.3389/fpsyg.2014.00512
- 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. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15971697>
- De Cesare, A., & Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychonomic Bulletin & Review*, 18(5), 840–847. doi:10.3758/s13423-011-0133-6
- De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545–559.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12), 2379–2394.
- Fradcourt, B., Peyrin, C., Baciú, M., & Campagne, A. (2013). Behavioral assessment of emotional and motivational appraisal during visual processing of emotional scenes depending on spatial frequencies. *Brain and Cognition*, 83(1), 104–113. doi:10.1016/j.bandc.2013.07.009
- Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: Implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1264–1273. doi:10.1037/a0013621
- Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision Research*, 18(11), 1579–1586. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24333686>
- Hegd , J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, 84(4), 405–439. doi:10.1016/j.pneurobio.2007.09.001
- Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, 8(3), 197–230. doi:10.1162/jocn.1996.8.3.197
- Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The neural bases of the semantic interference of spatial frequency-based information in scenes. *Journal of Cognitive Neuroscience*, 1–10. doi:10.1162/jocn
- 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. doi:10.1016/j.visres.2014.11.013
- Kauffmann, L., Ramano l, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, 112, 86–95. doi:10.1016/j.neuroimage.2015.02.058
- Kauffmann, L., Ramano l, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, 8(37), 1–14. doi:10.3389/fnint.2014.00037
- Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of Neuroscience*, 27(48), 13232–13240. doi:10.1523/JNEUROSCI.3481-07.2007
- Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review*, 12(1), 43–65. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15948283>
- Maunsell, J., Ghose, G. M., Assad, J. A., Mcadams, C. J., Boudreau, C. E., & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual Neuroscience*, 16, 1–14.
- McCotter, M., Gosselin, F., Sowden, P., & Schyns, P. (2005). The use of visual information in natural scenes. *Visual Cognition*, 12(6), 938–953. doi:10.1080/13506280444000599
- Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, 8(3), 454–469. doi:10.3758/BF03196180
- Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental Brain Research*, 227(3), 367–377. doi:10.1007/s00221-013-3517-1
- Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *Plos One*, 7(6), e38493. doi:10.1371/journal.pone.0038493
- Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the



- perception of complex visual stimuli. *Cognitive Psychology*, 34, 72–107. doi:0010-0285/97
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145–175.
- Ozgen, E., Payne, H. E., Sowden, P. T., & Schyns, P. G. (2006). Retinotopic sensitisation to spatial scale: Evidence for flexible spatial frequency processing in scene perception. *Vision Research*, 46(6-7), 1108–1119. doi:10.1016/j.visres.2005.07.015
- Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, 21(2), 147–160. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1513664>
- Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. *Journal of Experimental Psychology: Human Perception and Performance*, 22(6), 1448–1466. doi:10.1037/0096-1523.22.6.1448
- 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. doi:10.1162/jocn.2010.21424
- Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative Ophthalmology*, 11(5), 368–377. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4623891>
- Rotshtein, P., Schofield, A., Funes, M. J., & Humphreys, G. W. (2010). Effects of spatial frequency bands on perceptual decision: It is not the stimuli but the comparison. *Journal of Vision*, 10(10), 25–25. doi:10.1167/10.10.25.Introduction
- Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition*, 67(1-2), 147–179. doi:10.1016/S0010-0277(98)00016-X
- 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. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10193048>
- Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, 42(22), 2547–2554. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12445848>
- Shapley, R., & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience*, 8(1978), 547–581.
- Sowden, P. T., Özgen, E., Schyns, P. G., & Daoutis, C. (2003). Expectancy effects on spatial frequency processing. *Vision Research*, 43(26), 2759–2772. doi:10.1016/S0042-6989(03)00480-2
- Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Sciences*, 10, 538–545. doi:10.1016/j.tics.2006.10.007
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522.
- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, 1339(1), 190–198. doi:10.1111/nyas.12680
- Van Essen, D., & Deyoe, E. A. (1995). Concurrent processing in the primate visual cortex. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 383–400). Cambridge: Bradford Book.
- Vannucci, M., Viggiano, M. P., & Argenti, F. (2001). Identification of spatially filtered stimuli as function of the semantic category. *Cognitive Brain Research*, 12(3), 475–478. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11689308>