Mécanismes et bases cérébrales du traitement des fréquences spatiales lors de la catégorisation de scènes visuelles

Louise Kauffmann

▶ **To cite this version:**

Louise Kauffmann. Mécanismes et bases cérébrales du traitement des fréquences spatiales lors de la catégorisation de scènes visuelles. Psychologie. Université Grenoble Alpes, 2015. Français. NNT : 2015GREAS047 . tel-01693648

HAL Id: tel-01693648
https://theses.hal.science/tel-01693648

Submitted on 26 Jan 2018

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.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Sciences cognitives, Psychologie et Neurocognition**

Arrêté ministériel : 7 août 2006

Présentée par :

Louise KAUFFMANN

Thèse dirigée par **Carole PEYRIN** et **Alan CHAUVIN**

Préparée au sein du **Laboratoire de Psychologie et NeuroCognition (CNRS UMR 5105)**

Dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

MECANISMES ET BASES CEREBRALES DU TRAITEMENT DES FREQUENCES SPATIALES LORS DE LA CATEGORISATION DE SCENES VISUELLES

Thèse soutenue publiquement le **4 novembre 2015**,

devant le jury composé de :

Pr. Nicolas POIREL
Professeur, Université Paris Descartes, France. *Rapporteur*.
Dr. Emmanuel MELLET
Directeur de recherche CNRS, Université de Bordeaux, France. *Rapporteur.*
Dr. Valérie GOFFAUX
Chercheur, Université Catholique de Louvain, Belgique. *Examinatrice.*
Pr. Martial MERMILLOD
Professeur, Université Grenoble-Alpes, France*. Examinateur, Président.*
Dr. Carole PEYRIN
Chargée de Recherche CNRS, Université Grenoble-Alpes, France.
Directrice de Thèse.
Dr. Alan CHAUVIN
Maître de conférences, Université Grenoble-Alpes, France.
Directeur de Thèse

Université Joseph Fourier / Université Pierre Mendès France /
Université Stendhal / Université de Savoie / Grenoble INP

RESUME

L'analyse visuelle de scènes débute par l'extraction en parallèle de l'information à différentes fréquences spatiales, en suivant un mode de traitement par défaut « coarse-to-fine ». L'analyse rapide de l'information grossière (« coarse ») en basses fréquences spatiales fournirait un aperçu global de la scène, qui serait ensuite affiné par l'analyse des détails de la scène (« fine ») en hautes fréquences spatiales. L'objectif de cette thèse a été de préciser les bases cérébrales du traitement des fréquences spatiales lors de la catégorisation de scènes. A travers deux études comportementales, nous avons tout d'abord montré qu'une analyse « coarse-to-fine » est plus avantageuse pour la catégorisation rapide de scènes, et ce, indépendamment de la valeur de contraste de luminance associée aux différentes fréquences spatiales (Expériences 1 et 2). Des études en IRMf nous ont par la suite permis de mettre en évidence l'implication d'un large réseau cérébral lors de l'analyse « coarse-to-fine » des scènes, incluant les aires visuelles primaires et occipito-temporales, mais également le cortex frontal inférieur (Expérience 3). Une analyse de la connectivité a révélé que lors de cette analyse, le cortex frontal inférieur exercerait une influence « top-down » sur le cortex visuel primaire et les gyri fusiforme et parahippocampique au sein du cortex occipito-temporal. Ces résultats soulignent le rôle du cortex visuel primaire comme région intégrative, codant à la fois les afférences rétino-thalamiques et les influences « top-down » de régions supérieures. Nous avons également observé que le gyrus frontal inférieur et le gyrus fusiforme participaient activement à l'intégration de l'information sémantique contenue dans les basses et hautes fréquences spatiales d'une scène (Expérience 4). Enfin, nous avons spécifiquement étudié le traitement des fréquences spatiales au sein de régions occipito-temporales sélectives aux scènes : la « parahippocampal place area » (PPA), le cortex retrosplenial et l'« occipital place area ». Nous avons montré que ces trois régions participent de façon distincte au traitement des fréquences spatiales dans les scènes (Expérience 5) et qu'une stratégie d'analyse « coarse-to-fine » serait privilégiée par la PPA (Expérience 6). Les résultats de ces travaux nous permettent de conforter et de préciser les modèles actuels de la catégorisation visuelle de scènes basés sur un traitement fréquentiel de l'information visuelle.

Mots clés : Scènes naturelles, Fréquences spatiales, Coarse-to-fine, IRMf, Cortex visuel, Cortex frontal inférieur, Cortex inférotemporal, Régions sélectives aux scènes.

ABSTRACT

Visual analysis begins with the parallel extraction of different attributes at different spatial frequencies following a predominantly coarse-to-fine default processing sequence. Rapid processing of low spatial frequency information would permit a coarse parsing of the visual input, prior to the detailed analysis of fine information in high spatial frequencies. Our aim was to further address the neural bases of spatial frequency processing during scene categorization. We first demonstrated in two behavioral studies that a coarse-to-fine processing is indeed an advantageous strategy for rapid scene categorization, and is independent of the luminance contrast values associated with the different spatial frequencies (Experiments 1 and 2). In two fMRI studies, we showed first the involvement of a large cerebral network during coarse-to-fine processing of scenes, including early visual and occipito-temporal areas, but also the inferior frontal cortex. Effective connectivity analysis revealed that the inferior frontal gyrus exerts top-down influence on the early visual cortex as well as on the parahippocampal and fusiform gyri in the occipito-temporal cortex (Experiment 3). These results highlight the role of the primary visual cortex in integrating top-down influences from frontal areas to retino-thalamic incoming signals. We also evidenced that the inferior frontal and fusiform gyri actively participate to the integration of the semantic information contained in low and high spatial frequency (Experiment 4). Finally, we specifically investigated the spatial frequency processing of scenes within scene-selective areas of the occipito-temporal cortex: the parahippocampal place area (PPA), the retrosplenial cortex, and the occipital place area. We demonstrated that these regions participate differently in the spatial frequency processing of scenes (Experiment 5) and that a coarse-to-fine processing is favored within the PPA (Experiment 6). Overall, results allow us to refine current model of visual scene categorization based on a spatial frequency analysis.

Keywords: Natural scenes, Spatial frequencies, Coarse-to-fine, fMRI, Visual cortex, Inferior frontal cortex, Inferotemporal cortex, Scene-selective regions.

REMERCIEMENTS

Merci à Valérie Goffaux, Emmanuel Mellet, Martial Mermillod et Nicolas Poirel d'avoir accepté d'évaluer ce travail de thèse en faisant partie du jury.

Merci à mes deux directeurs de thèse, Alan Chauvin et Carole Peyrin. J'ai eu beaucoup de chance d'effectuer cette thèse dans d'aussi bonnes conditions d'entente et de collaboration avec vous.

Merci Alan pour m'avoir permis en premier lieu de travailler sur ces thématiques de recherches qui me passionnent toujours autant, et pour avoir continué de m'accompagner jusque-là. Merci pour ton aide, ta disponibilité, tes conseils avisés, ta gentillesse et ta bienveillance à mon égard.

Merci Carole pour m'avoir également permis de travailler sur tes thématiques de recherches et de me former à la neuroimagerie. Merci pour ton énergie, ton implication dans ce travail, ta rigueur et ton perfectionnisme, et pour la quantité astronomique d'autres choses que tu m'as apprises, de la rédaction d'articles à l'encadrement de mes futurs étudiants (Maître Yoda).

J'espère sincèrement que j'aurai de nouveau l'occasion de travailler avec vous.

Merci également à tous les étudiants qui ont participé à nos études et à tous ceux qui ont collaboré à ce travail.

Merci Stephen Ramanoël pour toute ton aide et pour avoir permis de rendre les sessions d'acquisitions moins ennuyeuses. Merci pour « la maison des oiseaux » et pour le Bongo Club. Je ne t'en veux pas tant que ça pour les 1000 acquisitions de rétinotopie...

Merci à Benoit Musel dont le travail de thèse a fourni les bases sur lesquelles nous nous sommes appuyés tout au long de ce travail et pour son aide au début de ma thèse.

Un énorme merci à Cédric Pichat, mon mentor de l'IRM. Merci pour ton aide, pour tout ce que tu m'as appris, et pour m'avoir écoutée me plaindre du DCM pendant plus d'un an. Je ne sais pas si tu te rends compte à quel point tu as été indispensable pour ce travail de thèse ! Merci pour ta gentillesse, ton amitié et pour le petit bonhomme en mousse.

Merci Nathalie Guyader pour ton aide précieuse dans de multiples aspects de ce travail et pour la clarté de tes explications.

Merci à Jessica Bourgin pour tout son travail, sur lequel nous nous sommes appuyés pour l'article relatif à l'Expérience 4 de cette thèse.

Merci aussi à Emilie Cousin, Laurent Lamalle, Irène Troprès et Johan Pietras de la plateforme IRMaGE pour leur aide lors de la mise en place et la préparation des acquisitions IRMf.

Je remercie également chaleureusement Guillaume Giraudet, Jocelyn Faubert et toute l'équipe du Laboratoire de Psychophysique et de Perception Visuelle à Montréal, pour m'avoir accueillie lors de mon séjour « Indoc ». Merci pour votre gentillesse, votre aide précieuse et vos conseils. J'espère que le projet que nous avons réalisé ensemble donnera lieu à d'autres collaborations.

Un grand merci au Laboratoire de Psychologie et Neurocognition et à tous ses membres, qui en font un lieu de travail vraiment appréciable. Merci à tous pour votre bonne humeur, votre soutien et pour l'intérêt que vous portez à notre avenir professionnel. Merci aussi pour les Martin's du vendredi et pour vos lobes frontaux qui lâchent de temps en temps. Un merci particulier à Martial, Richard, Elsa et David pour m'avoir donné l'opportunité d'assurer des cours dans le cadre de vos enseignements, et pour m'avoir donné envie de continuer. Un grand merci aussi à Claire Leroy, Guylaine Omnes et Karima Domingues pour votre efficacité et pour nous avoir grandement facilité les choses à plusieurs reprises.

Evidemment, un énorme merci aux doctorants et assimilés, passés (Nico, Benoit M., Benoit F., Sabine, Yanica, Solène, Marcela, Nico, Pauline, Mélanie, Audrey), présents (Fabrice, Lysianne, Laurie, Brice, Fanny, Thierry, Naila, Sylvain, Chloé, Cindy, Morgane, Violette, Prakhar, Faris, Sophie, Romain) et futurs (Amélie, Ladislas), qui ont largement contribué à la bonne réalisation de ce travail. Merci pour tout ce qui a pu se passer dans le 222bis, pour votre soutien et votre aide, et pour

avoir fait de ces trois années un plaisir. C'est quand même une énorme chance de travailler avec des amis. Une pensée aussi pour les anciens du Master : Anouk, Romain et Laura.

Merci à mes amis, pour votre soutien, vos encouragements et pour tout le reste.

Un très grand merci à ma famille. Merci à Odile, Philippe, Juliette, Pierre et Cécile pour votre soutien permanent, votre amour, et pour la fierté que vous exprimez à mon égard. Vous pouvez surtout être fiers de vous. Une pensée particulière à Paule et Jean-Maurice que je n'oublie pas, et à qui j'espère rendre hommage en devenant à mon tour un « Dr. Kauffmann ».

Enfin, le plus grand de tous les mercis à Thomas, qui me (sup)porte depuis maintenant deux mémoires et une thèse. Merci pour tout ce que tu m'apportes, ton amour, ton soutien et pour m'encourager dans ma carrière quoi qu'il arrive. Спасибо.

AVANT-PROPOS

Ce travail de thèse s'inscrit dans un axe de recherche pluridisciplinaire, existant depuis plusieurs années au sein du Laboratoire de Psychologie et Neurocognition (LPNC). Les travaux présentés dans cette thèse ont été réalisés en collaboration avec le laboratoire Grenoble Image Parole Signal Automatique (Gipsa-Lab), et la plateforme IRMaGe au Centre Universitaire Hospitalier de Grenoble. Cette thèse a été effectuée au sein du Laboratoire de Psychologie et Neurocognition (CNRS UMR 5105) et dans l'Ecole Doctorale d'Ingénierie pour la Santé, la Cognition et l'Environnement (EDISCE), grâce à un financement de la Région Rhône-Alpes (Allocation doctorale de recherche du dispositif « Cible »).

L'objet de ce travail de recherche a été de cartographier et modéliser au niveau cérébral les mécanismes sous-jacents à la perception et à la catégorisation de scènes chez le jeune adulte sain. Plus précisément, nous nous sommes attachés à étudier la façon dont le système visuel analyse et utilise une caractéristique élémentaire des scènes, les fréquences spatiales, afin de permettre une catégorisation rapide et fiable de l'environnement. Ce travail de thèse se situe donc dans l'objectif général d'élaborer un modèle neurocognitif biologiquement plausible de la catégorisation de scènes naturelles en utilisant de nouveaux paradigmes expérimentaux construits sur la base des propriétés neurophysiologiques des voies rétino-corticales.

ABREVIATIONS

ANOVA	Analyse de variance
BFS	Basses fréquences spatiales
BMA	*Bayesian model averaging*
BMS	*Bayesian model selection*
BOLD	*Blood oxygenation level dependant*
C	Cuneus
C	Contraste de Michelson
CGL	Corps genouillé latéral
cpd	Cycles par degré
CtF	*Coarse-to-fine*
DCM	*Dynamic causal modelling*
EEG	Electroencéphalographie
F	F de Fisher
FFA	*Fusiform face area*
FG	Gyrus fusiforme
FS	Fréquence spatiale
FSC	Fonction de sensibilité au contraste
FtC	*Fine-to-coarse*
HFS	Hautes fréquences spatiales
IFG	Gyrus frontal inférieur
IRMf	Imagerie par résonance magnétique fonctionnelle
K	Koniocellulaire
LO	Aire latérale occipitale
M	Magnocellulaire
MEG	Magnétoencéphalographie
mER	Taux d'erreur moyen
MFS	Moyennes fréquences spatiales
mRT	Temps de réponse moyens
ms	Millisecondes
MST	Aire temporale médiane supérieure
MVPA	*Multivariate pattern analysis*
MT	Aire temporale médiane
NF	Non filtré
OFC	Cortex orbitofrontal

OPA	*Occipital place area*
P	Parvocellulaire
p	Probabilité statistique sous l'hypothèse nulle
PPA	*Parahippocampal place area*
RMS	*Root mean square*
ROI	*Region of interest*
RSC	Cortex rétrosplenial
t	t de Student
TEP	Tomographie par émission de positrons
TMS	Stimulation magnétique transcrânienne
TR	Temps de réponse
V1	Cortex visuel primaire

TABLE DES MATIERES

REMERCIEMENTS 2
AVANT-PROPOS 4
ABREVIATIONS 5
TABLE DES MATIERES 7
PREAMBULE 10

CHAPITRE 1 - TRAITEMENT DE L'INFORMATION VISUELLE : UNE ANALYSE FREQUENTIELLE 13

1. Traitement de l'information visuelle de la rétine au cortex 13
1.1. La rétine 13
1.2. Les corps genouillés latéraux 14
1.3. Les aires visuelles striées et extra-striées 15
1.4. Les voies dorsale et ventrale 17

2. Modèles neurophysiologiques de l'intégration de l'information visuelle 18
2.1. Modèles hiérarchiques 18
2.2. Modèle par itération 20

3. Traitement des fréquences spatiales 23
3.1. Représentation et manipulation d'images dans le domaine fréquentiel 23
3.2. Les fréquences spatiales et la sensibilité au contraste 26
3.2.1. Les fréquences spatiales dans les réseaux sinusoïdaux 26
3.2.2. Les fréquences spatiales dans la perception globale et locale 27
3.2.3. Le contraste 29

4. Une analyse visuelle selon une séquence « coarse-to-fine » 32
4.1. Arguments neurophysiologiques 32
4.2. Arguments comportementaux 33
4.2.1. Une analyse « coarse-to-fine » résultant de contraintes perceptives ? 33
4.2.2. Une stratégie d'analyse flexible guidée par les contraintes de la tâche 38
4.3. Arguments computationnels 41

5. Problématique 45

CHAPITRE 2 – CATEGORISATION RAPIDE DE SCENES : ETUDE COMPORTEMENTALE DU TRAITEMENT DES FREQUENCES SPATIALES ET INFLUENCE DES PARAMETRES DE LA STIMULATION VISUELLE 49

1. Les scènes dynamiques 49

2. Influence de l'ordre, du mode d'accumulation et du contraste de luminance des fréquences spatiales pour la catégorisation rapide de scènes 50

2.1. EXPERIENCE 1 : Influence de l'ordre et du mode d'accumulation de l'information en fréquences spatiales pour la catégorisation de scènes 52
2.2. EXPERIENCE 2 : Influence du contraste de luminance pour l'avantage d'une analyse « coarse-to-fine » des scènes 54
2.3. Conclusions 75

CHAPITRE 3 - BASES CEREBRALES DE L'ANALYSE « COARSE-TO-FINE » 78

1. Organisation rétinotopique et latéralisée du traitement des fréquences spatiales au sein du cortex visuel 78

2. Bases cérébrales de l'analyse « coarse-to-fine » 84
2.1. Modèles neurocognitifs de la reconnaissance visuelle 84
2.2. EXPERIENCE 3 : Connectivité effective au sein du réseau cérébral impliqué dans l'analyse « coarse-to-fine » de scènes visuelles 91
2.3. EXPERIENCE 4 : Etude de l'influence du traitement des BFS sur le traitement des HFS via l'effet d'interférence sémantique 118
2.4. Conclusions 138

CHAPITRE 4 - TRAITEMENT DES FREQUENCES SPATIALES AU SEIN DES REGIONS OCCIPITO-TEMPORALES SELECTIVES AUX SCENES 141

1. Rôle des régions sélectives aux scènes 141
1.1. Parahippocampal place area ou PPA 141
1.2. Cortex retrosplenial ou RSC 143
1.3. Occipital place area ou OPA 144

2. Nature de l'information représentée au sein des régions sélectives aux scènes 146

3. EXPERIENCE 5 : Traitement des fréquences spatiales et du contraste de luminance au sein des régions sélectives aux scènes 151

4. EXPERIENCE 6 : Prédominance d'une stratégie d'analyse « coarse-to-fine » au sein des régions sélectives aux scènes 175

5. Conclusions 194

CHAPITRE 5 - DISCUSSION GENERALE ET PERSPECTIVES 198

1. Apports méthodologiques 199
1.1 Apport de la manipulation de l'ordre de présentation des fréquences spatiales d'une scène 199
1.2. Intérêt et conséquences de l'égalisation du contraste des scènes filtrées 201

2. Vers l'élaboration d'un modèle neurocognitif de la catégorisation de scènes 205
2.1. Le cortex visuel : région intégrative de l'information visuelle ? 205
2.2. Le cortex frontal inférieur : prédiction, intégration sémantique ou pilotage du traitement visuel ? 209
2.3. Les régions du cortex occipito-temporales : des relais majeurs pour l'analyse de scènes ? 213

3. Limites et perspectives 217
3.1. Précision du modèle via des aspects temporels et anatomiques 217

2.2 Influence de facteurs individuels, développementaux et pathologiques 219

4. Conclusion **223**

RÉFÉRENCES BIBLIOGRAPHIQUES 225

LISTE DES FIGURES 243

LISTE DES ANNEXES I

PREAMBULE

En un clin d'œil, nous sommes capables d'extraire une information suffisante sur l'environnement pour le catégoriser, connaître les objets qui le composent ainsi que leur structuration spatiale et pouvoir agir dessus. Pourtant, les scènes visuelles sont d'une variabilité et d'une complexité infinie si l'on considère les différents points de vue possibles d'une scène, la taille et la position dans l'espace des objets qui varient constamment, les ombres que ces objets projettent, ainsi que le fait qu'ils se masquent souvent les uns les autres. Le simple fait de voir apparaît donc comme une activité cognitive complexe, qui nécessite en permanence de reconstruire l'information pour aboutir à une perception porteuse de sens et cohérente du monde. Il est alors surprenant de constater la rapidité et la fiabilité avec laquelle nous percevons et faisons sens, sans effort, de l'environnement qui nous entoure.

L'objectif de ce travail de thèse a été de préciser les mécanismes visuels qui permettent une reconnaissance aussi rapide et fiable, ainsi que leurs bases cérébrales, lors de la perception et la catégorisation de scènes. Pour cela, nous nous sommes particulièrement intéressés à la façon dont le système visuel traite une caractéristique élémentaire du signal visuel, les fréquences spatiales. En effet, la plupart des travaux issus notamment de la neurophysiologie du système visuel, de la psychologie cognitive et de la neuroimagerie chez le primate non-humain et chez l'humain s'accordent à dire que l'analyse visuelle débute par l'extraction en parallèle des différents éléments de l'environnement visuel à différentes fréquences spatiales, selon un mode de traitement par défaut « coarse-to-fine ». L'analyse rapide de l'information grossière (« coarse ») véhiculée par les basses fréquences spatiales permettrait d'avoir accès à la structure globale de la scène, et serait ensuite affinée par l'analyse de l'information plus détaillée (« fine ») véhiculée par les hautes fréquences spatiales.

Dans le Chapitre 1, nous verrons dans un premier temps comment l'information de l'environnement est traitée par le système visuel, depuis la stimulation lumineuse des cellules de la rétine jusqu'au aires visuelles striées et extra-striées. Nous nous focaliserons tout particulièrement sur la façon dont le système visuel extrait et analyse l'information en fréquences spatiales lors de la perception de scènes. Les différents travaux de la neurophysiologie, de la psychologie cognitive et de modélisations computationnelles

présentés dans ce chapitre nous amènerons à nous interroger sur les mécanismes neurocognitifs à l'œuvre lors du traitement des fréquences spatiales.

Nous présenterons dans le Chapitre 2 deux études comportementales visant à préciser le traitement et l'intégration de l'information en fréquences spatiales lors de la catégorisation de scènes. Ces études ont eu pour objectif d'examiner les facteurs (et dans le cas présent, les paramètres de la stimulation visuelle) pouvant influencer le traitement des fréquences spatiales lors de la catégorisation rapide de scènes, tels que l'ordre de traitement des fréquences spatiales, leur mode d'accumulation par le système visuel, et le contraste de luminance associé aux différentes fréquences spatiales (**Expériences 1 et 2**). Ces deux expériences nous permettrons de sélectionner des stimuli appropriés pour l'étude des bases cérébrales du traitement des fréquences spatiales et de l'analyse « coarse-to-fine » dans le Chapitre 3.

Le Chapitre 3 sera consacré à une brève revue de littérature sur les bases cérébrales du traitement des fréquences spatiales au sein du cortex visuel. Ces travaux montrent notamment qu'au sein du cortex occipital le traitement des fréquences spatiales respecterait l'organisation rétinotopique du cortex visuel, et ce, de façon latéralisée entre les hémisphères. Nous présenterons ensuite les travaux récents qui se sont intéressés aux bases cérébrales d'une analyse « coarse-to-fine » des scènes. Dans la continuité de ces travaux, nous présenterons les résultats de deux études en IRMf que nous avons menées, la première ayant pour objectif de préciser les bases cérébrales de l'analyse « coarse-to-fine » ainsi que la connectivité au sein de ce réseau (**Expérience 3**), et la seconde visant à examiner la façon dont le traitement des basses fréquences spatiales peut influencer le traitement des hautes fréquences spatiales (**Expérience 4**). Nous verrons que les résultats de ces deux études montrent qu'une région du cortex inférotemporal connue pour être sélectivement impliquée dans la perception de scènes, la « parahippocampal place area » serait systématiquement impliquée dans l'analyse « coarse-to-fine » de scènes.

Dans le Chapitre 4, nous nous intéresserons donc plus spécifiquement à cette région, ainsi qu'à deux autres régions du cortex occipito-temporal également sélectives aux scènes. Après avoir rapidement décrit les principales fonctions qui leurs sont attribuées lors de la perception de scènes, nous examinerons à travers deux études en IRMf si ces régions sont également impliquées dans le traitement de l'information en fréquences spatiales contenue dans les scènes (**Expérience 5**) et si un mode de traitement « coarse-to-fine » est privilégié au sein de ces régions (**Expérience 6**).

Enfin, le Chapitre 5 sera consacré à la Discussion générale. Nous résumerons les principaux résultats des études effectuées pour ce travail de thèse et tenterons de les articuler dans un modèle neurocognitif global de la perception et de la catégorisation de scènes.

CHAPITRE 1 - TRAITEMENT DE L'INFORMATION VISUELLE : UNE ANALYSE FREQUENTIELLE

Cette thèse adresse la perception et la catégorisation rapide de scènes à travers l'étude des bases cérébrales du traitement des fréquences spatiales. Cela implique donc de comprendre en premier lieu comment l'information en fréquences spatiales est traitée par le système visuel. L'information lumineuse réfléchie par les objets de l'environnement est tout d'abord projetée sur la rétine. Elle est ensuite transmise, via un relais dans les corps genouillés latéraux, vers le cortex visuel primaire, puis vers des aires visuelles de plus haut niveau pour aboutir à la reconnaissance. Nous proposons donc de rappeler dans un premier temps les principales étapes de traitement de l'information visuelle, de la rétine aux aires visuelles striées et extra-striées.

1. Traitement de l'information visuelle de la rétine au cortex

1.1. La rétine

Le traitement de l'information visuelle débute par la stimulation lumineuse des photorécepteurs, situés dans la partie la plus profonde de la rétine (voir pour une revue Gollisch & Meister, 2010; Masland, 2012). On distingue deux types de photorécepteurs aux propriétés différentes : les cônes et les bâtonnets. Les cônes sont sensibles aux intensités lumineuses élevées et sont à l'origine de la vision en couleurs. Les bâtonnets, dont le nombre est environ 20 fois supérieur à celui des cônes, sont sensibles à de faibles intensités lumineuses et codent une information achromatique. La répartition des cônes et des bâtonnets sur la rétine n'est pas homogène, de telle sorte que le centre de la rétine (fovéa) est principalement composé de cônes, tandis que la zone périphérique de la rétine contient essentiellement des bâtonnets (Osterberg, 1935). En sortie des photorécepteurs, l'information visuelle est transmise aux cellules ganglionnaires via les cellules bipolaires, avant d'être véhiculée au système nerveux central le long du nerf optique. Les cellules ganglionnaires peuvent être classées sur la base de leurs propriétés anatomiques et physiologiques, selon lesquelles on distingue : les cellules de type magnocellulaire (M), les cellules de type

parvocellulaire (P) et les cellules de type koniocellulaire (K)[1], dénommées ainsi en raison de leurs projections respectives au niveau des couches M, P et K de structures en sortie de la rétine : les corps genouillés latéraux. Les cellules de type M (dites aussi « parasol » en raison de leur arborisation dendritique dense et étendue) ont un champ récepteur très étendu, ce qui fait que leur résolution spatiale est faible. Leur vitesse de conduction et par conséquent leur résolution temporelle est par contre très grande du fait de la taille importante du diamètre de leur axone et de la myélinisation de celui-ci. A l'inverse, les cellules de type P (dites « midget » ou naines) possèdent un petit champ récepteur et un axone fin. Les cellules M sont sensibles à des stimulations brèves et transitoires (réponse phasique), ainsi qu'à des stimulations soutenues (réponse tonique) tandis que les cellules P sont uniquement sensibles à des stimulations soutenues (Robert Shapley, Kaplan, & Soodak, 1981). Les cellules M répondent à une information achromatique, à de faibles variations de contraste, aux mouvements, aux basses fréquences spatiales (BFS) et aux hautes fréquences temporelles. A l'inverse, les cellules P répondent à des stimulations chromatiques, aux hautes fréquences spatiales (HFS) et aux basses fréquences temporelles (Curcio, Sloan, Kalina, & Hendrickson, 1990; Enroth-Cugell & Robson, 1966, 1984). Enfin, les cellules ganglionnaires de type M sont plus nombreuses en périphérie de la rétine, tandis que les cellules de type P sont plus nombreuses au centre de la rétine (Azzopardi, Jones, & Cowey, 1999; Curcio & Allen, 1990).

1.2. Les corps genouillés latéraux

Les axones des cellules ganglionnaires s'assemblent en sortie de la rétine pour former le nerf optique. La majorité des fibres du nerf optique (90%) est projetée sur différentes couches du corps genouillé latéral (CGL), situé dans la partie dorsale du thalamus. Les 10% restant sont projetés notamment vers le colliculus supérieur, une structure sous-corticale impliquée, entre autres, dans le contrôle des mouvements oculaires. Avant d'atteindre les CGL, les fibres optiques provenant des deux yeux subissent une décussation au niveau du chiasma optique. L'information visuelle du champ visuel gauche, en provenance de la rétine nasale de l'œil droit et de la rétine temporale de l'œil gauche, est véhiculée jusqu'au CGL droit (puis au cortex visuel primaire droit). De la même façon, l'information visuelle du champ visuel droit, en provenance de la rétine nasale de l'œil gauche et de la rétine temporale de l'œil droit, est véhiculée jusqu'au CGL gauche (puis au cortex visuel primaire gauche).

[1] Les caractéristiques des cellules K ne seront pas détaillées ici, de par leur intérêt moindre pour la suite de ces travaux.

L'information issue d'un hémi-champ visuel est donc envoyée au CGL, puis au cortex visuel primaire, controlatéral. Les afférences des cellules ganglionnaires M et P se projettent respectivement sur les couches M et P du CGL. Ainsi, les caractéristiques de l'information véhiculée par chaque type de cellule sont préservées au niveau de cette structure (Derrington & Lennie, 1984; Hicks, Lee, & Vidyasagar, 1983; Hubel & Wiesel, 1977; Maunsell et al., 1999; Merigan & Maunsell, 1993; Shapley, Kaplan, & Soodak, 1981; Shapley & Lennie, 1985). Les projections des axones des cellules ganglionnaires ont la propriété d'être rétinotopiques. Cela signifie que des points voisins dans le champ visuel sont codés par des cellules voisines de la rétine, et sont projetés sur des sites voisins au niveau du CGL. Ainsi, l'organisation spatiale du champ visuel et des informations rétiniennes est préservée au niveau du CGL. Dans les CGL, la séparation des afférences des cellules ganglionnaires M et P vont être à l'origine de deux voies distinctes jusqu'aux aires visuelles : la voie magnocellulaire (voie M) et la voie parvocellulaire (voie P).

1.3. Les aires visuelles striées et extra-striées

Après un relais au niveau du CGL, l'information visuelle est véhiculée vers le cortex visuel primaire (V1) ou cortex strié, correspondant à l'aire 17 dans la classification de Brodmann. Cette aire se situe dans la partie postérieure du cortex occipital et s'étend sur la surface interne de chaque hémisphère. Comme dans le CGL, un hémi-champ visuel est projeté sur le cortex visuel de l'hémisphère opposé. De plus, on retrouve au niveau de V1 une organisation rétinotopique de l'information visuelle (Daniel & Whitteridge, 1961; Engel et al., 1994; Engel, Glover, & Wandell, 1997). Chaque point voisin du champ visuel est donc projeté sur des régions voisines du cortex visuel (cf. Figure 1a). La partie supérieure du champ visuel se projette sur la partie inférieure du cortex visuel et inversement. Le champ visuel central est projeté sur la partie postérieure latérale du cortex visuel, tandis que le champ visuel périphérique se projette sur la partie antérieure médiale, en avant de la scissure calcarine. Il est à noter que l'information provenant de la fovéa, codant la partie centrale du champ visuel, subit une magnification au niveau cortical, de telle sorte que cette information qui occupe moins de 2% du champ visuel couvre près de 25% de la surface de V1 (cf. Figure 1b).

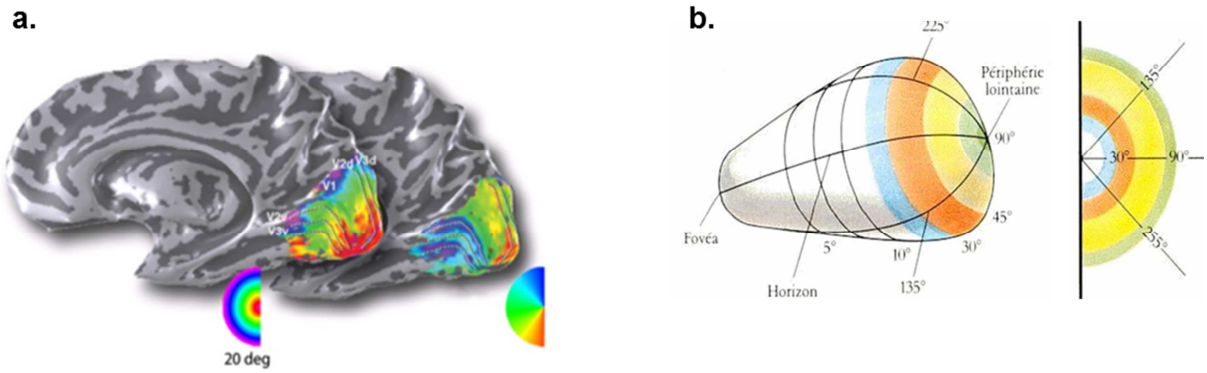


Figure 1 : (a) Organisation rétinotopique des aires visuelles V1, V2 et V3. Les couleurs indiquent la correspondance entre la position dans le champ visuel (en termes d'excentricité par rapport à la fovéa - au premier plan, ou d'angle polaire – au deuxième plan) et la projection au niveau cortex visuel. Par exemple, on peut remarquer que la partie centrale du champ visuel (en rouge) est projetée sur la partie postérieure du cortex visuel, tandis que la périphérie du champ visuel (en violet) est projetée sur la partie antérieure du cortex visuel. Figure adaptée de Wandell, Dumoulin, & Brewer (2007). **(b) Illustration de la magnification corticale dans V1.** La surface correspondant au champ visuel central (à droite) est surreprésentée au niveau cortical (à gauche). Figure adaptée de Hubel (1994).

Les afférences M et P en provenance du CGL se projettent sur des couches distinctes de V1. Ainsi, la séparation des informations en deux voies M et P reste préservée dans cette région. Le cortex visuel primaire se caractérise également par une organisation en colonnes verticales au sein desquelles les cellules répondent aux mêmes propriétés. On distingue les colonnes de dominance oculaire et les colonnes d'orientation. Les colonnes de dominance oculaire sont de larges bandes de tissu cérébral sur lesquelles se projettent de façon alternée les informations en provenance de l'œil gauche et de l'œil droit. Les colonnes d'orientation, orthogonales aux colonnes de dominance oculaire, sont plus fines, et chaque colonne contient des neurones sélectifs à une même orientation, sur 180° (Hubel & Wiesel, 1962, 1968, 1977). Les travaux de Hubel et Wiesel (Hubel & Wiesel, 1962, 1968, 1977) ont permis de mettre en évidence au sein de V1 deux types de cellules sélectives aux orientations : les cellules simples et les cellules complexes. De façon importante pour le raisonnement au cœur de ce travail de thèse, il a été montré que ces cellules sont également sensibles aux fréquences spatiales (De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Everson et al., 1998; Skottun et al., 1991), et qu'elles agiraient comme des filtres bidimensionnels, permettant d'extraire l'information des différents points du champ visuel à différentes orientations et différentes fréquences spatiales. L'information codée en V1 est donc une information locale, concernant l'orientation des contours des objets du champ visuel à différentes fréquences spatiales.

L'information visuelle en sortie de V1 est projetée vers les aires visuelles extra-striées, dont les principales sont les aires V2, V3, V4 et V5 (ou MT), impliquées dans les premières étapes du traitement visuel. Comme V1, ces régions sont rétinotopiques et sont spécialisées dans l'extraction de caractéristiques élémentaires spécifiques (Wandell et al., 2007). De manière générale, ces différentes régions se différencient donc par les propriétés de sélectivité de leurs cellules (e.g., contours, mouvement, couleur), la taille de leurs champs récepteurs, leurs liens de connectivité et leur représentation rétinotopique. Plus on s'élève dans la hiérarchie corticale, plus la taille des champs récepteurs des cellules augmente. Par conséquent, l'organisation rétinotopique devient de plus en plus grossière et les caractéristiques visuelles extraites sont de plus en plus complexes.

1.4. Les voies dorsale et ventrale

En sortie de V1, l'information visuelle est transmise à l'aire V2, à partir de laquelle on observe une dissociation anatomique en deux voies, la voie dorsale et la voie ventrale, véhiculant différent type d'information aux propriétés fonctionnelles différentes (de Haan & Cowey, 2011; Felleman & Essen, 1991; Goodale & Milner, 1992; Haxby et al., 1991; Ungerleider & Haxby, 1994). La voie dorsale véhicule l'information vers l'aire V5 (MT) principalement impliquée dans le traitement du mouvement (Maunsell & Van Essen, 1983a), avant d'être projetée vers le cortex pariétal. Du fait des propriétés des fibres M qui sont pour l'essentiel à l'origine de cette voie, les aires de la voie dorsale sont plutôt sensibles à une information achromatique, en mouvement et en BFS. Une autre conséquence de la présence massive de fibres M le long de la voie dorsale est la faible latence de réponse des neurones des aires de cette région. En effet, les fibres M ont la propriété de véhiculer l'information plus rapidement que les fibres P (Bullier, 2001; Merigan & Maunsell, 1993; Van Essen & Deyoe, 1995). L'information visuelle est donc conduite très rapidement le long de la voie dorsale.

La voie ventrale, faisant converger des afférences de type M et P, se projette vers l'aire V4, avant de parvenir au cortex inférotemporal. Les neurones de l'aire V4 sont sélectifs à la couleur et les neurones du cortex inférotemporal codent plutôt des formes complexes ou répondent à la combinaison de formes et de couleurs (Tanaka, Saito, Fukada, & Moriya, 1991; Tanaka, 1993, 1996). Les aires de la voie ventrale sont donc principalement impliquées dans le traitement de la couleur et de la forme des objets. Par ailleurs, les fibres de la voie ventrale étant majoritairement de type P, mais également, en proportion moindre, de type M et

K, les aires le long de cette voie sont principalement sensibles à une information en HFS, mais aussi en BFS.

La voie dorsale serait donc principalement impliquée dans le traitement du mouvement et de l'espace (voie du « Où »), tandis que la voie ventrale serait spécialisée dans le traitement de la forme et de la couleur, et serait impliquée dans les dernières étapes de la reconnaissance d'objets (voie du « Quoi »). Des données issues de la neuropsychologie corroborent cette dissociation en montrant, par exemple, des déficits spécifiques à la reconnaissance d'objet ou au traitement du mouvement, suite à une lésion respectivement localisée dans la voie ventrale ou dorsale (Damasio, Tranel, & Damasio, 1990; Zeki, 1990, 1991). Il est cependant à noter que la ségrégation en deux voies distinctes n'est pas aussi stricte qu'il n'y parait. En effet, des résultats issus de la neurophysiologie chez le primate non-humain suggèrent qu'il existe plus de 300 connexions entre les aires visuelles, dont des connexions entre des aires de la voie ventrale et de la voie dorsale (Breitmeyer, 2014; Felleman & Essen, 1991; Maunsell & Van Essen, 1983b; Merigan & Maunsell, 1993; Van Essen & Maunsell, 1983).

A la suite des travaux de Hubel et Wiesel (Hubel & Wiesel, 1962, 1968), pionniers dans l'étude des propriétés fonctionnelles du cortex visuel, de nombreuses études en neurophysiologie se sont attachées à recenser et à caractériser les aires cérébrales impliquées dans le traitement de l'information visuelle. Ces travaux ont permis d'identifier et de hiérarchiser chez le macaque une trentaine d'aires visuelles aux propriétés différentes avec plus de 300 liens de connectivité entre elles (Felleman & Essen, 1991; Van Essen & Deyoe, 1995; Van Essen & Maunsell, 1983). Sur la base de ce réseau d'aires visuelles, des modèles neurophysiologiques ont donc été proposés pour déterminer les différentes étapes de l'intégration de l'information visuelle avant d'aboutir à la reconnaissance d'objet.

2. Modèles neurophysiologiques de l'intégration de l'information visuelle

2.1. Modèles hiérarchiques

Les premiers modèles proposés sur la base de travaux en neurophysiologie réalisés chez le primate non-humain (Felleman & Essen, 1991; Marr, 1982; Van Essen & Deyoe, 1995; Van Essen & Maunsell, 1983) se caractérisent par une organisation hiérarchique des aires visuelles de telle sorte que les régions de plus haut niveau dans la hiérarchie intègrent une information visuelle issue des aires de plus bas niveau. Ainsi, le traitement de l'information

visuelle se ferait de manière ascendante, des premiers vers les derniers niveaux de la hiérarchie. Selon ces modèles, les neurones des aires visuelles de bas niveau telles que V1 et V2 effectuent des traitements locaux sur une information visuelle en deux dimensions (extraction de contours à différentes orientations), qui sont ensuite intégrés par les neurones des aires supérieures codant une information de plus en plus complexe (jonctions, formes), jusqu'à aboutir à une représentation globale et en trois dimensions de la scène visuelle au plus haut niveau de la hiérarchie.

La transmission de l'information visuelle de façon essentiellement ascendante constitue cependant l'une des principales limites de ces modèles hiérarchiques. En effet, le grand nombre d'aires et de niveaux impliqués dans la hiérarchie conduit à prédire que le temps nécessaire à l'intégration des différentes caractéristiques visuelles élémentaires pour aboutir à la reconnaissance devrait être long. Pourtant, de nombreuses études ont souligné la rapidité de la reconnaissance visuelle (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Crouzet, Kirchner, & Thorpe, 2010; Greene & Oliva, 2009; Johnson & Olshausen, 2003; Kirchner & Thorpe, 2006; Rolls & Tovee, 1994; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001a, 2001b). Par exemple, Thorpe et al. (1996) ont mesuré la rapidité de la reconnaissance visuelle en enregistrant les potentiels évoqués induit par la perception visuelle de scènes complexes. Les scènes étaient présentées seulement 30 ms et les participants avaient pour tâche d'appuyer sur un bouton si la scène contenait un animal et de ne rien faire sinon (réponse « go/no-go »). Les résultats ont montré que l'activité électrique cérébrale associée à la perception d'une scène contenant un animal se différenciait de l'activité associée à une scène ne contenant pas d'animal 150 ms après la présentation de la scène. Ces résultats suggèrent donc que seulement 150 ms suffisent pour analyser une scène et reconnaître les objets qui la composent, ce qui semble incompatible avec la longue chaine de traitements qu'impliquent les modèles hiérarchiques. Des travaux plus récents, qui enregistrent les mouvements oculaires rapides, montrent que lorsqu'un visage et un autre objet sont présentés de part et d'autre d'un point de fixation central, une saccade oculaire peut être déclenchée en direction du visage en seulement 100 ms (Crouzet et al., 2010). La rapidité de cette réponse motrice indique qu'en fait moins de 100 ms suffisent pour reconnaître le visage.

Une autre limite majeure des modèles hiérarchiques réside dans le fait que ces modèles fonctionnent et permettent la reconnaissance seulement dans la situation très peu fréquente où le stimulus visuel serait univoque et non ambigu (Bullier, 2001). En effet, Bullier (2001) remarque que si que le stimulus visuel est par exemple masqué, l'extraction de caractéristiques locales par les aires de bas niveau puis leur intégration dans les aires de haut

niveau devrait conduire à une perception globale erronée. Pourtant, nous reconnaissons sans effort les objets de notre environnement indépendamment de la complexité de celui-ci, et parvenons à les envisager de façon globale et complète, malgré le fait qu'ils soient souvent masqués par d'autres. Dès lors, l'analyse locale des aires de bas niveau doit être contrainte par des représentations plus globales pour permettre la reconnaissance d'un stimulus ambigu. Ces considérations ont donc conduit plusieurs auteurs à proposer des modèles ne fonctionnant pas de manière uniquement ascendante, mais selon lesquels des représentations globales issues d'aires de haut niveau pourraient influencer le traitement d'informations plus locales dès les premiers niveaux de la hiérarchie.

2.2. Modèle par itération

Plusieurs auteurs ont proposé des modèles de la perception visuelle accordant un rôle prépondérant aux connexions descendantes, permettant de contraindre les traitements effectués par les aires de bas niveau sur la base de représentations plus globales au sein d'aires de plus haut niveau (e.g., Bullier, 2001; Lamme et al., 1998; Lee et al., 1998; Lee, 2003). En particulier, Bullier (2001) propose un modèle de l'intégration de l'information visuelle sur la base d'un mécanisme de rétro-injection de l'information traitée par des aires de haut niveau vers les aires de plus bas niveau. A l'appui de ce modèle, ses travaux en neurophysiologie chez le macaque (Bullier, Hupé, James, & Girard, 1996, 2001; Bullier, 1998; Nowak, Munk, Girard, & Bullier, 1995) suggèrent que, contrairement à ce qui pourrait être prédit par les modèles hiérarchiques, les latences de réponse dans des aires visuelles dites de « haut niveau » dans la hiérarchie sont comparables, voire plus courtes que celles des cellules d'aires de plus bas niveau. Par exemple, les latences de réponse dans des aires frontales telles que le « frontal eye field » ou pariétales telles que l'aire MST, sont de l'ordre de 40 à 60 ms, comme les latences de réponse de l'aire V1 (cf. Figure 2).

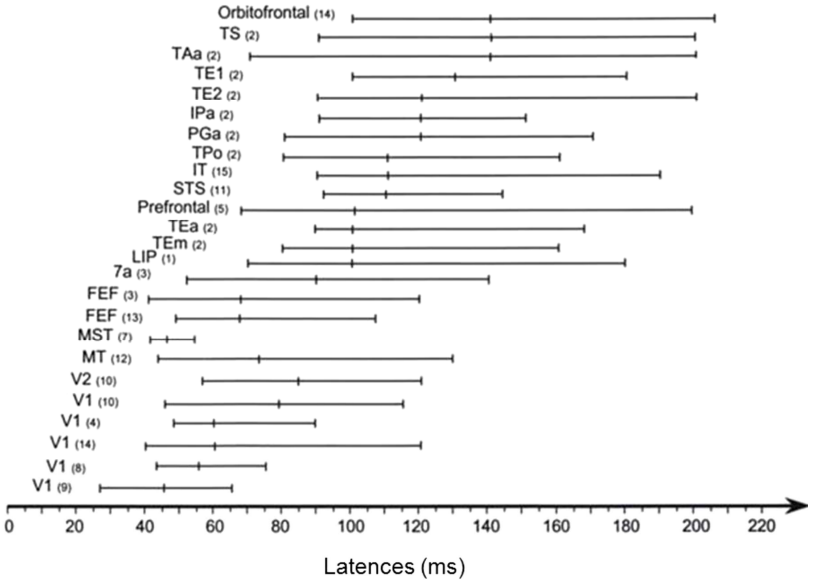


Figure 2 : Latences de réponse visuelle de neurones de différentes aires corticales. Pour chaque aire, les barres verticales centrales représentent la latence de réponse médiane (en millisecondes) et les barres latérales représentent les centiles de 10 à 90%. Les chiffres entre parenthèse renvoient aux références citées dans Bullier et al. (2001). On peut noter que les latences de réponse les plus courtes ne correspondent pas forcément aux niveaux les plus bas de la hiérarchie. En particulier, les aires MT, MST, ou FEF ont des latences de réponse comparables à celles de V1, bien qu'elles se situent plus haut dans la hiérarchie. Figure adaptée de Bullier (2001).

Ces aires de haut niveau aux latences de réponse très courtes pourraient donc être activées assez rapidement pour influencer la réponse d'aires de plus bas niveau telles que V1 et V2. Par ailleurs, le nombre de connexions descendantes est plus important que le nombre de connexions ascendantes, ce qui suggère une forte influence des aires de haut niveau sur les aires de bas niveau. On peut également noter que les aires aux latences de réponse les plus courtes se situent principalement le long de la voie dorsale, tandis que les aires de la voie ventrale ont des latences de réponse plus longues. Cela s'explique par le fait que, comme nous l'avons vu précédemment, la voie dorsale est dominée par des fibres M dont la vitesse de conduction est plus importante que les fibres P, majoritaires le long de la voie ventrale. Dans son modèle intégré du traitement de l'information visuelle, Bullier (2001) propose donc que le traitement de l'information visuelle ne se ferait pas selon une propagation uniquement ascendante, mais selon trois vagues d'activation (cf. Figure 3). La première vague d'activation correspond à la projection vers le cortex visuel primaire de l'information en sortie du CGL véhiculée rapidement par la voie M. Cette information serait propagée vers les aires de la voie dorsale mais également ventrale, toujours par le biais des fibres de type M. Lors de la deuxième vague d'activation, les traitements effectués dans ces régions de haut niveau seraient alors rétro-injectés, via des connexions descendantes rapides, vers les aires de plus

bas niveau telles que V1 et V2. Au même moment, l'information véhiculée par la voie P en provenance du CGL arriverait également aux aires visuelles de bas niveau, constituant la troisième vague d'activation. Ce mécanisme de rétro-injection permettrait donc de moduler et guider le traitement des afférences parvocellulaires dès leurs premières étapes de traitement au niveau du cortex visuel.

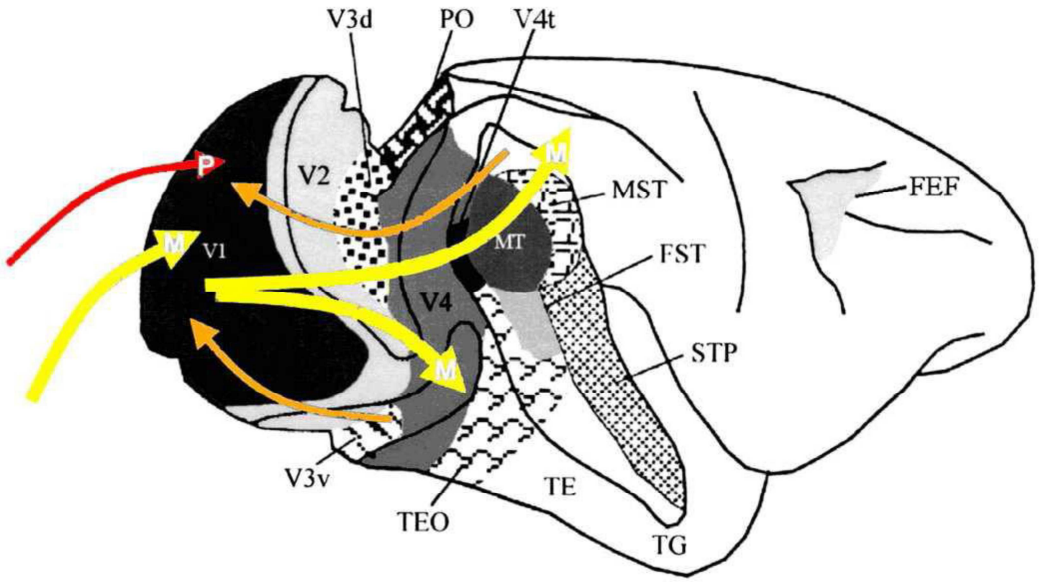


Figure 3 : Modèle intégré du traitement de l'information visuelle de Bullier (2001). Les flèches jaunes représentent la première vague d'activation véhiculée par la voie M, propagée des aires visuelle de bas niveau vers des aires visuelles de haut niveau pariétales et temporales. Les flèches orange représentent la vague d'activation descendante (rétro-injection), propagée des aires pariétales et temporales vers les aires visuelle de plus bas niveau. Cette rétro-injection a lieu au moment de l'arrivée plus tardive de l'information portée par la voie P (flèche rouge) sur le cortex visuel primaire, qui constitue la troisième vague d'activation. Figure adaptée de Bullier (2001).

Ainsi, les traitements effectués au sein des aires de haut niveau dont les champs récepteurs sont plus importants, donnant lieu à des représentations plus globales, pourraient contraindre les traitements plus locaux effectués dans les aires V1 et V2 (Bullier et al., 2001; Hupé et al., 1998, 2001). Dans ce modèle, le cortex visuel primaire joue un rôle intégratif plutôt que purement sensoriel, en combinant des représentations globales issues d'influences descendantes au traitement d'une information locale issue d'influences ascendantes.

Cependant, ce modèle fournit peu de précisions quant à la nature de l'information visuelle véhiculée au cours des différentes vagues d'activation. Une réponse peut cependant être apportée sur la base des propriétés des voies M et P que nous avons évoquées précédemment. En effet, ces deux voies répondent de façon spécifique à différentes propriétés physiques de l'information visuelle, telles que le mouvement (voie M) ou la couleur (voie P),

mais aussi – et de façon plus importante pour la suite de ces travaux – aux fréquences spatiales. La voie M est sélective à une information en BFS tandis que la voie P répond plutôt à une information en HFS. Ainsi, dans le modèle de Bullier (2001), la première vague d'activation pourrait correspondre à la propagation d'une information en BFS, qui viendrait guider, en correspondance avec la deuxième vague d'activation, le traitement de l'information en HFS véhiculée plus tardivement par la voie P, en correspondance avec la troisième vague d'activation. Dans ce contexte, nous verrons plus en détails dans la suite de ce manuscrit comment le traitement rapide d'une information en BFS pourrait guider le traitement plus tardif de l'information en HFS lors de la catégorisation de scènes. Il est cependant important dans un premier temps de voir comment se définissent les fréquences spatiales, et la façon dont elles sont traitées par le système visuel.

3. Traitement des fréquences spatiales

Comme nous l'avons mentionné précédemment, les travaux en neurophysiologie réalisés sur le cortex visuel primaire du macaque et du chat ont montré que celui-ci est composé de cellules simples et complexes dont les champs récepteurs sont sélectifs à différentes orientations et différentes fréquences spatiales (De Valois, Albrecht, et al., 1982; De Valois, Yund, et al., 1982; Hubel & Wiesel, 1962, 1968, 1977). Les fréquences spatiales constitueraient donc une caractéristique élémentaire du signal visuel et joueraient un rôle crucial dès les premières étapes de traitement. Nous verrons dans cette partie comment se définissent et se caractérisent les fréquences spatiales, et comment cette propriété du signal visuel peut être manipulée lorsque l'on s'intéresse à des stimuli visuels complexes tels que les scènes. Nous présenterons par la suite les principaux travaux portant sur le traitement des fréquences spatiales qui suggèrent une précédence temporelle du traitement des BFS sur le traitement des HFS.

3.1. Représentation et manipulation d'images dans le domaine fréquentiel

Une image peut être considérée, dans le domaine spatial, comme une matrice de pixels de différentes valeurs d'intensité (i.e. luminance). Une image peut également être représentée dans le domaine fréquentiel, grâce à la transformée de Fourier. Cette opération mathématique permet de décomposer le signal en une somme de sinusoïdes qui peuvent être caractérisées par leur fréquence, leur amplitude (contraste), leur phase et leur orientation. Dans le domaine

fréquentiel (ou domaine de Fourier), une image peut être résumée par son spectre d'amplitude et son spectre de phase. Le spectre d'amplitude donne la distribution du contraste de luminance (amplitude) en fonction des fréquences spatiales et des orientations (cf. Figure 4). Le spectre de phase caractérise la position relative des fréquences spatiales dans l'espace.

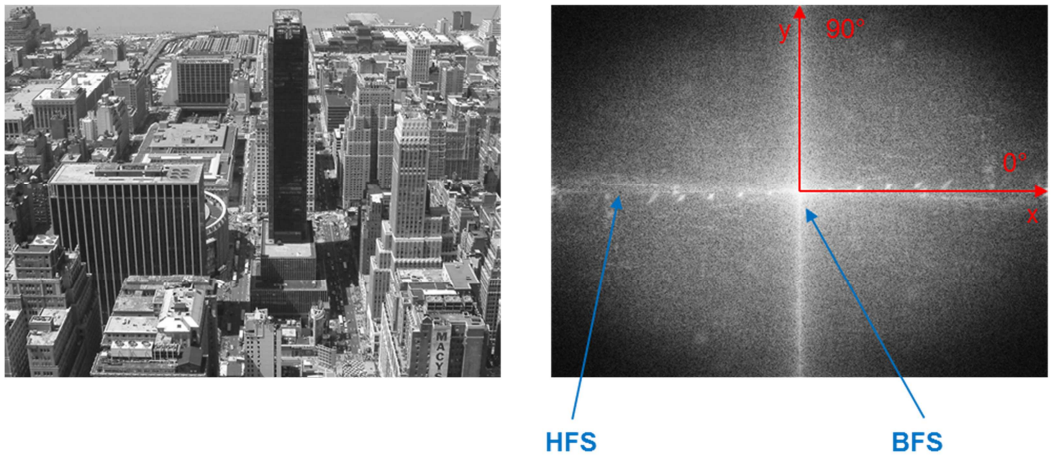


Figure 4 : Exemple de la représentation d'une scène de ville dans le domaine spatial (à gauche) et de son spectre d'amplitude dans le domaine de Fourier (à droite). Sur le spectre d'amplitude, l'axe *x* correspond aux orientations verticales et l'axe *y* correspond aux orientations horizontales. Les basses fréquences sont codées au centre et les hautes fréquences sont codées en périphérie. Le spectre d'amplitude est symétrique selon les axes *x* et *y*. Sur cet exemple, on peut constater qu'une scène de ville se caractérise par une répartition de l'information en basses et hautes fréquences spatiales en majorité sur des orientations verticales et horizontales.

Afin d'étudier le traitement des fréquences spatiales lors de la perception d'images complexes telles que les scènes, il est possible de manipuler et de filtrer les images dans le domaine fréquentiel (Ginsburg, 1986; cf. Figure 5) afin d'obtenir dans le domaine spatial, par la transformée de Fourier inverse, des images dont le contenu en fréquences spatiales est contrôlé. Par exemple, un filtre passe-bas appliqué au spectre d'amplitude permet de ne laisser passer qu'un signal en BFS et d'atténuer le signal en HFS supérieur à une fréquence spatiale donnée (i.e. fréquence de coupure). A l'inverse, un filtre passe-haut permet de ne conserver que le signal en HSF et d'atténuer le signal en BFS, inférieur la fréquence de coupure. Enfin, un filtre passe-bande permet de ne conserver que le signal compris entre deux fréquences de coupure et d'atténuer le signal dont la fréquence spatiale est supérieure ou inférieure. Notons que lors de ces manipulations, le spectre de phase n'est pas altéré, afin de préserver la position relative des fréquences spatiales dans l'espace et donc, l'organisation spatiale de l'image.

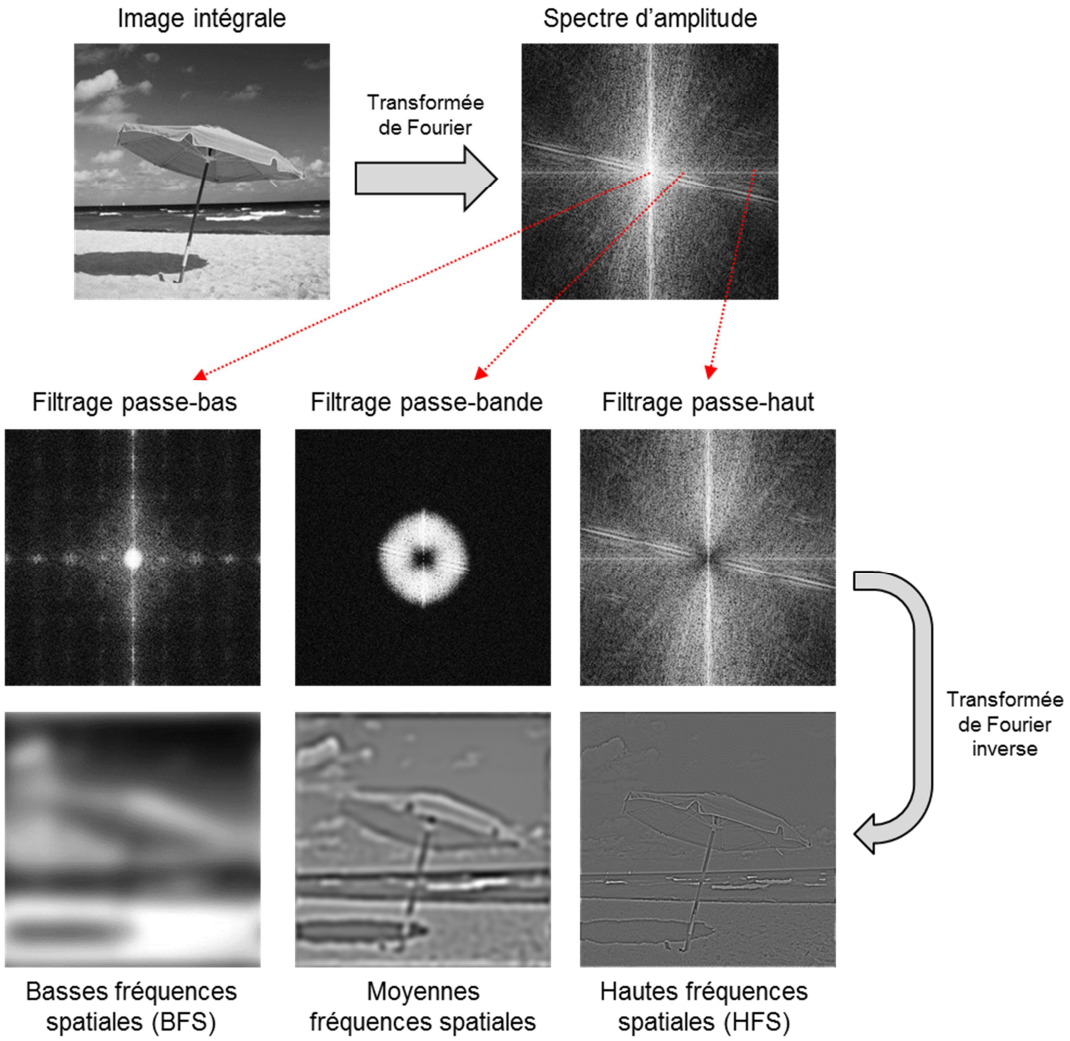


Figure 5 : Illustration de la procédure de filtrage d'une scène. Le spectre d'amplitude d'une scène est obtenu par la transformée de Fourier. On peut ainsi effectuer sur le spectre d'amplitude un filtrage passe-bas, passe-bande ou passe-haut pour ne conserver dans le domaine spatial (via la transformée de Fourier inverse) que l'information en basses, moyennes ou hautes fréquences spatiales, respectivement. Comme on peut le constater les BFS d'une scène contiennent une information grossière, floue, permettant d'avoir accès à sa structure globale. Les HFS contiennent quant à elles une information fine, permettant de distinguer les contours des objets et les détails de la scène.

Sur la base des travaux neurophysiologiques réalisés sur le primate non humain (De Valois, Albrecht, et al., 1982; De Valois & De Valois, 1988), plusieurs auteurs proposent que le système visuel fonctionnerait comme un analyseur de Fourier au niveau local, et qu'il extrairait en parallèle l'information du champ visuel à différentes orientations et différentes fréquences spatiales (Maffei & Fiorentini, 1973; Marendaz, Rousset, & Charnallet, 2003; Robert Shapley & Lennie, 1985; Westheimer, 2001). Nous allons à présent voir comment se définissent les fréquences spatiales, ainsi que le rôle du contraste pour leur perception.

3.2. Les fréquences spatiales et la sensibilité au contraste

3.2.1. Les fréquences spatiales dans les réseaux sinusoïdaux

La fréquence spatiale caractérise la structure périodique des variations de luminance dans le champ visuel. Elle correspond au nombre de cycles de variation de luminance par unité de distance. L'exemple le plus simple pour comprendre la notion de fréquence spatiale est le réseau sinusoïdal, stimulus visuel souvent utilisé dans les études psychophysiques. Un réseau sinusoïdal se compose d'une alternance de bandes sombres et claires dont les valeurs de luminance suivent une fonction sinusoïdale. Dans un réseau sinusoïdal, un cycle correspond à l'intervalle entre deux pics de luminance (i.e. une bande claire et une bande sombre forment un cycle). La fréquence spatiale correspond au nombre de cycles par degré d'angle visuel de l'image rétinienne (cpd). L'angle visuel exprime la taille de l'image rétinienne en fonction de la distance de l'observateur au stimulus visuel. Par conséquent, la fréquence spatiale correspond au nombre de cycles contenus dans l'image rétinienne pour un degré d'angle visuel (cf. Figure 6). Ainsi, plus l'observateur s'éloigne d'un stimulus, plus le nombre de cycles, et donc la fréquence spatiale contenue par degré d'angle visuel, augmente. La fréquence spatiale (*FS*) est donnée par la formule $FS = \frac{1}{2\arctan(\frac{L}{2d})}$, avec L correspondant à la largeur d'un cycle en cm et d la distance en cm entre l'œil de l'observateur et le stimulus. La formule $2\arctan(\frac{L}{2d})$ exprime la taille angulaire d'un cycle.

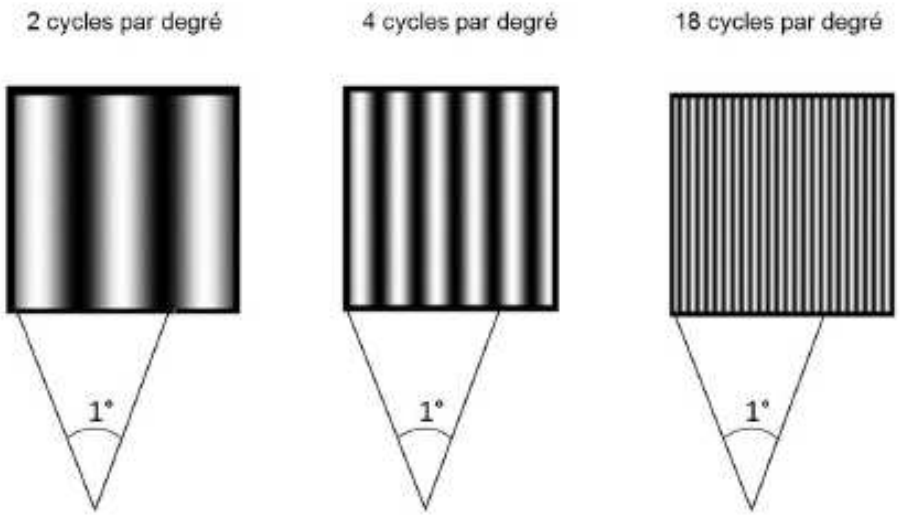


Figure 6 : Exemples de réseaux sinusoïdaux à différentes fréquences spatiales.

Les réseaux sinusoïdaux ont été utilisé dans plusieurs travaux en psychophysique chez l'humain (Blakemore & Campbell, 1969; Breitmeyer, 1975) afin d'étudier le traitement des différentes fréquences spatiales au niveau comportemental. En particulier, Breitmeyer (1975) a mesuré les temps de réaction pour détecter des réseaux sinusoïdaux à différentes fréquences spatiales et a observé que les réseaux en BFS (0,5 cpd) étaient détectés plus rapidement que les réseaux en HFS (11 cpd).

Les travaux psychophysiques utilisant des réseaux sinusoïdaux fournissent des données de référence pour la compréhension du traitement des fréquences spatiales par le système visuel. Il est cependant important de remarquer que les nombreux travaux comportementaux réalisés sur la perception des formes visuelles globale et locale ont également contribués à une meilleure compréhension des processus cognitifs sous-jacents au traitement des fréquences spatiales.

3.2.2. Les fréquences spatiales dans la perception globale et locale

Ces travaux, initiés par un courant théorique fondé au début du XXème siècle, le courant de la psychologie de la forme ou psychologie de la Gestalt, reposent sur le postulat que les éléments du champ visuel seraient traités de façon prédominante comme un tout global, et non comme une simple association d'éléments locaux. Le « tout » serait donc supérieur à la « somme des parties » pour l'organisation perceptive (Kohler, 1964). Dans ce contexte, de nombreux travaux se sont attachés à étudier le décours temporel d'une perception globale et locale des éléments visuels. Ces études utilisaient des figures hiérarchiques permettant de manipuler l'information présente aux niveaux global et local d'un stimulus visuel. Ces figures hiérarchiques sont généralement de grandes lettres (élément global) constituées de petites lettres (élément local, cf. Figure 7). L'information globale et locale peuvent être soit congruentes (e.g., un grand « H » constitué de petits « H »), soit non congruentes (e.g., un grand « H » constitué de petits « E »). La tâche des participants est d'identifier le plus rapidement possible soit la lettre globale, soit la lettre locale.

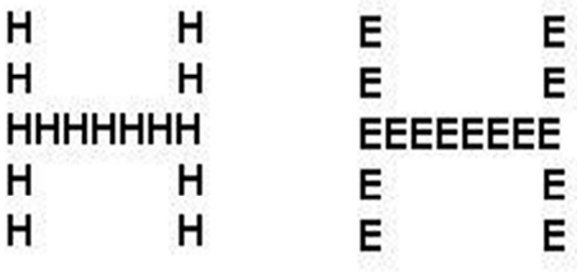


Figure 7 : Exemple de figures hiérarchiques. A gauche, la forme globale (grand « H ») est congruente avec la forme locale (petit « H »). A droite, la forme globale (grand « H ») est non congruente avec la forme locale (petit « E »).

Deux résultats principaux émergent avec ce type de paradigme (Navon, 1977). Tout d'abord, les participants sont plus rapides pour identifier la grande lettre (information globale) que la petite lettre (information locale). Cette prédominance de l'information globale a été appelée effet de précédence globale. De plus, on observe que l'identification de l'information locale conduit à des temps de réponse plus long lorsque celle-ci est non congruente que lorsqu'elle est congruente avec l'information globale, tandis que l'identification de l'information globale n'est pas influencée par la congruence entre l'information globale et locale. Ce dernier résultat suggère que lorsque les informations globales et locales ne sont pas congruentes, seule l'information globale interfère avec le traitement de l'information locale et gène son identification. C'est l'effet d'interférence globale.

De nombreuses études ont par la suite cherché à établir un lien entre le traitement de l'information globale/locale et le traitement des BFS/HFS (Badcock, Whitworth, Badcock, & Lovegrove, 1990; Hughes, Fendrich, & Reuter-Lorenz, 1990; Hughes, Nozawa, & Kitterle, 1996; Hughes, 1986; Lagasse, 1993; Lamb & Yund, 1993; Lamb, 1996; Shulman, Sullivan, Gish, & Sakoda, 1986; Shulman & Wilson, 1987). Par exemple, Shulman, Sullivan, Gish, & Sakoda (1986) ont montré que l'exposition prolongée à des réseaux sinusoïdaux en BFS affectait par la suite la détection de l'orientation d'une lettre au niveau global d'une figure hiérarchique, tandis que l'exposition prolongée à des réseaux sinusoïdaux en HFS affectait ensuite la détection de l'orientation d'une lettre au niveau local d'une figure hiérarchique. Ce résultat suggère donc que les BFS et les HFS permettraient de coder respectivement les niveaux global et local. Dans une autre étude, Shulman & Wilson (1987) ont testé si l'orientation de l'attention au niveau global ou local d'un stimulus affectait le traitement de réseaux sinusoïdaux de différentes fréquences spatiales. Ces auteurs ont observé que l'orientation de l'attention au niveau global facilitait le traitement des BFS tandis que

l'orientation de l'attention au niveau local facilitait le traitement des HFS. Ces résultats suggèrent donc encore une fois un lien entre le traitement de l'information globale et locale et le traitement des BFS et des HFS, respectivement. Par la suite, Badcock, Whitworth, Badcock et Lovegrove (1990) ont directement manipulé le contenu fréquentiel de stimuli hiérarchiques et ont mis en évidence que lorsqu'une figure hiérarchique était filtrée en HFS (et ne contenait pas d'information en BFS), le traitement de l'information globale était plus long que lorsque la figure n'était pas filtrée. De plus, l'information globale n'interférait pas avec le traitement de l'information locale. Ces résultats, qui montrent que la suppression de l'information en BFS réduit les effets de précédence et d'interférence globale, suggèrent à nouveau que l'information globale de formes hiérarchiques serait contenue dans les BFS. Il est à noter que cette étude ne comprenait pas de condition inverse dans laquelle la figure hiérarchique était filtrée en BFS (et privée de HFS) afin de tester le lien entre l'information locale et les HFS. En effet, cette condition n'était pas envisageable car un filtrage passe-bas floute les éléments locaux d'une figure hiérarchique, empêchant leur identification (ce qui n'est pas le cas du filtrage passe-haut qui laisse la possibilité d'identifier une figure hiérarchique aux niveaux local et global). On peut d'ailleurs considérer cette limite d'ordre méthodologique comme une mise en garde pour quiconque voudrait étudier les mécanismes neurocognitifs du traitement des fréquences en utilisant des figures hiérarchiques. En effet, comme nous l'avons souligné ci-dessus, l'information globale peut être identifiée sur la base des BFS, mais elle reste disponible même dans une image ne conservant que les HFS. Cela montre bien que le lien entre l'information globale et les BFS d'une part, et l'information locale et les HFS d'autre part, ne signifie pas que ces notions soient interchangeables (Hughes et al., 1996; Kimchi, 1992; Morrison & Schyns, 2001; Oliva & Schyns, 1997).

Cependant, pris dans leur ensemble, ces travaux indiquent clairement une précédence du traitement des BFS sur le traitement des HFS. Nous verrons dans la partie suivante que le traitement des fréquences spatiales est cependant intimement lié à une autre caractéristique du signal visuel, le contraste.

3.2.3. Le contraste

Outre sa fréquence spatiale, un réseau sinusoïdal se caractérise également par son amplitude. L'amplitude d'un réseau sinusoïdal correspond à la différence d'intensité lumineuse entre le point le plus clair et le point le plus sombre dans le réseau, et fournit une

estimation de son contraste. Classiquement, le contraste d'un réseau sinusoïdal est cependant estimé selon la formule *C* de Michelson:

$$C_{Michelson} = \frac{L_{\max -} L_{min}}{L_{\max +} L_{min}}$$

avec L_{max} l'intensité lumineuse maximum et L_{min} l'intensité lumineuse minimum.

Il existe d'autres estimations du contraste, telles que le contraste de Weber qui correspond à la différence entre L_{max} et L_{min} , divisée par l'intensité lumineuse du fond, ou le contraste « root-mean-square » (RMS) qui correspond à l'écart type des valeurs de luminance, divisé par la luminance moyenne. L'utilisation préférentielle de ces différentes estimations du contraste varie selon le type de stimulus dont on veut évaluer la visibilité (Bex & Makous, 2002). Le contraste *C* de Michelson est préférentiellement utilisé pour les réseaux, tandis que le contraste de Weber est plus approprié pour des stimuli tels que les lettres. De façon importante pour la suite de ce travail, le contraste RMS serait le meilleur indicateur de la visibilité de stimuli complexes tels que les scènes. Cette estimation du contraste a l'avantage de ne pas dépendre du contenu en fréquences spatiales ou de la distribution spatiale du contraste dans l'image. Nous verrons par la suite que l'égalisation du contraste RMS est également largement utilisée comme un contrôle expérimental pour étudier les bases cérébrales du traitement des fréquences spatiales dans des stimuli naturels tels que les scènes ou les visages (Goffaux et al., 2011; Vlamings, Goffaux, & Kemner, 2009).

La capacité de discriminer les variations de luminance (contraste) est indispensable pour la perception des fréquences spatiales. En effet, lorsque le contraste entre les bandes d'un réseau sinusoïdal est par exemple proche de zéro, le réseau sera perçu comme une surface uniforme et non comme une alternance de bandes sombre et claire. La plus petite différence d'intensités lumineuses perceptible entre deux bandes correspond au seuil de sensibilité au contraste (i.e. valeur de contraste minimum pour percevoir des réseaux de fréquences spatiales différentes). La sensibilité au contraste en fonction des fréquences spatiales donne ainsi la fonction de sensibilité au contraste (FSC, cf. Figure 8). Cette fonction varie de façon curvilinéaire avec la fréquence spatiale.

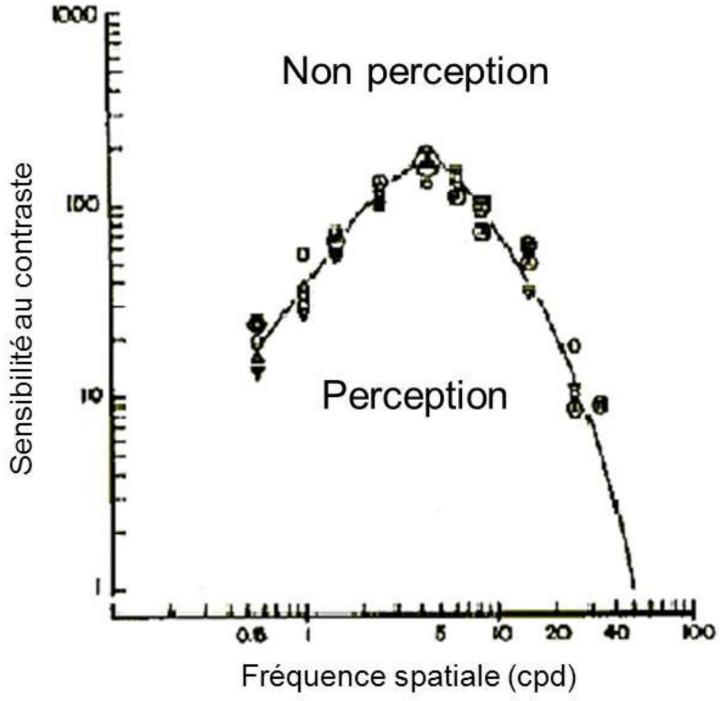


Figure 8 : Fonction de sensibilité au contraste chez l'adulte pour un réseau statique à différentes fréquences spatiales présenté en vision centrale. L'axe des abscisses représente la fréquence spatiale en cpd. L'axe des ordonnées correspond à la sensibilité au contraste (inverse du contraste : 1/c). L'échelle des axes est logarithmique. La surface sous la courbe correspond à la perception d'une alternance de bandes sombres et claires et la surface au-dessus de la courbe correspond à la perception d'une surface uniforme. Adapté de De Valois & De Valois (1988).

Chez l'adulte sain, la sensibilité au contraste est maximale pour des fréquences spatiales entre 2 et 6 cpd. Au-delà de 6 cpd, la sensibilité au contraste chute drastiquement. On estime la fréquence spatiale la plus élevée perceptible à un contraste maximum autour de 60 cpd (De Valois & De Valois, 1988). Le contraste influence non seulement la visibilité de réseaux à différentes fréquences spatiales, mais également la rapidité de détection de ceux-ci. En effet, plusieurs travaux montrent une influence du contraste sur le traitement des fréquences. Par exemple, une première série de travaux rapporte que le temps de réaction pour détecter un réseau quelle que soit sa fréquence spatiale augmente lorsque le contraste du réseau diminue (Harwerth & Levi, 1978; Lupp, Hauske, & Wolf, 1976; Parker & Dutch, 1987; Parker, 1980; Vassilev & Mitov, 1976). Par ailleurs, Vassilev et Mitov (1976) ont directement testé si la précédence temporelle des BFS sur les HFS pouvait être expliquée par un sensiblité au contraste plus faible pour les HFS que les BFS (i.. plus la fréquence spatiale est haute et moins le réseaux est visible). Dans cette expérience, les auteurs utilisaient des réseaux en BFS (de 2 à 4 cpd) et HFS (au-delà de 6 cpd) dont le contraste était soit faible (*C*

de Michelson de 0,2), soit réhaussé (*C* de Michelson de 0,6). Les résulats ont montré que la précédence temporelle de la détection des réseaux en BFS sur les réseaux en HFS observée en condition de contrastre faible était réduite en condition de contraste réhaussé. Plus précisément, le réhaussement du contraste améliorait significativement la détection des réseaux en HFS.

La visibilité des fréquences spatiales et leur vitesse de traitement étant directement liées au niveau de contraste, nous verrons dans la suite de ce manuscrit que le contraste doit nécessairement être pris en considération pour l'étude du traitement des fréquences spatiales.

Dans l'ensemble, les études comportementales réalisées à l'aide de réseaux sinusoïdaux ou de figures hiérarchiques soutiennent donc l'idée d'une précédence du traitement de l'information globale et des BFS sur le traitement de l'information locale et des HFS. Cela apparaît compatible avec les propriétés temporelles et fonctionnelles des voies M et P qui véhiculent respectivement l'information en BFS de façon très rapide et l'information en HFS de façon plus lente (Bullier, 2001; Derrington & Lennie, 1984; Enroth-Cugell & Robson, 1966, 1984; Hicks et al., 1983; Maunsell et al., 1999; Merigan & Maunsell, 1993; Van Essen & Deyoe, 1995). Les modèles actuels de la reconnaissance visuelle postulent donc que l'analyse visuelle se ferait selon une séquence de traitement par défaut « coarse-to-fine ». L'analyse rapide de l'information grossière (« coarse ») en BFS fournirait un aperçu global du stimulus visuel, qui serait ensuite affiné par l'analyse plus détaillée de l'information fine (« fine ») en HFS.

Nous développerons donc dans la partie suivante les différents arguments neurophysiologiques, comportementaux et computationnels en faveur de la prédominance d'une analyse « coarse-to-fine ». Ces travaux se sont particulièrement intéressés au traitement des fréquences spatiales lors de la perception et la catégorisation de stimuli plus complexes et écologiques que les réseaux sinusoïdaux, tels que les scènes naturelles.

4. Une analyse visuelle selon une séquence « coarse-to-fine »

4.1. Arguments neurophysiologiques

Comme nous l'avons vu jusqu'ici au niveau neurophysiologique, la précédence du traitement des BFS sur le traitement des HFS est majoritairement inférée des propriétés temporelles et fonctionnelles des voies M et P. En manipulant directement les fréquences

spatiales de réseaux sinusoïdaux, des travaux en neurophysiologie chez le chat et le singe (Allen & Freeman, 2006; Bredfeldt & Ringach, 2002; Einevoll, Jurkus, & Heggelund, 2011; Frazor, Albrecht, Geisler, & Crane, 2004; Mazer, Vinje, McDermott, Schiller, & Gallant, 2002; Nirody, 2014) ont également pu mettre en évidence une dynamique de traitement « coarse-to-fine » en mesurant la latence de réponse des neurones du LGN et du cortex visuel primaire induite par la présentation des réseaux. Ces études ont par exemple montré que la sélectivité des neurones de ces régions changeait au cours du temps (passant d'une sélectivité à une large gamme de fréquences spatiales après présentation des stimuli à une sélectivité plus restreinte au cours du temps), mais aussi que la fréquence spatiale privilégiée était de plus en plus haute au cours du temps. Chez l'humain, Parker et Salzen (1977) ont mesuré sur les sites occipitaux les potentiels électriques évoqués par la présentation de réseaux sinusoïdaux à différentes fréquences spatiales. Ces auteurs ont observé que plus la fréquence spatiale du réseau augmentait, plus la latence du pic d'activité générée par la présentation du stimulus augmentait. Ces résultats fournissent donc un substrat neurophysiologique à la prédominance d'une analyse « coarse-to-fine ».

4.2. Arguments comportementaux

Comme nous l'avons vu précédemment, les premiers arguments comportementaux en faveur d'une analyse « coarse-to-fine » proviennent d'étude utilisant des réseaux sinusoïdaux ou des figures hiérarchiques comme stimuli. Par la suite, de nombreux travaux en psychologie cognitive se sont attachés à étudier le rôle des fréquences spatiales et leur traitement pour la perception et la catégorisation de stimuli plus complexes tels que les scènes naturelles, les objets ou les visages (Hughes et al., 1996; Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012; Parker, Lishman, & Hughes, 1992, 1996, 1997; Schyns & Oliva, 1994, 1999). Ces études ont montré qu'une analyse visuelle « coarse-to-fine » serait la plus efficace pour percevoir correctement et catégoriser rapidement un stimulus visuel. Dans cette optique, la prédominance d'une analyse « coarse-to-fine » lors de la perception visuelle résulterait de contraintes d'ordre perceptives.

4.2.1. Une analyse « coarse-to-fine » résultant de contraintes perceptives ?

Selon Parker et al. (1992, 1997), la différence de latence entre la detection d'un réseau en BFS et celle d'un réseau HFS pourrait être secondaire à la façon dont le système visuel va

intègrer l'information en fréquences spatiales pour arriver à un percept correct du stimulus visuel, et ce, notamment dans le cas de stimuli plus complexes et écologiques tels que les scènes. Si la reconnaissance visuelle dépend seulement de la vitesse de traitement des BFS et HFS, le temps nécessaire pour arriver à une perception complète du stimulus serait déterminé par la composante en fréquences spatiales la plus longue à traiter. Dans ce cas, il n'y aurait pas de biais dans la façon dont le système visuel analyserait l'information en fréquences spatiales. Il accumulerait l'information comme elle arriverait et concluerait au même percept quel que soit l'ordre d'arrivée des BFS et HFS (hypothèse d'intégration isotropique). Par contre, si l'ordre d'accumulation ou d'intégration des fréquences spatiales est crucial, une transgression de l'ordre naturel par lequel l'information est naturellement accumulée (des basses aux hautes fréquences spaitales) gênerait alors l'intégration de l'information. Dans ce cas, le système visuel ne concluerait pas au même percept suivant l'ordre d'arrivée des BFS et HFS (hypothèse d'intégration anisotropique). Pour tester ces deux hypothèses alternatives, les auteurs ont présenté dans une première étude des séquences composées de trois versions filtrées d'une même scène, passant des BFS aux HFS ou des HFS aux BFS (Parker et al., 1992). Les séquences étaient présentées 120 ms (40 ms par image filtrée dans la séquence) et les participants devaient évaluer la qualité perçue de la scène sur une échelle de 1 (mauvaise qualité) à 4 (bonne qualité). Sous l'hypothèse d'une intégration isotropique des fréquences spatiales, les auteurs s'attendaient à ce que les deux types de séquences soient jugés de qualité similaire. A l'inverse, sous l'hypothèse d'une intégration anisotropique, ils s'attendaient à ce que les séquences pour lesquelles les BFS seraient présentées après les HFS, transgressant alors l'ordre naturel du traitement des fréquences spatiales, soient jugées comme de moins bonne qualité. Les auteurs ont constaté que les séquences passant des BFS aux HFS étaient jugées comme de meilleure qualité que les séquences suivant l'ordre inverse. Dans une deuxième étude (Parker et al., 1997), ces auteurs ont à nouveau présenté des séquences composées de trois versions filtrées d'une image (scènes et visages), assemblées des BFS aux HFS ou inversement, ainsi que des séquences composées de trois versions non-filtrées de l'image. Les participants devaient cette fois-ci indiquer s'ils percevaient la séquence comme contenant des images intactes (i.e. non filtrées) ou dégradées (i.e. filtrées). Les résultats ont montré que pour les scènes, comme pour les visages, les participants percevaient plus souvent la séquence comme contenant des images intactes lorsque celle-ci était présentée des BFS aux HFS que dans l'ordre inverse. Les résultats de ces deux études ont été interprétés comme reflétant une meilleure intégration de l'information visuelle (conduisant à la perception d'une image de meilleure qualité et non dégradée) lorsque celle-ci était accumulée des BFS aux

HFS, en faveur de l'hypothèse d'une intégration anisotropique de l'information en fréquences spatiales. Selon les auteurs, l'accumulation de l'information visuelle selon l'ordre inverse (des HFS aux BFS) ne serait pas efficace et gênerait sa bonne intégration. Ces études ont été parmi les premières à mettre en évidence, au niveau comportemental, le rôle de l'ordre relatif du traitement des fréquences spatiales lors de la perception visuelle. Cependant, si cette étude montre le rôle de l'ordre de présentation des fréquences spatiales pour aboutir à une intégration optimale de l'information visuelle dans le cas d'un jugement qualitatif, elle ne remet nullement en question le fait que la différence de vitesse de traitement des BFS et HFS pourrait expliquer à elle seule la rapidité avec laquelle nous reconnaissons notre environnement visuel.

Schyns et Oliva (Oliva & Schyns, 1997; Schyns & Oliva, 1994) ont fait l'hypothèse que la rapidité et la fiabilité de la reconnaissance de scènes visuelles, en dépit de leur complexité, pouvaient être expliquées par une stratégie d'analyse « coarse-to-fine » de l'information visuelle. Selon eux, l'analyse rapide des BFS permettrait d'extraire une information grossière (« coarse ») composée de formes floues (« blobs ») orientées et arrangées de manière particulière, qui fournirait un aperçu global sur la structure spatiale de la scène suffisant pour inférer sa catégorie. Ce traitement grossier serait ensuite affiné par l'analyse plus tardive des HFS, qui permettrait cette fois-ci d'avoir accès aux détails fins (« fine ») de l'image et d'identifier précisément les objets et les contours de la scène. Afin de tester cette hypothèse, ces auteurs ont utilisé comme stimuli des images hybrides (Schyns & Oliva, 1994). Les images hybrides sont obtenues en superposant une scène filtrée en BFS et une autre scène filtrée en HFS, les deux scènes appartenant à des catégories sémantiques différentes (cf. Figure 9).

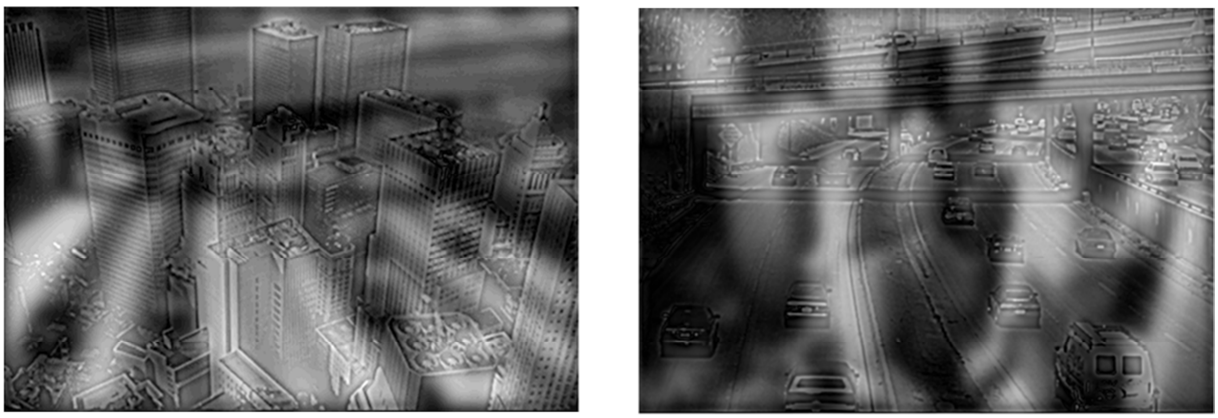


Figure 9 : Exemples d'image hybrides. A gauche, l'image hybride est composée des BFS d'une autoroute et des HFS d'une ville. A droite, l'image hybride est composée des BFS d'une ville et des HFS d'une autoroute. Adapté de Schyns et Oliva (1994).

Dans une première expérience, les auteurs ont utilisé une tâche d'appariement entre une scène amorce et une scène cible. La scène amorce était soit (1) une scène non filtrée, (2) une scène filtrée en BFS, (3) une scène filtrée en HFS ou (4) une scène hybride (BFS d'une scène et HFS d'une autre scène). L'amorce était présentée 30 ou 150 ms et suivie d'un masque. La scène cible était toujours une scène non filtrée. Dans le cas où la scène amorce était une image hybride, la scène cible pouvait correspondre soit à la partie en BFS, soit à la partie en HFS de l'image hybride. Ainsi, pour la condition hybride, le type d'appariement permettait de connaître l'information (BFS ou HFS) préférentiellement utilisée par les participants dans l'image hybride. Tout d'abord, les résultats ont montré que lorsque l'amorce était une scène filtrée en BFS, en HFS ou non filtrée, les participants appariaient correctement l'amorce et la cible, quel que soit le temps de présentation de l'amorce (30 ou 150 ms). Cela indique donc que même avec un temps de présentation très court, les informations en BFS et en HFS peuvent être correctement perçues. Lorsque l'amorce était une scène hybride, les auteurs ont observé que le type d'appariement de la scène cible, et donc l'utilisation soit des BFS, soit des HFS de l'image hybride, dépendait du temps de présentation de l'amorce. Lorsque l'amorce hybride était présentée 30 ms, la scène cible était majoritairement appariée à la scène en BFS (63 % des essais), suggérant que la perception de l'hybride était dominée par les BFS. Lorsque l'amorce hybride était présentée 150 ms, la scène cible était cette fois-ci majoritairement appariée à la scène en HFS (86 % des essais), suggérant que la perception de l'hybride était dominée par les HFS. Ces résultats indiquent que, bien que l'information véhiculée par les BFS et les HFS puisse être perçue en très peu de temps, l'information grossière véhiculée par les BFS est privilégiée pour un traitement rapide de la scène tandis que l'information fine véhiculée par les HFS est préférentiellement utilisée lorsque plus de temps est disponible pour analyser la scène. Ces résultats sont donc cohérents avec une logique de traitement « coarse-to-fine ». A noter que la tâche d'appariement utilisée dans cette expérience n'imposait pas une catégorisation explicite des scènes. Les résultats ne garantissent donc pas qu'une stratégie d'analyse « coarse-to-fine » soit effectivement utilisée lors de la catégorisation rapide des scènes. Les auteurs ont donc réalisé une deuxième expérience dans laquelle les participants devaient effectuer une tâche de catégorisation explicite de scènes (salons, chambres, villes ou autoroutes). Dans cette expérience, les stimuli étaient des séquences de deux images hybrides présentées successivement durant 45 ms chacune. Si l'on considère la séquence formée par les deux images présentées dans la Figure 9, l'utilisation d'une stratégie d'analyse « coarse-to-fine » devait conduire à catégoriser l'image comme une scène d'autoroute, tandis que l'utilisation de la stratégie inverse devait

conduire à catégoriser l'image comme une scène de ville. Les résultats ont montré une proportion plus importante de catégorisation basée sur une analyse « coarse-to-fine » (67%) que « fine-to-coarse » (29%). Ces résultats soutiennent donc encore une fois l'hypothèse d'une stratégie d'analyse « coarse-to-fine » privilégiée lors de la perception et la catégorisation rapide de scènes.

Dans une étude récente, Musel et al. (2012) ont explicitement manipulé l'ordre de présentation de scènes filtrées en fréquences spatiales. Les stimuli étaient de courts films de 150 ms composés de six versions filtrées d'une même scène (filtrage passe-bande) assemblées soit des BFS aux HFS soit des HFS aux BFS. Ainsi, ces séquences imitaient et permettaient d'imposer au système visuel une analyse « coarse-to-fine » (CtF) ou « fine-to-coarse » (FtC) des scènes, respectivement. Les participants étaient de jeunes adultes et des personnes âgées qui avaient pour tâche de catégoriser les séquences comme étant des scènes d'intérieur ou d'extérieur. Les résultats obtenus chez les participants jeunes ont montré que les séquences CtF étaient catégorisées plus rapidement que les séquences FtC, suggérant qu'une stratégie d'analyse « coarse-to-fine » est plus avantageuse pour la catégorisation de scènes chez le jeune adulte. Les résultats obtenus chez les personnes âgées ont révélé une catégorisation plus rapide des séquences CtF que FtC pour les scènes d'extérieur mais cette différence tendait à s'inverser en faveur des séquences FtC pour les scènes d'intérieur.

Les résultats présentés dans cette partie suggèrent donc qu'une stratégie d'analyse « coarse-to-fine » est privilégiée pour catégoriser rapidement une scène. Cependant, on remarque que cette stratégie n'est pas une condition *sine qua non* pour permettre la catégorisation rapide des scènes. En effet, dans l'étude de Schyns et Oliva (1994), une stratégie d'analyse « fine-to-coarse » est privilégiée dans environ 30% des cas. De plus, les informations en BFS et en HFS semblent toutes deux disponibles assez tôt au niveau du système visuel et peuvent permettre, seules, un appariement correct. Similairement, dans l'étude de Musel et al. (2012), les séquences FtC étaient catégorisées correctement par tous les participants et semblaient être privilégiées pour la catégorisation de scènes d'intérieur chez les personnes âgées. Il apparaît alors que l'utilisation préférentielle d'une stratégie d'analyse des fréquences spatiales « coarse-to-fine » ou « fine-to-coarse » pourrait différer indépendamment des contraintes perceptives de la tâche visuelle (e.g., le temps de présentation).

4.2.2. Une stratégie d'analyse flexible guidée par les contraintes de la tâche

Oliva et Schyns (1997; Schyns & Oliva, 1999) ont, par la suite, mené une série d'expériences afin de déterminer (1) si le traitement des BFS précédait nécessairement le traitement des HFS, et (2) dans quelle mesure la stratégie d'analyse des fréquences spatiales pouvait être flexible et varier en fonction des contraintes de la tâche visuelle. Dans une première expérience (Oliva & Schyns, 1997), les auteurs ont testé si à la fois les informations en BFS et en HFS d'une image hybride pouvaient être disponibles très tôt pour le système visuel (i.e. 30 ms) ou si seule l'information en BFS pouvait être disponible aussi rapidement. Le protocole expérimental était similaire à celui utilisé dans leur étude princeps (Schyns & Oliva, 1994). Une scène hybride était présentée pendant 30 ms en amorce, suivie d'une scène cible qui correspondait soit à l'information en BFS de l'amorce, soit à l'information en HFS de l'amorce, soit à une scène non présente dans l'amorce. Selon les auteurs, l'utilisation préférentielle de l'information en BFS pour des temps de présentations courts (i.e. 30 ms) mise en évidence dans leur première étude (Schyns & Oliva, 1994) pouvait être due au fait que la tâche d'appariement utilisée nécessitait de traiter explicitement l'image hybride. Cela aurait conduit à engager des processus sémantiques de haut niveau, qui auraient pu contraindre le type d'information préférentiellement traité dans l'image hybride. Dans cette nouvelle expérience, les auteurs ont donc utilisé une tâche de dénomination de la catégorie de la scène cible. Cette nouvelle tâche permettait aux participants de ne pas avoir à traiter explicitement l'amorce et la cible pour effectuer la tâche, et donc, d'examiner le traitement de l'information aux deux fréquences spatiales contenues dans l'hybride en termes purement perceptifs. Les résultats ont montré que la dénomination de la catégorie de la scène cible était plus rapide lorsque cette scène avait été présentée dans l'image hybride que lorsqu'il s'agissait d'une scène différente en amorce. Cet effet d'amorçage était le même, que la scène cible ait été présentée dans la partie en BFS ou HFS de l'image hybride. Ces premiers résultats ont permis de montrer que les informations véhiculées par les BFS et les HFS peuvent être toutes deux disponibles très tôt pour le système visuel, et que la prédominance d'une analyse « coarse-to-fine » mise en évidence dans les études précédentes (Parker et al., 1992, 1997; Schyns & Oliva, 1994) ne résulte pas nécessairement d'un traitement plus précoce des BFS par rapport aux HFS.

Dans une expérience suivante, les auteurs ont tenté d'influencer l'utilisation préférentielle de l'une ou l'autre de ces échelles spatiales. Ils ont, par exemple, montré que la sensibilisation à une bande de fréquence particulière déterminait celle qui était privilégiée

ensuite pour catégoriser des images hybrides. Cette expérience était réalisée en deux phases. Dans la première phase, les participants étaient exposés à des images pseudo-hybrides (phase de sensibilisation) qu'ils devaient catégoriser. Les images pseudo-hybrides étaient composées d'une scène en BFS et de bruit en HFS pour un groupe de participants et d'une scène en HFS et de bruit en BFS pour un autre groupe de participants. Pour effectuer la tâche de catégorisation, les participants devaient donc s'appuyer sur la partie en fréquences spatiales de l'image pseudo-hybride contenant la scène (i.e. BFS ou HFS selon le groupe). Cette phase de sensibilisation devait donc induire, pour chaque groupe, la fréquence spatiale à privilégier (ou « diagnostique » selon les mots des auteurs) pour effectuer la tâche. Dans la deuxième phase, les participants devaient catégoriser des images hybrides contenant cette fois-ci deux scènes, l'une en BFS et l'autre en HFS, appartenant à des catégories différentes. Les auteurs ont observé que les participants ayant été sensibilisés aux images pseudo-hybrides contenant la scène dans la partie en BFS privilégiaient les BFS pour catégoriser les images hybrides lors de la deuxième phase. De la même façon, les participants ayant été sensibilisé aux images pseudo-hybrides contenant la scène dans la partie en HFS privilégiaient les HFS pour catégoriser les hybrides lors de la deuxième phase.

Dans l'ensemble, les résultats de ces différentes expériences ont permis de montrer que l'utilisation de l'information en fréquences spatiales pour la catégorisation n'est pas figée et imposée par des contraintes perceptives, mais varierait en fonction de leur pertinence pour la tâche visuelle. L'influence de la tâche visuelle a été directement étudiée dans une autre série d'expériences réalisées par Schyns et Oliva (1999). Les auteurs utilisaient cette fois-ci des visages hybrides comme stimuli (cf. Figure 10). Ces visages hybrides étaient composés d'un visage d'homme ou de femme en BFS et d'un visage du sexe opposé en HFS. De plus, les visages de chaque bande de fréquence exprimaient une émotion différente (i.e. joie, colère ou neutre). Les participants avaient pour tâche de catégoriser les visages hybrides, présentés pendant un temps très court de 50 ms, soit comme étant un homme ou une femme (tâche « GENDER »), soit comme étant expressif ou non-expressif (tâche « EX-NEX »), soit comme exprimant la joie, la colère ou une expression neutre (tâche « CATEX »). Comme dans leur étude princeps (Schyns & Oliva, 1994), la réponse donnée par les participants permettait de connaître l'information en fréquences spatiales de l'image hybride privilégiée pour effectuer la tâche.

Figure 10 : Exemples de visages hybrides utilisés dans l'étude de Schyns et Oliva (1999). A gauche, un visage de femme avec une expression neutre en BFS superposé à un visage d'homme en colère en HFS. A droite, un visage d'homme en colère en BFS superposé à un visage de femme neutre en HFS.

Les résultats de la tâche « GENDER » ne montraient aucun recours préférentiel à une bande de fréquence particulière : les participants utilisaient de façon équivalente l'information en BFS et en HFS pour catégoriser le visage. En revanche, les résultats de la tâche « EX-NEX » révélaient une utilisation préférentielle de l'information en HFS pour la catégorisation, tandis que les résultats de la tâche « CATEX » montraient une utilisation préférentielle de l'information en BFS. Ces résultats indiquent donc que le type de catégorisation et les attentes liées à celle-ci influencent la façon d'analyser un stimulus. La catégorisation de l'expressivité ou non d'un visage nécessiterait l'analyse d'une information fine telle que les contours de la bouche, véhiculée par les HFS, tandis que la catégorisation de l'expression émotionnelle nécessiterait l'utilisation d'une information plus globale comprenant les yeux, la bouche, et le nez, et portée par les BFS.

Les résultats de l'ensemble des travaux présentés ci-dessus semblent donc en faveur d'une utilisation flexible de l'information en fréquences spatiales lors de la perception et la catégorisation visuelle. L'information en BFS et en HSF serait disponible très tôt pour le système visuel et l'une ou l'autre de ces informations serait privilégiée en fonction des contraintes de la tâche. Plusieurs travaux ont en effet mis en évidence par la suite que la stratégie d'analyse des fréquences spatiales pouvait dépendre de nombreux facteurs tels que l'allocation de l'attention à des fréquences spatiales particulières (Abrams, Barbot, & Carrasco, 2010; Ozgen et al., 2006; Özgen et al., 2005; Sowden et al., 2003), l'âge (Musel et al., 2012; Viggiano, Righi, & Galli, 2006) ou encore la catégorie des stimuli (Awasthi, Sowman, Friedman, & Williams, 2013; Collin & McMullen, 2005; Morrison & Schyns, 2001;

Rotshtein, Schofield, Funes, & Humphreys, 2010; Vannucci, Viggiano, & Argenti, 2001). Ozgen et al. (2005) ont, par exemple, montré que l'orientation de l'attention vers une bande de fréquence spatiale particulière sur la base d'un indice sonore (e.g., le mot « coarse » était prédictif d'une information en BFS et le mot « fine » était prédictif d'une information en BFS) influençait la bande de fréquence privilégiée pour catégoriser ensuite des images hybrides. Les images hybrides étaient préférentiellement catégorisées sur la base des BFS lorsqu'elles étaient précédées du mot « coarse », mais sur la base des HFS lorsqu'elles étaient précédées du mot « fine ». Des travaux sur la catégorisation d'objets (Vannucci et al., 2001) ont montré que la fréquence spatiale privilégiée pour la catégorisation d'objets différait en fonction de leur catégorie. Par exemple, les BFS seraient privilégiées pour la catégorisation d'animaux et d'outils, tandis que les HFS seraient privilégiées pour la catégorisation de légumes. Plus récemment, Rotshtein et al. (2010) ont montré que l'information en fréquences spatiales privilégiée pour catégoriser des images hybrides ne serait pas spécifique à la catégorie du stimulus visuel mais varierait en fonction du type de catégorisation à effectuer. Par exemple, les BFS seraient privilégiées pour des tâches de catégorisation de type fleur *versus* visage ou visage d'homme *versus* de femme, tandis que les HFS seraient privilégiées pour des tâches de catégorisation de type maison *versus* fleur, visage *versus* maison ou expression positive *versus* négative dans des visages. Une analyse « coarse-to-fine » ne serait donc pas contrainte et nécessaire pour percevoir correctement un stimulus visuel, mais serait un mode de traitement par défaut, privilégié lors de la catégorisation rapide.

L'ensemble de ces travaux soulignent donc le rôle d'influences de traitements de haut niveau sur les traitements plus bas niveau de l'information visuelle (i.e. influences « top-down »). De nombreux travaux se sont attachés par la suite à modéliser de façon computationnelle la catégorisation de scènes naturelles, en se basant sur l'extraction de propriétés visuelles de bas niveau telles que les fréquences spatiales. Ces travaux se sont beaucoup basés sur les propriétés ou « statistiques » des scènes naturelles, qui présentent des régularités sur lesquelles le système visuel pourrait s'appuyer pour la catégorisation.

4.3. Arguments computationnels

Schyns et Oliva (1994) postulaient que, lors de la perception de scènes, l'extraction rapide de l'information grossière portée par les BFS fournirait une sorte de canevas, permettant une première catégorisation perceptive, qui serait ensuite complété par l'analyse

fine des détails et des contours des objets, véhiculés par les HFS. De façon intuitive, il est cependant difficile de concevoir comment l'information floue, grossière, portée par les BFS pourrait renseigner sur la catégorie sémantique de la scène visuelle tant la variabilité d'une scène à l'autre est importante. La Figure 11, qui représente l'information moyennée de 100 images de visages et de scènes de différentes catégories, illustre bien ce point. Contrairement aux visages qui présentent de nombreuses régularités dans le domaine spatial, les scènes se caractérisent par une multitude d'orientations, de couleurs et de configurations spatiales différentes. Il semble alors impossible sur la seule base de caractéristiques grossières de pouvoir discriminer une scène de montagne d'une scène d'autoroute ou d'intérieur**.**

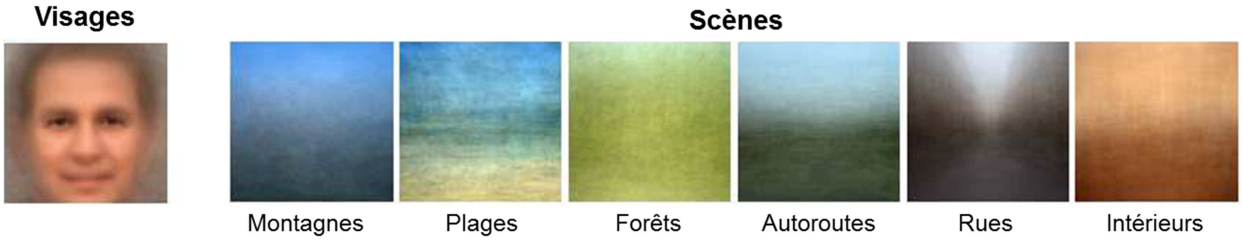


Figure 11 : Résultat du moyennage de 100 exemplaires de visages et de scènes appartenant à différentes catégories. On peut constater que les visages contiennent plus de régularités dans le domaine spatial que les scènes, qui sont difficilement reconnaissables une fois moyennées. Adapté de Torralba et Oliva (2003).

Les scènes naturelles présentent pourtant de nombreuses régularités caractéristiques de leur catégorie dans le domaine de Fourier, et notamment dans leur représentation par le spectre d'amplitude. Tout d'abord, les scènes naturelles, relativement à d'autres stimuli visuels tels que les visages ou les objets, ont la propriété d'avoir un spectre d'amplitude qui décroit avec les fréquences spatiales, selon une fonction en $1/f^{\alpha}$, avec f correspondant à la fréquence spatiale et α compris entre 0,8 et 1,5 (Burton & Moorhead, 1987; Field, 1987; Tolhurst, Tadmor, & Chao, 1992; Van Der Schaaf & Van Hateren, 1996). Cela implique que les BFS des scènes naturelles se caractérisent par un fort contraste de luminance tandis que les HFS se caractérisent par un faible contraste de luminance (nous verrons par la suite que cette caractéristique des scènes est importante à prendre en compte dans le cadre de nos travaux). La majorité de l'énergie (i.e. carré de l'amplitude) des scènes naturelles est donc contenue dans les basses et moyennes fréquences spatiales. Dès lors, il devient plus facile de concevoir qu'une quantité suffisante d'information soit contenue dans les BFS d'une scène pour permettre tout au moins une catégorisation grossière. Des travaux ont spécifiquement montré que la distribution de l'information en fréquences spatiales en fonction des orientations dans

le spectre d'amplitude permettrait de distinguer différentes catégories. Par exemple, Switkes, Mayer et Sloan (1978) ont étudié les spectres d'amplitude de scènes de différentes catégories (intérieurs, villes et paysages) et ont observé que le spectre des scènes d'intérieur et de villes était dominé par une répartition des fréquences spatiales de 1 à 25 cpd dans les orientations verticales et horizontales, tandis que le spectre des paysages étaient caractérisés par une répartition de l'information en fréquences spatiales sur toutes les orientations. De plus, les auteurs ont montré que la majorité des orientations pour les paysages et les orientations horizontales pour les scènes d'intérieur et de villes étaient principalement véhiculées par les BSF de 1 à 5 cpd. Ces résultats ont donc suggéré, d'une part, que la majorité de l'information d'une scène est contenue dans les BFS et, d'autre part, que les scènes d'une catégorie possèdent un spectre d'amplitude caractéristique et qu'elles se différencient donc des scènes d'une autre catégorie par une distribution différente de l'information en fréquences spatiales en fonction des orientations. Par la suite, Torralba et Oliva (2003) ont étudié plus systématiquement les propriétés spectrales des scènes en moyennant les spectres d'amplitude de scènes de 14 catégories différentes et ont pu mettre en évidence différentes signatures spectrales caractéristiques de chaque catégorie (cf. Figure 12). De manière générale, on observe que les scènes contenant des éléments artificiels (e.g., intérieurs, villes, autoroutes) se caractérisent par un spectre d'amplitude en forme de « croix », dominé par des orientations verticales et horizontales. Les scènes naturelles ouvertes telles que les plages, champs ou côtes se caractérisent par un spectre d'amplitude dominé par des orientations horizontales. Enfin, les scènes d'environnements texturés ou contenant des éléments proches de l'observateur (e.g., montagnes, forêts, objets, portraits) se caractérisent par une répartition des fréquences spatiales sur toutes les orientations (Oliva, Torralba, Guérin-dugué, & Hérault, 1999; Oliva & Torralba, 2001; Torralba & Oliva, 2003).

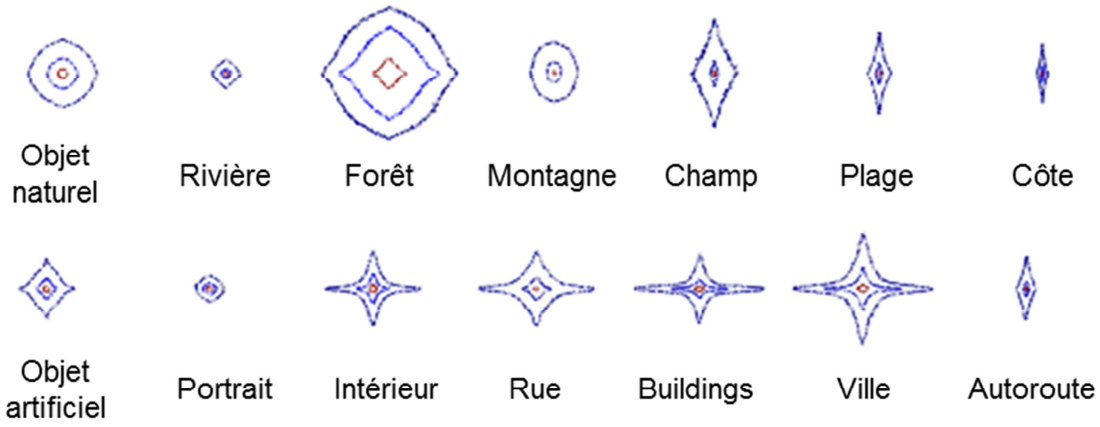


Figure 12 : Spectres d'amplitude moyens d'images de scènes de 14 catégories. L'axe horizontal dans le domaine de Fourier représente les orientations verticales dans le domaine spatial et l'axe vertical dans le domaine de Fourier représente des orientations horizontales dans le domaine spatial. Les trois contours correspondent, de l'extérieur vers l'intérieur, à 90, 80 et 60% de l'énergie du spectre. Adapté de Torralba et Oliva (2003).

Ces études indiquent donc qu'en dépit de la grande variabilité entre les scènes d'une même catégorie dans le domaine spatial, des régularités émergent dans le domaine de Fourier en termes de répartition des fréquences spatiales en fonction des orientations. Ces travaux soulignent également l'importance des BFS qui véhiculent la majorité de l'information de la scène visuelle.

Les travaux en modélisation computationnelle utilisent des systèmes visuels artificiels bio-inspirés simulant les propriétés neurophysiologiques du système visuel humain. Field (1987) propose que les cellules du cortex visuel primaire agissent comme des filtres, permettant d'extraire la fréquence et la phase d'une partie locale du signal visuel selon différentes orientations. Herault, Oliva et Guerin-Dugué (1997) se sont appuyés sur ce principe pour simuler le fonctionnement du système visuel (i.e. échantillonnage des scènes au niveau local par des filtres passe-bandes orientés ou filtres de Gabor) et ont montré que ce système artificiel était en mesure de classer les scènes en catégories sémantiques distinctes sur la base de l'information en basses et moyennes fréquences, et ce, de façon similaire à des observateurs humains. D'autres travaux ont conduit à des résultats similaires, en utilisant également des systèmes artificiels classant des scènes naturelles sur la base d'orientations locales (Guérin-dugué & Oliva, 2000) ou du spectre d'amplitude global (Guyader, Chauvin, Peyrin, Hérault, & Marendaz, 2004; Guyader & Hérault, 2001; Oliva et al., 1999; Oliva & Torralba, 2001; Torralba & Oliva, 1999, 2003). L'ensemble de ces travaux soutient donc

l'idée que l'information véhiculée principalement par les BFS et liée aux orientations dominantes des scènes est suffisante pour permettre leur catégorisation.

Dans la continuité de ce travail, Mermillod et al. (2005) se sont intéressés à la bande de fréquence spatiale la plus utile pour la catégorisation de scènes appartenant à six catégories (forêts, villes, intérieurs, montagnes, villages et plages). Ces auteurs ont utilisé un système artificiel composé de filtres de Gabor échantillonnant le signal de la scène sur six orientations et cinq bandes de fréquences spatiales. Les résultats ont montré que le système classait les scènes de villes, de plages et de villages sur la base des BFS, et les scènes de forêts, d'intérieur et de montagnes sur la base des moyennes et HFS. Ces résultats indiquent donc que l'information en BFS n'est pas forcément la bande de fréquences la plus pertinente pour la catégorisation de scènes de différentes catégories et viennent à l'appui d'études comportementales en faveur d'une utilisation flexible de l'information en fréquences spatiales en fonction de leur pertinence pour la tâche (Oliva & Schyns, 1997; Schyns & Oliva, 1999).

5. Problématique

L'ensemble des travaux présentés dans ce chapitre indique que le système visuel analyse l'information visuelle en termes de fréquences spatiales selon un mode de traitement, par défaut, « coarse-to-fine ». Cependant, si cette stratégie d'analyse a été très étudiée au niveau comportemental, ses bases cérébrales sont encore méconnues. L'objectif de cette thèse a été d'étudier les bases cérébrales de l'analyse « coarse-to-fine » lors de la catégorisation de scènes, au moyen d'études comportementales et en IRMf.

En utilisant des films qui simulent des séquences « coarse-to-fine » et « fine-to-coarse » des scènes, Musel et al. (2012) ont récemment montré que les séquences CtF sont catégorisées plus rapidement que les séquences FtC chez le jeune adulte sain, suggérant que l'ordre de présentation des fréquences spatiales est déterminant pour la catégorisation rapide de scènes. D'autres travaux utilisant également des séquences de scènes filtrées ont cependant montré que l'efficacité de la catégorisation serait plus liée à la quantité d'information en fréquences spatiales disponible dans le stimulus, plutôt qu'à l'ordre relatif du traitement des fréquences spatiales (i.e. traitement des BSF avant les HFS ou inversement). Dans une première étude comportementale (**Expérience 1**), nous avons donc tenté de déterminer le rôle relatif de l'ordre de traitement des fréquences spatiales et de la quantité d'information en fréquences spatiales pour la catégorisation rapide de scènes. Par ailleurs, nous avons vu que le traitement des fréquences spatiales est intimement lié à la perception du contraste. Si le rôle

relatif des fréquences spatiales et du contraste de luminance ont été établis pour la détection de réseaux sinusoïdaux dans des études psychophysiques, cela n'a jamais été étudié dans le contexte de la perception des scènes. L'information de contraste doit pourtant être considérée avec attention lors de la catégorisation de scènes puisque ces stimuli se caractérisent par une répartition asymétrique du contraste de luminance en fonction des fréquences spatiales. Nous avons donc mené une deuxième étude comportementale (**Expérience 2**) afin de déterminer le rôle du contraste de luminance associé aux fréquences spatiales pour la catégorisation rapide de scènes.

Au niveau neurobiologique, les bases cérébrales du traitement des fréquences spatiales restent à préciser. En effet, des travaux suggèrent une organisation rétinotopique du traitement des fréquences spatiales au niveau du cortex visuel, mais également une spécialisation hémisphérique du traitement des fréquences spatiales, chaque hémisphère traitant de façon prédominante une bande de fréquences spatiales. Par ailleurs, les modèles actuels de la perception visuelle et de la reconnaissance postulent que l'analyse « coarse-to-fine » de l'information visuelle impliquerait un large réseau d'aires cérébrales au-delà du cortex visuel primaire. Selon ces modèles, l'analyse rapide des BSF permettrait d'activer de façon précoce des aires cérébrales de haut niveau (pariétales, temporales et frontales), afin d'activer des représentations grossières du stimulus visuel. Cela permettrait de déclencher des influences « top-down » sur des régions de plus bas niveau (occipito-temporales), afin de guider le traitement plus tardif des HFS et de faciliter la catégorisation et la reconnaissance. Cependant, les mécanismes et les interactions cérébrales sous-jacents sont encore méconnus. Nous présenterons donc dans le Chapitre 3 les résultats de deux études IRMf visant à identifier les bases cérébrales de l'analyse « coarse-to-fine » et à préciser la connectivité au sein du réseau cérébral impliqué dans cette stratégie d'analyse « coarse-to-fine » des scènes (**Expérience 3**). Nous porterons une attention particulière aux corrélats cérébraux d'une potentielle influence du traitement des BFS sur le traitement des HFS (**Expérience 4**).

Enfin, il existe au niveau du cortex occipito-temporal une multitude de régions sélectives à différentes catégories de stimuli. En particulier, trois régions ont été mises en évidence comme sélectives aux scènes : la « parahippocampal place area », le cortex retrosplenial et « l'occipital place area ». De nombreuses études ont montré que ces régions sont impliquées dans des traitements de haut niveau lors de la perception de scènes, telles que la navigation, la reconnaissance ou dans le traitement d'informations sémantiques liées au contexte spatial ou non spatial de l'environnement visuel. Cependant, très peu de travaux se sont penchés sur la sensibilité de ces régions aux fréquences spatiales dans les scènes. Afin de

préciser les bases cérébrales du traitement des fréquences spatiales lors de la perception de scènes, nous avons donc directement étudié à travers deux études IRMf (**Expériences 6 et 7)** comment ces régions dédiées au traitement des scènes répondent aux fréquences spatiales lors de la catégorisation de scènes, et si un mode de traitement « coarse-to-fine » est privilégié par ces régions.

Ce travail de thèse se situe dans l'objectif général d'élaborer un modèle neurocognitif biologiquement plausible de la catégorisation de scènes naturelles en utilisant de nouveaux paradigmes expérimentaux construits sur la base des propriétés neurophysiologiques des voies rétino-corticales.

RESUME

L'ensemble des travaux présentés dans ce chapitre indique que le système visuel extrait en parallèle au niveau local une information à différentes fréquences spatiales et différentes orientations. Du fait des propriétés neurophysiologiques des voies visuelles, l'information en BFS, véhiculée par la voie M serait traitée plus rapidement que l'information en HFS véhiculée par la voie P. L'analyse visuelle se ferait donc selon un mode de traitement par défaut « coarse-to-fine ». L'analyse rapide de l'information grossière véhiculée par les BFS fournirait un aperçu global de la scène, permettant une première catégorisation perceptive, qui serait ensuite affinée, validée ou rejetée sur la base de l'analyse plus tardive de l'information détaillée véhiculée par les HFS. Une analyse « coarse-to-fine », bien qu'avantageuse pour la catégorisation rapide, ne serait cependant pas systématique. En effet, les informations véhiculées par les BFS et les HFS pourraient être disponibles très rapidement pour le système visuel et différentes stratégies d'analyse des fréquences spatiales pourraient être utilisées en fonction des contraintes de la tâche. Le traitement des fréquences spatiales serait donc flexible et guidé par des processus de plus haut niveau. Cependant, si cette stratégie d'analyse a été très étudiée au niveau comportemental, les bases cérébrales de l'analyse « coarse-to-fine » sont encore méconnues. L'objectif de cette thèse qui s'inscrit dans l'objectif plus général d'élaborer un modèle biologiquement plausible de la perception visuelle de scènes a été d'identifier le réseau cortical sous-jacent à une analyse « coarse-to-fine », ainsi que les interactions et les influences entre les différentes régions de ce réseau.

CHAPITRE 2 – CATEGORISATION RAPIDE DE SCENES : ETUDE COMPORTEMENTALE DU TRAITEMENT DES FREQUENCES SPATIALES ET INFLUENCE DES PARAMETRES DE LA STIMULATION VISUELLE

En utilisant des films qui simulent des séquences « coarse-to-fine » et « fine-to-coarse », Musel et al. (2012) ont montré que les séquences « coarse-to-fine » étaient catégorisées plus rapidement que les séquences « fine-to-coarse » chez le jeune adulte sain, suggérant que cette stratégie d'analyse est avantageuse pour la catégorisation rapide de scènes. Le paradigme utilisé dans cette étude semble donc approprié pour imposer une analyse « coarse-to-fine » ou « fine-to-coarse » des scènes afin d'en étudier les bases cérébrales. Les stimuli tels qu'utilisés dans l'étude de Musel et al. (2012) présentent cependant quelques problèmes méthodologiques. Dans le cadre de ce travail de thèse, nous avons donc adapté ces stimuli comme décrit ci-après.

1. Les scènes dynamiques

Musel et al. (2012) ont utilisé des scènes dynamiques permettant d'imposer au système visuel une analyse « coarse-to-fine » ou « fine-to-coarse » des scènes. Ces stimuli sont des séquences construites à partir de six filtrages passe-bande d'une même scène, centrés sur différentes fréquences spatiales définies selon une échelle linéaire (i.e. filtres passe-bande centrés sur des fréquences spatiales correspondant à 1, 2, 3, 4, 5 et 6 cpd). Cependant, dans cette étude, le filtre centré sur les plus basses fréquences laissait passer toute l'information en BFS inférieure à 2,7 cpd et correspondait donc à un filtre passe-bas. Par ailleurs, nous avons vu que les scènes naturelles ont la propriété d'avoir un spectre d'amplitude qui décroit avec les fréquences spatiales, de telle sorte que la majorité de l'information véhiculée par la scène est contenue dans les BFS (Field, 1987, 1989; Tolhurst et al., 1992). Dans l'étude de Musel et al. (2012), l'utilisation d'une échelle linéaire pour définir les fréquences centrales des filtres avait donc pour conséquence de présenter la majorité de l'information d'une scène dans les deux premières images des séquences CtF (i.e. deux dernières images des séquences FtC). Comme nous l'avons vu précédemment, l'information contenue dans les BFS des scènes est

suffisante pour permettre leur catégorisation (Hérault et al., 1997; Torralba & Oliva, 2003). Il est donc possible que l'avantage des séquences « coarse-to-fine » dans l'étude de Musel et al. (2012) soit dû au fait que les premières images de ces séquences contenaient assez d'information pour permettre la catégorisation, indépendamment de l'information en fréquences spatiales contenues dans les images suivantes. Ces stimuli ont donc été modifiés, afin de pallier ces biais potentiels. Tout d'abord, nous avons modifié les paramètres de filtrage des scènes composant les séquences de telle sorte que le filtre centré sur les plus basses fréquences spatiales soit un filtre passe-bande et ne laisse pas passer l'information en BFS inférieure à 0,3 cpd. De plus, nous avons modifié l'échelle utilisée pour définir les fréquences centrales des filtres afin d'obtenir un meilleur échantillonnage de l'information en fréquences spatiales. Nous avons donc utilisé une échelle logarithmique qui permet d'avoir plus de filtres centrés sur les BFS (Willenbockel et al., 2010). Les fréquences centrales des filtres utilisés pour construire les séquences utilisées dans le cadre de ces travaux correspondaient donc à 1 ; 1,4 ; 2 ; 2,9 ; 4,2 ; et 6 cpd. Il est à noter que ces modifications des stimuli ont été évaluées à travers des études comportementales préalables à ce travail de thèse qui ne seront pas détaillées ici. Ces nouveaux stimuli ont été utilisés dans deux études comportementales afin d'étudier l'influence de facteurs tels que l'ordre de présentation, le mode d'accumulation et le contraste de luminance des fréquences spatiales pour la catégorisation rapide de scènes.

2. Influence de l'ordre, du mode d'accumulation et du contraste de luminance des fréquences spatiales pour la catégorisation rapide de scènes

Les différents travaux rapportés dans le Chapitre 2 indiquaient que, par défaut, une analyse « coarse-to-fine » est privilégiée pour la catégorisation rapide de scènes (Schyns & Oliva, 1994). En effet, nous avons vu que l'information issue des BFS serait suffisante pour avoir rapidement accès à la catégorie de la scène, avant une analyse plus détaillée de l'information fine véhiculée par les HFS permettant d'identifier précisément la scène et les objets qui la composent. De plus, les travaux de Musel et al. (2012) ont mis en évidence le rôle fondamental de l'ordre de traitement des fréquences spatiales pour la catégorisation rapide de scènes en démontrant que la catégorisation était avantagée par le traitement des BFS avant les HFS par rapport à un traitement dans l'ordre inverse. Cependant, les mécanismes d'intégration de l'information en fréquences spatiales lors de la catégorisation de scènes sont peu connus. Des études récentes en électroencéphalographie (EEG) se sont intéressées à la

quantité minimum d'information en fréquences spatiales nécessaire pour la reconnaissance de scènes (De Cesarei & Codispoti, 2011; De Cesarei, Mastria, & Codispoti, 2013). Dans ces études, les auteurs ont présenté des séquences de 16 scènes filtrées, au cours desquelles le contenu d'une scène était progressivement révélé, via l'ajout progressif d'information en fréquences spatiales d'une image à l'autre dans la séquence (cf. Figure 13). Ainsi, les séquences débutaient soit par une scène filtrée en BFS, à laquelle de l'information en HFS était progressivement ajoutée (i.e. augmentation de la fréquence de coupure du filtrage), soit par une scène en HFS, à laquelle de l'information en BFS était progressivement ajoutée (i.e. diminution de la fréquence de coupure du filtrage).

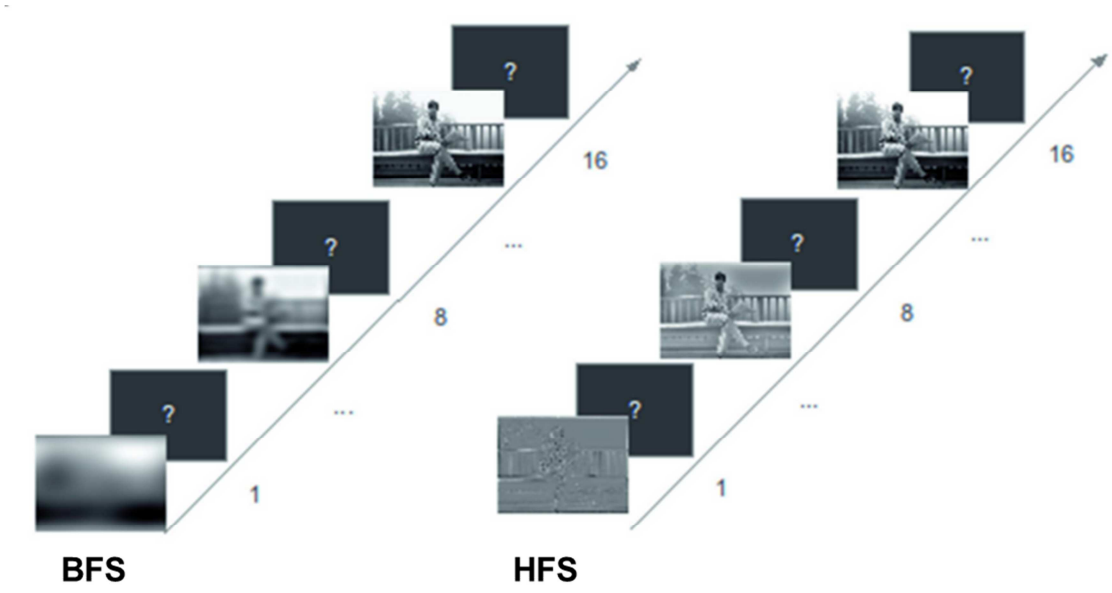


Figure 13 : Exemple de stimuli utilisés par De Cesarei et al. (2011, 2013). Les séquences sont composées de 16 versions filtrées d'une même scène. Les séquences débutent soit par une image en BFS, à laquelle de l'information en HFS est progressivement ajoutée, soit par une image en HFS à laquelle de l'information en BFS est progressivement ajoutée. Adapté de De Cesarei et al. (2013).

Les résultats de ces études ont montré que plus on ajoutait de l'information en fréquences spatiales, plus le taux d'identification des scènes augmentait, indépendamment du contenu initial en fréquences spatiales des séquences (i.e. BFS ou HFS). Ces résultats ont donc suggéré que l'identification des scènes ne dépend pas crucialement du traitement d'une information en BFS avant une information en HFS, mais plutôt de l'accumulation de façon additive de l'information en fréquences spatiales, indépendamment de l'ordre de traitement de celles-ci. Dans une première étude comportementale (**Expérience 1**), nous avons donc testé dans quelle mesure la catégorisation rapide de scènes était facilitée par l'ordre de présentation de l'information en fréquences spatiales (séquence « coarse-to-fine » *versus* « fine-to-

coarse ») et/ou par le mode d'accumulation des fréquences spatiales (accumulation successive *versus* additive).

Par ailleurs, comme nous l'avons vu dans le chapitre précédent, les propriétés statistiques des scènes sont telles que les BFS sont caractérisées par un fort contraste de luminance et les HFS par un faible contraste de luminance, du fait de la décroissance de l'énergie avec les fréquences spatiales. Or, plusieurs études ont montré qu'un stimulus visuel à fort contraste est détecté plus rapidement qu'un stimulus à faible contraste (Harwerth & Levi, 1978; Lupp et al., 1976; Parker, 1980). Dès lors, il est donc possible que l'avantage à traiter les BFS avant les HFS pour la catégorisation (Musel et al. 2012) résulte simplement d'un avantage à traiter une information très contrastée avant une information moins contrastée. Dans une deuxième étude comportementale (**Expérience 2**) nous avons donc directement testé le rôle respectif des fréquences spatiales et du contraste de luminance pour la catégorisation rapide de scènes.

2.1. EXPERIENCE 1 : Influence de l'ordre et du mode d'accumulation de l'information en fréquences spatiales pour la catégorisation de scènes

L'**Expérience 1** est présentée en détail dans l'article ci-après publié dans la revue *Vision Research* (Kauffmann, Chauvin, Guyader & Peyrin, 2015). Dans cette première étude comportementale, nous avons examiné l'influence de l'ordre de présentation des fréquences spatiales (i.e. des BSF aux HFS ou des HFS aux BFS) pour la catégorisation de scènes. De plus, nous avons testé si une analyse séparée et successive de différentes bandes de fréquences spatiales était plus avantageuse pour la catégorisation rapide de scènes qu'une analyse au cours de laquelle le contenu en fréquences spatiales est accumulé de façon additive. Pour cela, nous avons utilisé, comme dans Musel et al. (2012), des séquences dynamiques composées de six versions filtrées d'une même scène. Dans la condition Successive (Figure 14), les scènes étaient filtrées par six filtres passe-bandes de même taille, mais centrés sur différentes fréquences spatiales. Le contenu en fréquences spatiales changeait donc d'une image à l'autre et celles-ci étaient assemblées soit des BFS aux HFS (séquences « coarse-to-fine », CtF), soit des HFS aux BFS (séquences « fine-to-coarse », FtC). Dans la condition Additive (Figure 14), les scènes étaient filtrées par six filtres passe-bandes dont la largeur augmentait. Le contenu en fréquences spatiales augmentait donc progressivement d'une image à l'autre et celles-ci étaient assemblées soit des BFS à toute l'information en fréquences spatiales (séquences CtF), soit des HFS à toute l'information en fréquences spatiales (séquences FtC). Les scènes

dynamiques étaient présentées 150 ms (i.e. 25 ms par image dans la séquence), suivies d'un masque présenté 30 ms afin d'éviter une persistance rétinienne de la dernière scène de la séquence. Les participants avaient pour tâche de catégoriser les scènes comme étant soit des scènes d'intérieur, soit des scènes d'extérieur.

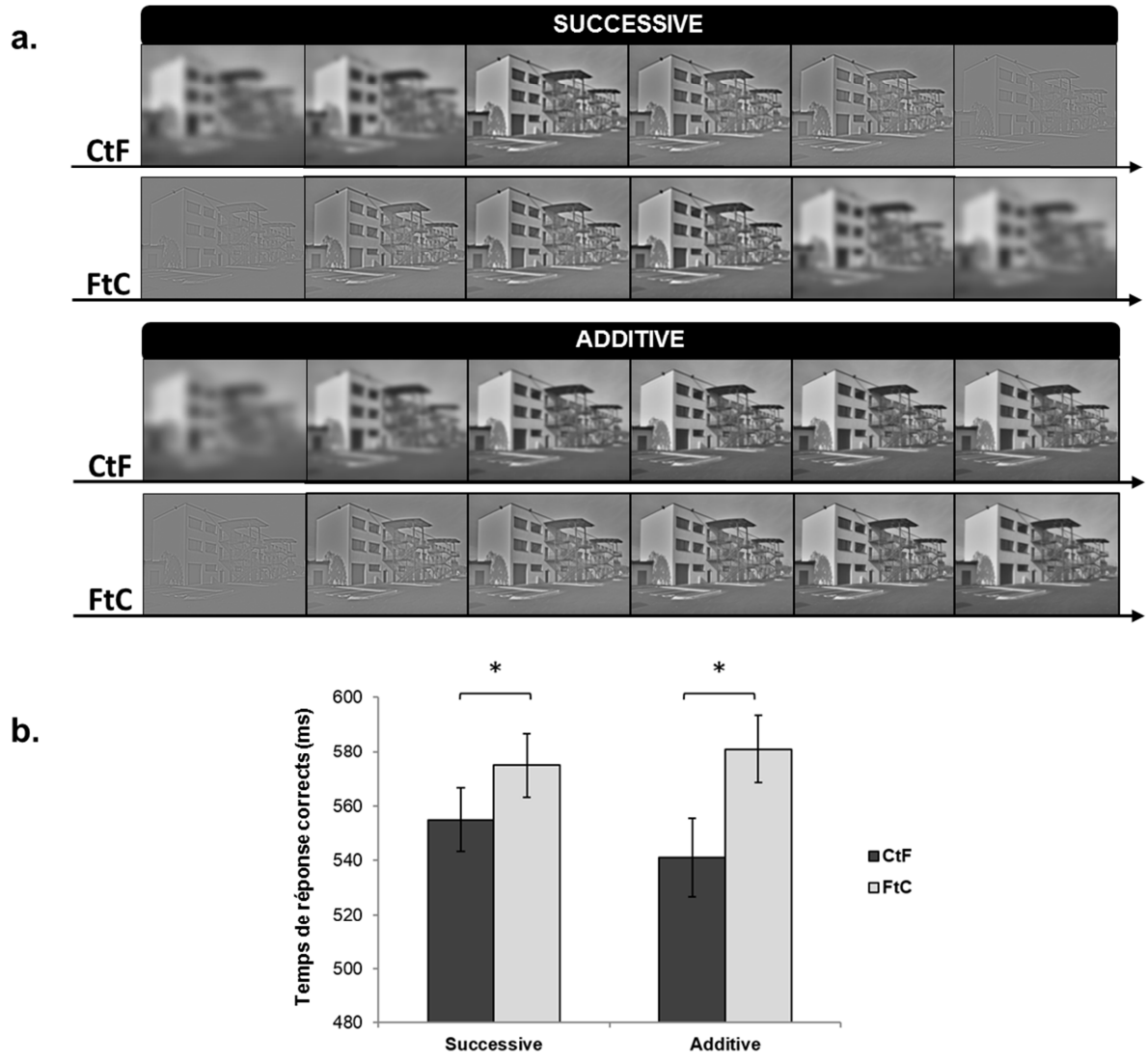


Figure 14 : Résumé des résultats de l'Expérience 1. (a) Exemples de séquences utilisées dans l'Expérience 1. Pour les séquences Successives, l'information en fréquences spatiales change d'une image à l'autre dans la séquence et passe des BFS aux HFS (séquence CtF) ou des HFS aux BFS (séquence FtC). Pour les séquences Additives, de l'information en fréquences spatiales est ajoutée d'une image à l'autre dans la séquence et passe soit des BFS à une scène intacte (séquence CtF), soit des HFS à une image intacte (séquence FtC). (b) Temps de réponse corrects (en millisecondes) pour catégoriser les séquences CtF et FtC dans les conditions Successives et Additives.

Les résultats ont montré que les séquences CtF étaient catégorisées plus rapidement que les séquences FtC, quelle que soit la progression des fréquences spatiales au cours de la séquence (i.e. Successive ou Additive). Cependant, nous n'avons pas observé de différence

entre les catégorisations de scènes Successives et Additives, ni d'interaction entre l'ordre de présentation et le mode d'accumulation des fréquences spatiales. Ces résultats suggèrent que l'accumulation de façon additive de l'information en fréquences spatiales (condition Additive), et donc la réinjection au cours du temps d'une information précédemment traitée n'avantage pas le processus de catégorisation par rapport à une analyse séparée de différentes bandes de fréquences spatiales (condition Successive). La catégorisation rapide des scènes serait en revanche principalement facilitée par l'analyse des BFS avant les HFS, en accord avec les travaux précédemment mentionnés (e.g., Musel et al., 2012; Schyns & Oliva, 1994).

2.2. EXPERIENCE 2 : Influence du contraste de luminance pour l'avantage d'une analyse « coarse-to-fine » des scènes

L'**Expérience 2** est présentée en détail dans l'article ci-dessous publié dans la revue *Vision Research* (Kauffmann et al., 2015). Dans cette deuxième étude comportementale, nous avons cherché à comprendre si des différences de contraste entre les fréquences spatiales seraient en fait le facteur déterminant d'une analyse « coarse-to-fine ». En effet, les résultats de l'**Expérience 1** suggèrent que le traitement des BFS avant les HFS est plus avantageux pour la catégorisation. Cependant, étant donné que les BFS et les HFS se caractérisent respectivement par un fort et un faible contraste de luminance, il est possible que l'avantage de l'analyse « coarse-to-fine » sur l'analyse « fine-to-coarse » soit dû à un avantage à traiter une information très contrastée avant une information peu contrastée, plutôt qu'à un avantage à traiter les BFS avant les HFS. Il est donc difficile de déterminer dans quelle mesure l'avantage des séquences « coarse-to-fine » résulte d'un effet des fréquences spatiales uniquement, d'un effet du contraste, ou de la combinaison de ces deux facteurs. Afin d'éviter une confusion entre l'effet des fréquences spatiales et l'effet du contraste, des études récentes sur la catégorisation de visages utilisant des stimuli filtrés en fréquences spatiales égalisent le contraste RMS entre les stimuli filtrées (e.g., Goffaux et al., 2011; Vlamings, Goffaux, & Kemner, 2009). Cependant, à notre connaissance, l'effet d'une égalisation du contraste des stimuli en BFS et HFS n'a jamais été directement testé dans le cadre de la catégorisation de scènes.

Dans cette deuxième étude comportementale, nous avons utilisé un paradigme de catégorisation « coarse-to-fine » ou « »fine-to-coarse » de scènes nous permettant de tester de façon isolée l'effet des fréquences spatiales, d'une part, et celui du contraste, d'autre part. Pour cela nous avons utilisé comme stimuli des scènes dynamiques similaires à la condition

Successive de l'**Expérience 1**, composées de six versions d'une même scène, selon trois conditions (cf. Figure 15a). Dans une première condition expérimentale (FS-CON), les séquences étaient composées de six versions filtrées d'une même scène, et le contraste des images filtrées n'était pas modifié (dans cette condition, les séquences étaient les mêmes que les séquences Successives de l'**Expérience 1**). Les séquences passaient donc des BFS avec un fort contraste aux HFS avec un faible contraste (séquences CtF), ou des HFS avec un faible contraste aux BFS avec un fort contraste (séquences FtC). Cette condition nous permettait donc de tester l'effet combiné des fréquences spatiales et du contraste. Dans une deuxième condition (FS), les séquences étaient composées de six versions filtrées d'une même scène, et le contraste RMS des images filtrées était égalisé (i.e. les scènes filtrées composant la séquence avaient toutes le même écart-type de valeurs de luminance). Les séquences passaient donc des BFS aux HFS (séquences CtF) ou des HFS aux BFS (séquences FtC). Cette condition nous permettait de tester l'effet des fréquences spatiales indépendamment de l'effet du contraste. Enfin, dans la troisième condition (CON), les séquences étaient composées de six versions non filtrées d'une même scène Pour chaque scène on attribuait la valeur du contraste RMS correspondant à celle des scènes filtrées des séquences de la condition FS-CON. Les séquences passaient donc d'un fort à un faible niveau de contraste (séquences CtF) ou d'un faible à un fort niveau de contraste (séquences FtC). Cette dernière condition nous permettait de tester l'effet du contraste, indépendamment de l'effet des fréquences spatiales.

Les résultats (cf. Figure 15b) ont montré que les séquences CtF étaient catégorisées plus rapidement que les séquences FtC quelle que soit la condition de contraste (FS-CON, FS et CON). Cependant, nous avons observé que l'avantage des séquences CtF sur les séquences FtC était plus important dans la condition SF-CON que dans la condition CON. Ces résultats suggèrent que si la différence de contraste entre les fréquences spatiales explique en partie l'avantage d'une analyse CtF pour la catégorisation rapide, celui-ci résulte de façon prédominante d'un effet des fréquences spatiales. Ces résultats remettent donc en question l'utilisation de l'égalisation du contraste entre des scènes filtrées pour l'étude du traitement des fréquences spatiales.

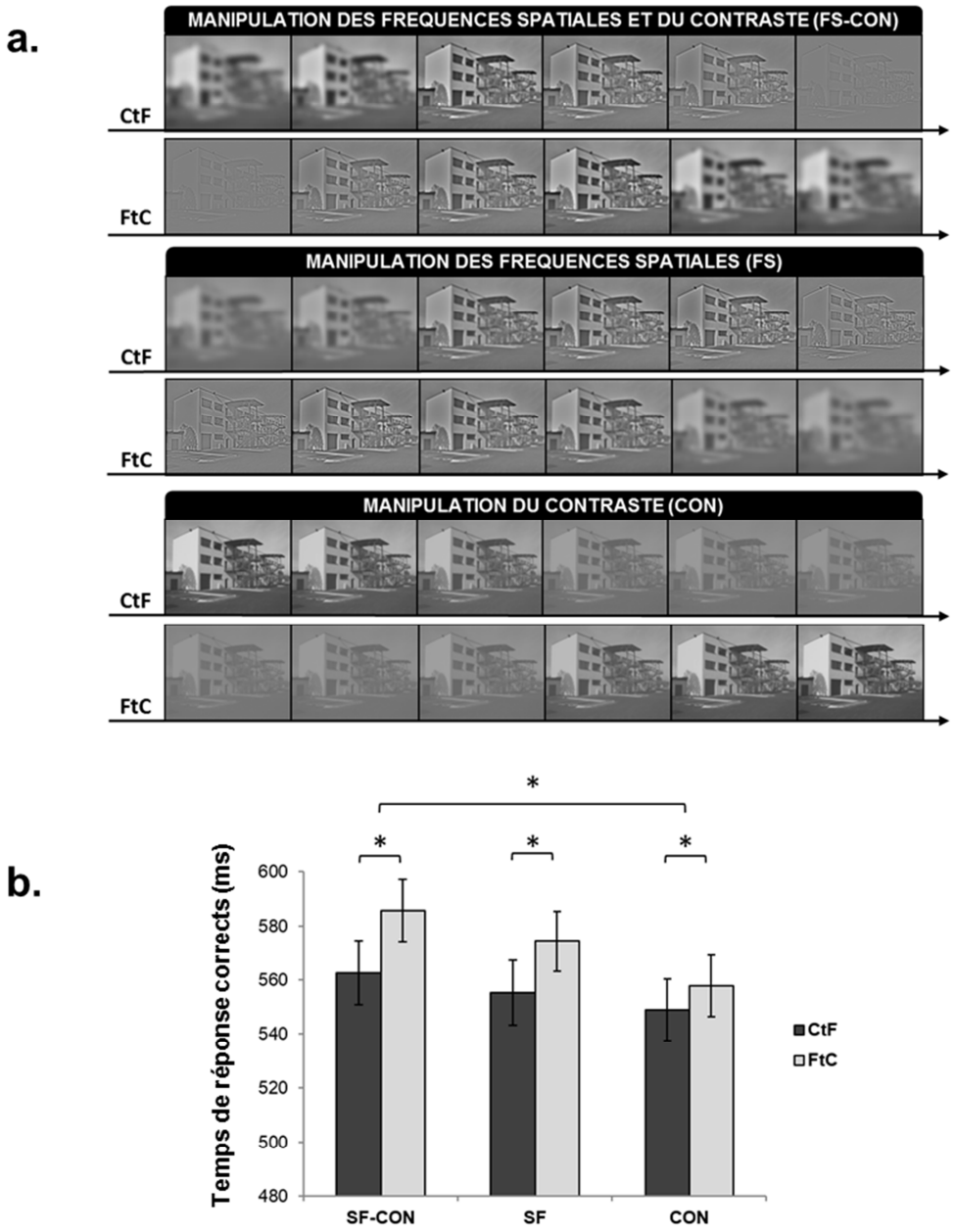


Figure 15 : Résumé des résultats de l'Expérience 2. (a) Exemple de stimuli utilisés dans l'Expérience 2. Dans la condition FS-CON, le contraste des scènes filtrées n'est pas modifié. Les séquences passent des BFS avec un fort contraste aux HFS avec un faible contraste (séquences CtF) ou inversement (séquences FtC). Dans la condition FS, le contraste est égalisé entre les scènes filtrées. Les séquences passent des BFS aux HFS (séquences CtF) ou des HFS aux BFS (séquences FtC). Dans la condition CON les scènes composant les séquences ne sont pas filtrées mais le contraste est modifié pour correspondre à celui des scènes de la condition SF-CON. Les séquences passent d'un fort à un faible contraste (séquences CtF) ou d'un faible à un fort contraste (séquence FtC). (b) Temps de réponse corrects (en millisecondes) pour catégoriser les séquences CtF et FtC des conditions SF-CON, SF et CON.

Article publié dans la revue *Vision Research* (2015)

Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast

Louise Kauffmann[1,2*], Alan Chauvin[1,2] , Nathalie Guyader[3], and Carole Peyrin[1,2*]

[1] Univ. Grenoble Alpes, LPNC, F-38040 Grenoble
[2] CNRS, LPNC, F-38040 Grenoble
[3] Univ. Grenoble Alpes, GIPSA-lab, F-38402 Grenoble

Abstract

Visual analysis follows a default processing sequence which is predominantly coarse-to-fine. Low spatial frequencies (LSF) are processed more rapidly than high spatial frequencies (HSF), allowing an initial coarse parsing of visual input, prior to analysis of finer information. Our study investigated the influence of spatial frequency processing order on rapid scene categorization. We also examined the effects of the accumulation mode of spatial frequency information throughout processing, and of differences in luminance contrast between spatial frequencies. In Experiment 1, we used sequences composed of six filtered scenes, assembled from LSF to HSF (coarse-to-fine) or from HSF to LSF (fine-to-coarse) to test the effects of spatial frequency order. Spatial frequencies were either successive or additive within sequences to test the effects of spatial frequency accumulation mode. Results showed that participants categorized coarse-to-fine sequences more rapidly than fine-to-coarse sequences, irrespective of spatial frequency accumulation in the sequences. In Experiment 2, we investigated the extent to which differences in luminance contrast rather than in spatial frequency account for the advantage of coarse-to-fine over fine-to-coarse processing. Results showed that both spatial frequencies and luminance contrast account for a predominant coarse-to-fine processing, but that the coarse-to-fine advantage stems mainly from differences in spatial frequencies. Our study cautions against the use of contrast normalization in studies investigating spatial frequency processing. We argue that this type of experimental manipulation can impair the intrinsic properties of a visual stimulus. As the visual system relies on these to enable recognition, bias may be induced in strategies of visual analysis.

Keywords: coarse-to-fine; fine-to-coarse; low spatial frequency; high spatial frequency; root mean square contrast normalization.

1. Introduction

The human visual system is constantly involved in the perception and categorization of complex stimuli such as natural scenes. In the spatial domain, scenes are classically described in terms of pixel luminance. In the Fourier domain, a dual representation of a scene is created using the amplitude and phase spectra (Field, 1987; Ginsburg, 1986; Hughes, Nozawa, & Kitterle, 1996; Tolhurst, Tadmor, & Chao, 1992). The amplitude spectrum corresponds to the distribution of luminance contrast across spatial frequencies and orientations, and the phase spectrum corresponds to the spatial relation between spatial frequencies. Luminance contrast refers to the magnitude of luminance variation in a stimulus relative to its mean luminance (Robert Shapley & Enroth-Cugell, 1984). The visual system uses low-level features, such as spatial frequencies and luminance contrast to enable recognition, and from a neurobiological point of view, we now know that cells in the primary visual cortex respond to luminance contrast, spatial frequencies and orientations (Boynton, 2005; De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Poggio, 1972; Shapley & Lam, 1993). Many studies have also highlighted the importance of the amplitude spectrum in scene categorization (Guyader et al., 2004; Oliva & Torralba, 2001; Torralba & Oliva, 2003). Overall, these studies support current influential models of scene perception (Bar et al., 2006; Bar, 2003; Bullier, 2001; Hegdé, 2008; Kauffmann, Ramanoël, & Peyrin, 2014; Peyrin et al., 2010; Schyns & Oliva, 1994). According to these models, visual analysis is based on the parallel extraction of different attributes at different spatial frequencies in scenes, and follows a predominant and default coarse-to-fine processing sequence. Low spatial frequencies (LSF), containing the coarse information on a visual stimulus, are rapidly conveyed by magnocellular pathways to the occipital cortex and then access high-order cortical areas, in order to activate plausible interpretations of the visual input. This initial coarse analysis is then used to guide the later processing of high spatial frequencies (HSF) which are conveyed more slowly by parvocellular pathways and provide finer information.

Several behavioral studies, and a few neuroimaging studies, have investigated coarse-to-fine processing during scene perception (De Cesarei & Loftus, 2011; Musel et al., 2014, 2012; Parker et al., 1992; Peyrin et al., 2010; Schyns & Oliva, 1994). Evidence of a predominant coarse-to-fine processing was originally provided by psychophysical studies using hybrid scenes (Schyns & Oliva, 1994). Hybrid stimuli are made up by superimposing two images of scenes that contain different spatial frequencies and different semantic information (e.g., a highway scene in LSF superimposed on a city scene in HSF). When presentation times were very short (30 ms), perception of these hybrid scenes was dominated by LSF information. When presentation times were longer (150 ms), perception of hybrids was dominated by HSF information. This suggests that LSF take precedence over HSF in the visual time-course. More recent studies (Musel et al., 2014, 2012; Peyrin et al., 2005, 2010) explicitly simulated different time courses of spatial frequency processing during scene categorization using sequences of scene images in which the spatial frequency content differed from one image to the other, going either from low-to-high spatial frequencies (coarse-to-fine processing) or from high-to-low spatial frequencies (fine-to-coarse processing). These studies showed that sequences depicting coarse-to-fine processing were categorized more rapidly than those depicting fine-to-coarse processing. This suggests that the presentation order of spatial frequencies strongly influences the speed of scene categorization, and when LSF are presented first in the sequence, this may particularly facilitate the process. Recent event-related brain potential (ERP) studies have suggested that the accumulation of spatial frequency

information could also influence the perception of scenes, irrespective of the presentation order of spatial frequencies. De Cesarei and colleagues (De Cesarei & Codispoti, 2011; De Cesarei et al., 2013) investigated how spatial frequencies influence the identification of neutral and emotional scenes. They also used sequences containing images of scenes in which the amount of spatial frequency information increased progressively from one image to the next. To be precise, they presented sequences in which the first scene was either LSF or HSF, and the entire scene was gradually revealed by progressively adding either HSF or LSF information. This procedure allowed the authors to investigate the effects of the addition of spatial frequency information, according to the type of spatial frequency content (either LSF or HSF) which had been initially processed. Behavioral results showed that the identification rate of scenes increased as spatial frequency information was added, irrespective of the spatial frequency content initially presented in the sequence (LSF or HSF), and no differences in behavioral performances were observed between sequences starting with LSF and HSF information. These results suggested that scene identification did not critically depend on the initial processing of LSF – it appeared to rely more on the addition of spatial frequency information. However, divergences between the above mentioned studies may result from methodological differences in the accumulation mode of spatial frequency in the sequence of scene images. The accumulation mode refers here to how spatial frequency information is received as an input by the visual system, throughout the sequences (e.g., in the previously cited examples, either successively or additively).They also raised the question of whether the visual system would benefit from the reinjection of spatial frequency information relating to previous inputs during the processing of spatial frequency sequences. To our knowledge, no study has as yet directly investigated how the accumulation mode of spatial frequency in the sequence influences the well-established advantage of LSF over HSF during rapid scene categorization.

The first experiment in the present study aimed to investigate rapid scene categorization depending on both the presentation order of spatial frequencies (LSF before HSF or HSF before LSF) and the accumulation mode of spatial frequencies in the sequence (successive or additive presentation of different spatial frequency bands). In order to do this, we used dynamic scenes composed of six images of a scene filtered in different spatial frequencies, going from LSF to HSF or from HSF to LSF. This allowed us to test the effects of spatial frequency order. Dynamic scenes also depicted either a successive or additive processing. This allowed us to test the effects of spatial frequency accumulation mode. Successive sequences therefore started with either an LSF or an HSF filtered scene. They subsequently shifted to a higher or a lower spatial frequency band, and ended with an HSF or LSF filtered scene, respectively. Additive sequences started with an LSF or an HSF filtered scene, but this time, either HSF or LSF information was added. The spatial frequency content therefore increased from one image to the next, and sequences ended with an almost intact (or non-filtered) scene. Participants were asked to perform a categorization task on these stimuli (indoor vs. outdoor). Based on recent models of visual perception, we expected to observe a coarse-to-fine advantage, i.e. more rapid categorization when LSF (rather than HSF) were presented first. However, if the addition of spatial frequencies throughout the sequence constitutes an advantage irrespective of the order of presentation of spatial frequency, additive sequences should be categorized more rapidly than successive sequences in sequences which begin with either LSF or HSF information. Interaction between the accumulation mode of spatial frequencies and the order of spatial frequencies during sequences ought to reduce the coarse-to-fine

advantage in additive sequences (compared to successive sequences).

In the same context, differences in luminance contrast have been shown to exert a strong influence on speed of visual processing. For example, reaction times decrease as luminance contrast increases (Harwerth & Levi, 1978). The luminance contrast in scenes decreases as spatial frequency increases, following a $1/f^{\alpha}$ function (Field, 1987). LSF are characterized by a high luminance contrast, and HSF are characterized by a lower luminance contrast. The temporal precedence of LSF over HSF (i.e. coarse-to-fine processing) during scene categorization could therefore be explained by differences in contrast rather than in spatial frequency content. In order to avoid any confusion between the influence of spatial frequency content and that of luminance contrast in scene perception, recent studies equalize both the mean luminance and the luminance contrast of the filtered stimuli used (see, for example, Goffaux et al., 2011; Mu & Li, 2013; Vlamings, Goffaux, & Kemner, 2009) by attributing the same mean luminance and the same root mean square (RMS) contrast to all filtered stimuli. The RMS contrast corresponds to the standard deviation of luminance values and has been shown to be the most reliable indicator of the visibility of broadband filtered images (Bex & Makous, 2002). However, the specific role of luminance contrast in the spatial frequency processing of scenes has never been systematically investigated. In a second experiment, we investigated the specific role of luminance contrast, spatial frequencies, and their interaction during the coarse-to-fine processing of scenes. In order to do so, we used dynamic scenes adapted from Experiment 1, and manipulated the spatial frequency content and luminance contrast of the images composing the sequences.

2. Experiment 1

The first experiment aimed to determine whether scene categorization relies more on the processing of LSF prior to that of HSF, or whether it is favored more by the addition of information from different spatial frequencies rather than by the successive and separate processing of different spatial frequency bands. We used scene sequences composed of six filtered versions of a scene, as stimuli. First of all, we manipulated the order of spatial frequencies in our sequences. In the coarse-to-fine condition, filtered scenes were assembled following an LSF to HSF order. In the fine-to-coarse condition, filtered scenes were assembled following a reverse HSF to LSF order. Next, we manipulated the accumulation mode of spatial frequencies in the sequences. In the successive condition, filtered scenes which made up the sequences were built from band-pass filters centered on different spatial frequencies. Each frame therefore contained different spatial frequency information. In the additive condition, sequences began with the same frame as in the successive condition, but spatial frequency information was progressively added to the previous filtered scene of the sequence. Each frame therefore contained an increasing amount of spatial frequency information (while retaining the prior information) until an almost intact scene was presented. If scene categorization is mainly favored by the initial processing of LSF information, we expected to observe a more rapid categorization of coarse-to-fine sequences than of fine-to-coarse sequences, irrespective of the accumulation mode of spatial frequency information (successive or additive). Moreover, if the addition of spatial frequency information is more advantageous for scene categorization than the successive processing of differential spatial frequency band information, we expected additive sequences to be categorized more rapidly than successive sequences in all of the spatial frequency order conditions. Finally, if the addition of different spatial frequency bands is more advantageous irrespective of presentation order, we expected to observe an interaction between the accumulation mode of spatial frequencies and their order, and anticipated that this would be expressed

in a reduction of the coarse-to-fine advantage in additive sequences (compared to successive sequences).

2.1. Materials and Methods

2.1.1. Participants

Thirty right-handed participants (4 males; Mean age ± SD: 23 ± 4 year) with normal or corrected-to-normal vision and no history of neurological disorder were included in the experiment. All participants gave their informed written consent before participating in the study which was in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.1.2. Stimuli and Procedure

Stimuli were constructed from 40 black and white photographs (256-level grey-scales, 1024 x 768 pixels) of natural scenes, belonging to two categories (20 indoor scenes and 20 outdoor scenes) with a visual angle of 24 × 18 degrees[1]. Exemplars from the two categories (outdoor and indoor) were chosen in order to ensure that dominant orientations in the mean amplitude spectrum were similar and to avoid categorization based on this type of visual cue. Furthermore, outdoor and indoor categories were equivalent in terms of visual cluttering (Subband Entropy measures; see Rosenholtz, Li, & Nakano; 2007). Mean sub-band entropy did not differ between outdoors and indoors (2.95 ± 0.16 and 2.95 ± 0.14, respectively; $F_{1,38} < 1$). Stimuli were elaborated using the MATLAB image processing toolbox (Mathworks Inc., Sherborn, MA, USA).

In our Successive sequences, each scene was filtered with six band-pass filters with central frequencies corresponding to 24, 34, 49, 71, 101 and 144 cycles per image and a standard deviation of 25.6 cycles per image (i.e., 1, 1.4, 2, 2.9, 4.2 and 6 cycles per degree and a standard deviation of 1.07 cycles per degree; see Figure 1c). The central frequencies of filters were not linearly spaced - they followed a logarithmic scale[2]. This enabled us to obtain a better sampling of the amplitude spectrum of natural scenes, in which luminance contrast decreases as spatial frequency increases ("1/f" shape; Field, 1987) and a better sampling of low spatial frequencies with more filters centered on low spatial frequencies (for a similar approach, see Willenbockel et al., 2010). The information contained in frequencies below 7.2 cycles per image (or 0.3 cycles per degree) was removed. Filtered scenes were then normalized to obtain a mean luminance of 0.5 (for luminance values of between 0 and 1; i.e. mean luminance of 128 on a 256 grey-level scale).

For the Additive sequences, the six filters used in the Successive condition were iteratively added from LSF to HSF and from HSF to LSF. For sequences starting with LSF, each scene was filtered with six filters. The first filter was centred on 24 cycles per image. The second filter corresponded to the addition of two filters, one centred on 24 cycles per image and the other one on 34 cycles per image (see Figure 1c). The third filter corresponded to the addition of three filters centred on 24, 34, and 49 cycles per images. The fourth filter corresponded to the addition of four filters centred on 24, 34, 49, and 71 cycles per image. The fifth filter corresponded to the addition of five filters centred on 24, 34, 49, 71, and 101 cycles per image. The sixth filter corresponded to the addition of six filters centred on 24, 34, 49, 71, 101, and 144 cycles per image. For sequences starting with HSF, the filtering order was reversed. Each filtering level therefore corresponded to an increasing truncation of the frequency spectrum of images along high or low spatial frequencies. This additive procedure used to build filters was preferred to a procedure based on increasing width, in order to maintain previously-received information following the addition of spatial frequencies. Since the addition of band-pass filters resulted in a filter with a larger band-pass, the resulting filter was normalized to a maximum value of 1[3]. Filtered scenes were then normalized to obtain a mean luminance

of 0.5 (for luminance values of between 0 and 1; i.e. mean luminance of 128 on a 256 grey-level scale).

The resulting filtered scenes in both the Successive and Additive conditions were then assembled in order to create short movies, from LSF to HSF (CtF sequence) or from HSF to LSF (FtC sequence, see Figure 1). Each movie lasted 150 ms and was composed of six filtered versions of the same scene (presented for 25 ms each). These dynamic sequences (for both Successive and Additive condition) allowed us to imitate and impose either a coarse-to-fine (CtF) or a fine-to-coarse (FtC) processing of scenes. It should be noted that the first LSF image in CtF sequences was the same in the Successive and Additive conditions. Similarly, the first HSF image in FtC sequences was the same in the Successive and Additive conditions[4].

Stimuli were displayed using E-Prime software (E-Prime Psychology Software Tools Inc., Pittsburg, PA) on a 17' monitor, with a resolution of 1024 × 768 pixels, at a refreshing rate of 80 Hz and with a viewing distance of 70 cm. In order to respect the distance and the central position, participants' heads were supported by a chinrest. Each participant performed two sessions according to the two accumulation conditions (Successive and Additive), which consisted of 80 trials each. The order of the sessions was counterbalanced between participants. Within each session, the sequences (CtF and FtC) were displayed randomly. Each trial began with a central fixation point that was presented for 500 ms (in order to control the gaze direction to the center of the screen), immediately followed by the dynamic sequence lasting 150 ms and a backward mask built with 1/f noise and presented during 30 ms to prevent retinal persistence. Participants were asked to categorize the dynamic scenes as fast and as accurately as possible by pressing a button with their forefinger or their middle finger according to the category of the scene. Response buttons were counterbalanced across participants Response accuracy and response times (in milliseconds) were recorded. The experiment lasted about 15 minutes.

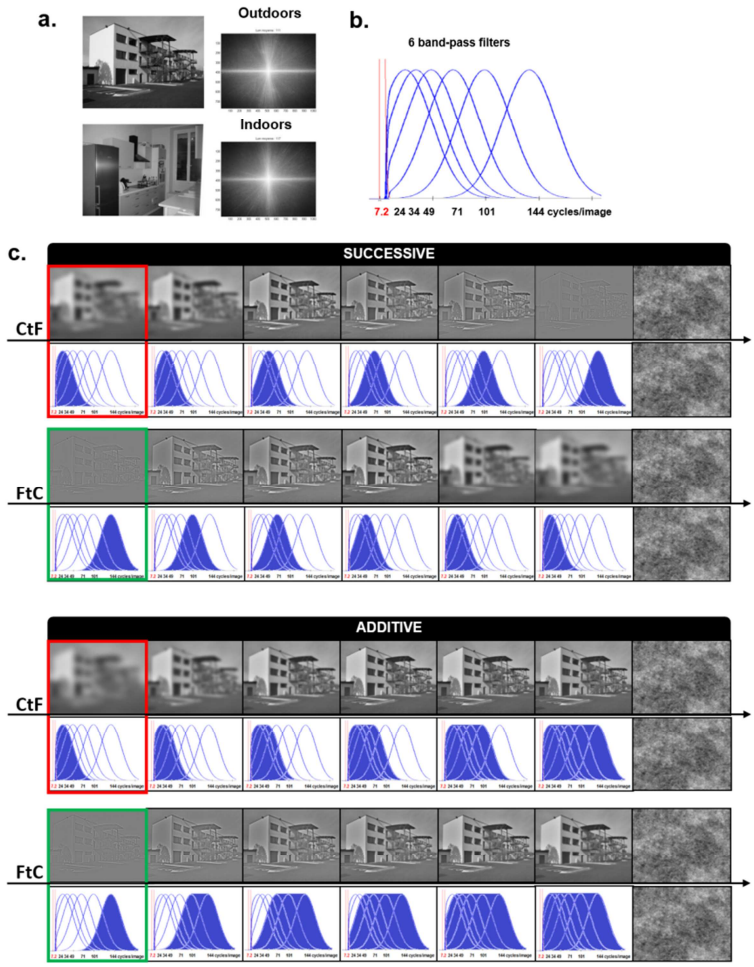


Figure 1: (a) Example of scenes from different categories (outdoors and indoors). Mean amplitude spectra of each category. On each amplitude spectrum, low spatial frequencies are close to the center, and high spatial frequencies are on the periphery. Vertical orientations are represented on the x-axis and horizontal orientations on the y-axis. (b) Six band-pass filters used to filter the scenes. (c) Example of six spatial frequency filtered images in the Successive (top) and the Additive (bottom) conditions depicting a coarse-to-fine (CtF) and fine-to-coarse (FtC) sequence, with the corresponding band-pass filter underneath. On the band-pass filter, we illustrate how spatial frequency information progresses in either a successive or additive manner. For each Order condition (CtF and FtC), successive and additive sequences started with exactly the same frame. It should be noted that the perception of spatial frequencies could be affected by the reduction in picture size of scenes for the illustrative purposes. Picture size is approximately 10 times smaller than the actual picture size on the screen used in the experiment. Central spatial frequencies used to filter the scenes were therefore reduced for illustrative purposes.

2.2. Results

Two 2 × 2 × 2 ANOVAs with Order of spatial frequencies (CtF and FtC), Accumulation mode of spatial frequencies (Successive and Additive), and Category of scenes (outdoor and indoor) as within subject factors were performed on mean error rates (mER) and mean correct response times (mRT, in milliseconds) using Statistica 10.0 software (Statsoft, Tulsa, USA). The significance threshold was set at $\alpha = 0.05$. To reduce the effect of extreme values, RT for each subject's correct response in each condition was trimmed by removing responses inferior and superior to three times the interquartile interval. This resulted in removing an average of 0.92% of the trials.

Whatever the experimental condition, mER was low (under 10 %). The ANOVA performed on mER revealed no main effect of Order (Mean ± SD; CtF: 7.25 ± 5.50 %; FtC: 6.79 ± 4.77 %, $F_{1, 29} = 0.42$, $p = 0.522$,

partial eta square – η_p^2 = 0.014), Accumulation (Successive: 6.59 ± 4.99 %; Additive: 7.46 ± 5.24 %; $F_{1, 29}$ = 1.69, p = 0.204, η_p^2 = 0.055), and Category (outdoor: 6.92 ± 4.96 %; indoor: 7.13 ± 5.36 %, $F_{1, 29}$ = 0.08, p = 0.775, η_p^2 = 0.003). There was no interaction between Accumulation and Order ($F_{1, 29}$ = 0.07, p = 0.797, η_p^2 = 0.002), between Accumulation and Category ($F_{1, 29}$ = 1.70, p = 0.203, η_p^2 = 0.055), or Order and Category ($F_{1, 29}$ = 0.25, p = 0.623, η_p^2 = 0.008). However, there was a significant interaction between these three factors ($F_{1, 29}$ = 7.59, p = 0.010, η_p^2 = 0.207), due to a significant Order × Category interaction for Successive scenes only (Successive: $F_{1, 29}$ = 5.93, p = 0.021, η_p^2 = 0.170; Additive: $F_{1, 29}$ = 2.04, p = 0.164, η_p^2 = 0.066). Post-hoc comparisons (Tukey's HSD) for the Successive condition revealed that participants made more errors when categorizing indoor (8.00 ± 7.02 %) than outdoor scenes (4.50 ± 5.47 %) for FtC sequences (p = 0.020), whereas there was no difference between indoor (6.33 ± 6.94 %) and outdoor (7.50 ± 8.17 %) categorizations for CtF sequences (p = 0.893).

The ANOVA performed on mRT revealed no main effect of the Accumulation (Successive: 565 ± 61 ms; Additive: 561 ± 69 ms; $F_{1, 29}$ = 0.50, p = 0.486, η_p^2 = 0.017) and no main effect of Category (Mean ± SD; outdoor: 563 ± 65 ms; indoor: 563 ± 64 ms, $F_{1, 29}$ = 0.02, p = 0.892, η_p^2 = 0.001) but there was a significant main effect of Order ($F_{1, 29}$ = 37.39, p < 0.001, η_p^2 = 0.563). CtF sequences (548 ± 68 ms) were categorized faster than FtC sequences (578 ± 62 ms), irrespective of the Accumulation of spatial frequencies in the sequence. The expected interaction between Accumulation and Order was not significant ($F_{1, 29}$ = 2.13, p = 0.155, η_p^2 = 0.068). However, with respect to our hypothesis of a faster categorization when LSF are presented first, we compared CtF and FtC categorization for each Accumulation condition. Planned comparisons revealed that CtF sequences were categorized faster than FtC sequences for both Successive (CtF: 555 ± 63 ms; FtC: 575 ± 63 ms; $F_{1, 29}$ = 7.21, p = 0.012, η_p^2 = 0.200) and Additive sequence (CtF: 541 ± 78 ms; FtC: 581 ± 67 ms; $F_{1, 29}$ = 18.77, p < 0.001, η_p^2 = 0.393). It should be noted that there was no difference between Successive and Additive scenes for CtF (Successive: 555 ± 63 ms; Additive: 541 ± 78 ms; $F_{1, 29}$ = 2.45, p = 0.128, η_p^2 = 0.078) or FtC categorization (Successive: 575 ± 63 ms; Additive: 581 ± 67 ms; $F_{1, 29}$ = 0.47, p = 0.498, η_p^2 = 0.016). There was no interaction between either of the other manipulated factors (Accumulation × Category: $F_{1, 29}$ = 0.02, p = 0.896, η_p^2 = 0.001; Order × Category: $F_{1, 29}$ = 1.15, p = 0.293, η_p^2 = 0.038; Accumulation × Order × Category: $F_{1, 29}$ = 2.18, p = 0.151, η_p^2 = 0.070).

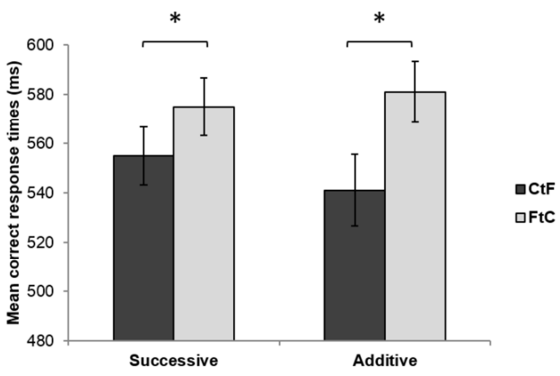


Figure 2: Mean correct reaction times (in milliseconds) to categorize coarse-to-fine (CtF) and fine-to-coarse (FtC) sequences in the Successive and Additive conditions. Error bars correspond to standard errors. * indicates significant differences.

2.3. Discussion

In Experiment 1, our aim was to investigate rapid scene categorization depending on the sequence order of spatial frequencies (LSF before HSF or HSF before LSF) and on the accumulation mode of spatial frequency throughout the sequence. More specifically, we tried to find out whether the successive or the additive presentation of spatial frequency information is more advantageous for coarse-to-fine and fine-to coarse scene categorization. Results showed that coarse-to-fine sequences were categorized more rapidly than fine-to-coarse sequences, irrespective of the accumulation mode of spatial frequency information. This

suggests that the initial processing of LSF (compared to HSF) information is more advantageous in rapid scene categorization. This result is consistent with some behavioral data (Musel et al., 2012; 2014; Parker et al., 1992; Schyns & Oliva, 1994) and therefore provides a new experimental argument for coarse-to-fine processing as a useful default strategy in scene categorization. However, contrary to what might have been expected (De Cesarei & Codispoti, 2011; De Cesarei et al., 2013), the additive presentation of spatial frequency information during coarse-to-fine processing did not induce a significant increase in speed of categorization. This result suggests that the addition of spatial frequency information – and thus the reinjection of the previously-received spatial frequency information over sequence processing – is not more advantageous than the successive processing of information from different spatial frequency bands (going from LSF to HSF) in the categorization of scenes.

Overall, the results from Experiment 1 suggest that the role of the presentation order of spatial frequencies (i.e. LSF before HSF) is more pronounced than that of the accumulation of spatial frequencies in rapid scene categorization. Greater speed of scene categorization when LSF presentation precedes that of HSF could however be explained by differences in luminance contrast rather than by differences in spatial frequency content. In Experiment 2, we therefore looked more closely at the specific role of luminance contrast, spatial frequencies, and their interaction during coarse-to-fine processing of scenes.

3. Experiment 2

Experiment 2 aimed to further explore the extent to which differences in spatial frequencies and luminance contrast might explain the advantage of coarse-to-fine over fine-to-coarse processing. Our stimuli were dynamic scenes, composed of six filtered versions of a scene, assembled from LSF to HSF (coarse-to-fine processing) or from HSF to LSF (fine-to-coarse processing). In the first experimental condition (SF-CON), dynamic scenes corresponded to those used in the Successive condition in Experiment 1. The movie stimuli were composed of 6 images, each with a different spatial frequency content and contrast value, depicting two different sequences: a low-to-high spatial frequency with high-to-low contrast sequence (CtF) and a high-to-low spatial frequency with low-to-high contrast sequence (FtC). This condition allowed us to test the effects of the interaction between spatial frequencies and contrast during scene categorization. The second experimental condition (SF) allowed us to investigate how differences in spatial frequencies might explain the coarse-to-fine advantage irrespective of luminance contrast differences. The movies were composed of 6 images, each with a different spatial frequency content. Contrast was however identical in all images. The movies depicted low-to-high spatial frequency (CtF) and high-to-low spatial frequency (FtC) sequences. The third experimental condition (CON) allowed us to investigate how differences in luminance contrast might explain the coarse-to-fine advantage irrespective of spatial frequency differences. The movie stimuli were composed of 6 non-filtered images, and in each image contrast was modified in order to correspond to the values of the CtF or FtC sequences used in the SF-CON condition. In this way, we obtained movies depicting high-to-low contrast (CtF) and low-to-high contrast (FtC) sequences. As in Experiment 1, participants performed a categorization task (indoor vs. outdoor) on these stimuli. It should be noted that the fine-to-coarse sequences in the additive condition tested in Experiment 1 had lower overall luminance contrast values than the coarse-to-fine sequences (mean RMS contrast of fine-to-coarse sequences : 0.07; mean RMS contrast of coarse-to-fine sequences : 0.12). We did not therefore use additive sequences when specifically investigating the role of contrast in Experiment 2. We compared sequences in

which the same images were displayed but in reverse order.

If the advantage of coarse-to-fine over fine-to-coarse categorization is mainly explained by differences in spatial frequencies, we expected CtF sequences to be categorized more rapidly than FtC sequences in the SF-CON and SF conditions only. In this case, we should observe a significant interaction between SF-CON/CON and CtF/FtC conditions. Conversely, if the advantage of coarse-to-fine processing is mainly due to differences in luminance contrast rather than in spatial frequency content, we expected CtF sequences to be categorized more rapidly than FtC sequences only in the SF-CON and the CON conditions. We should observe a significant interaction between SF-CON/SF and CtF/FtC conditions.

3.1. Material and Methods

3.1.1. Participants

Thirty-eight right-handed participants (8 males Mean age ± SD: 22 ± 5 year) with normal or corrected-to-normal vision and no history of neurological disorder were included in the experiment. Participants did not participate to Experiment 1. All participants gave their informed written consent before participating in the study which was in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

3.1.2. Stimuli and Procedure

Stimuli were built from the same 40 black and white photographs of indoor and outdoor scenes used in Experiment 1. In the SF-CON condition, we used exactly the same stimuli as in the Successive condition of Experiment 1 (see Section *2.1.2. Stimuli and Procedure* of Experiment 1). Filtered scenes were then normalized to obtain a mean luminance of 0.5 (for luminance values of between 0 and 1; i.e. mean luminance of 128 on a 256 grey-level scale). Luminance contrast was not modified. This resulted in six images of each scene with different spatial frequency content associated with a particular RMS contrast value (Table 1).

In the SF condition, each scene was filtered with the same six band-pass filters used in the SF-CON condition. This time, images were then normalized to obtain a mean luminance of 0.5 (for luminance values of between 0 and 1; i.e. mean luminance of 128 on a grey-level scale) with a standard deviation of 0.1 (i.e., 25.6 on a grey-level scale; root mean square [RMS] contrast). This resulted in six images of each scene with different spatial frequency content but identical RMS contrast value (Table 1).

In the CON condition, the spatial frequency content of scenes was not manipulated. The luminance contrast of each scene was normalized using six standard deviations of luminance values corresponding to those obtained after scenes were filtered using the six band-pass filters from the SF-CON condition (Table 1). This resulted in six images of each scene with identical spatial frequency content and different RMS contrast values.

The resulting filtered scenes in each Contrast condition were then assembled in order to create short movies (Figure 3), from low-to-high spatial frequencies with high-to-low contrast (CtF sequences) or from high-to-low spatial frequencies with low-to-high contrast (FtC sequences) in the SF-CON condition; from low-to-high spatial frequencies (CtF sequences) and from high-to-low spatial frequencies (FtC sequences) in the SF condition; and from high-to-low contrast (CtF sequences) and from low-to-high contrast (FtC sequences) in the CON condition. Each movie lasted 150 ms and was composed of six filtered versions of the same scene (presented for 25 ms each).

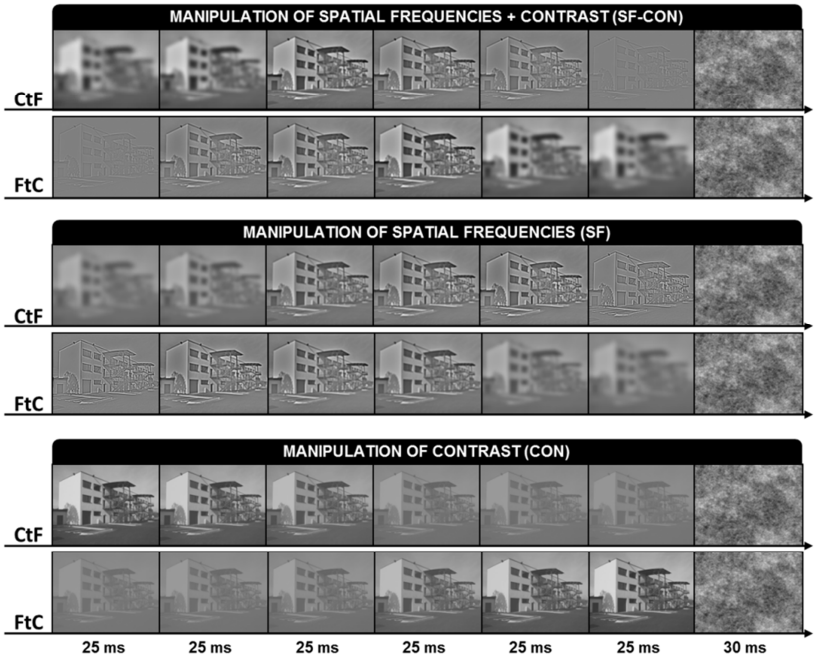


Figure 3: Example of six images which make up the movies and depicting a coarse-to-fine (CtF) and a fine-to coarse (FtC) sequence. In the SF-CON condition, both spatial frequencies and luminance contrast are manipulated. In the **SF** condition, only spatial frequencies are manipulated. In the CON condition, only luminance contrast is manipulated. It should be noted that perception of spatial frequencies could be affected by the reduction in picture size of scenes for the purposes of illustration. Picture size is approximately 10 times smaller than the actual picture size on the screen used in the experiment. Central spatial frequencies used to filter the scenes were therefore reduced for illustrative purposes.

Table 1: Mean and standard deviation (± SD) of RMS contrast of the six images composing the sequence for each contrast condition (SF-CON, SF, and CON) and each sequence order (CtF and FtC) **in Experiment 2**, for luminance values between 0 and 1. The mean luminance value of all images was equal to 0.5.

		Image 1		**Image 2**		**Image 3**		**Image 4**		**Image 5**		**Image 6**	
Condition		Mean	±*SD*	Mean	±*SD*	Mean	±*SD*	Mean	±*SD*	Mean	±*SD*	Mean	±*SD*
SF-CON	CTF	0.13	*0.01*	0.10	*0.01*	0.07	*0.01*	0.05	*0.01*	0.05	*0.01*	0.04	*0.01*
	FTC	0.04	*0.01*	0.05	*0.01*	0.05	*0.01*	0.07	*0.01*	0.10	*0.01*	0.13	*0.01*
SF	CTF	0.10	*0*	0.10	*0*	0.10	*0*	0.10	*0*	0.10	*0*	0.10	*0*
	FTC	0.10	*0*	0.10	*0*	0.10	*0*	0.10	*0*	0.10	*0*	0.10	*0*
CON	CTF	0.13	*0.01*	0.10	*0.01*	0.07	*0.01*	0.05	*0.01*	0.05	*0.01*	0.04	*0.01*
	FTC	0.04	*0.01*	0.05	*0.01*	0.05	*0.01*	0.07	*0.01*	0.10	*0.01*	0.13	*0.01*

Each participant performed three sessions according to the three Contrast conditions (SF-CON, SF, and CON), which consisted of 80 trials each. The order of the sessions was counterbalanced between participants. Within each session, the sequences (CtF and FtC) were displayed randomly. The procedure and task were the same as in Experiment 1. Response accuracy and response times (in milliseconds) were recorded. The experiment lasted about 20 minutes.

3.2. Results

Two 3 × 2 × 2 ANOVAs with Contrast condition (SF-CON, SF, and CON), Order (CtF and FtC), and Category (outdoor and indoor) as within subject factors were

performed on mean error rates (mER) and mean correct response times (mRT, in milliseconds). To reduce the effect of extreme values, RT for each subject's correct response in each condition was trimmed by removing responses inferior and superior to three times the interquartile interval. This resulted in removing an average of 0.67% of the trials.

Whatever the experimental condition, mER was low (under 5 %). The ANOVA conducted on mER revealed no main effect of Contrast (Mean ± SD; SF-CON: 5.03 ± 4.75 %; SF: 5.00 ± 4.21 %; CON: 4.57 ± 4.73 %; $F_{2, 74} = 0.42$, $p = 0.656$; $\eta_p^2 = 0.011$), Order (CtF: 5.35 ± 4.98 %; FtC: 4.39 ± 3.84 %; $F_{1, 37} = 3.05$; $p = 0.089$; $\eta_p^2 = 0.076$), and Category (outdoor: 5.09 ± 4.93 %; indoor: 4.65 ± 3.82 %; $F_{1, 37} = 0.71$, $p = 0.404$; $\eta_p^2 = 0.019$). There was no significant interaction between Contrast and Order $F_{2, 74} = 0.60$, $p = 0.551$, $\eta_p^2 = 0.016$) or between SF-CON/CON and CtF/FtC ($F_{1, 37} = 0.83$, $p = 0.367$, $\eta_p^2 = 0.022$) or SF-CON/SF and CtF/FtC conditions ($F_{1, 37} = 0.04$, $p = 0.842$, $\eta_p^2 = 0.001$). There was a significant interaction between Contrast and Category ($F_{2, 74} = 3.16$, $p = 0.048$; $\eta_p^2 = 0.079$). However, post-hoc comparisons (Tukey's HSD) revealed no difference between any conditions (all ps > 0.05). All other interactions were not significant (Order × Category: $F_{1, 37} = 0.01$, $p = 0.998$, $\eta_p^2 = 0.000$; Contrast × Order × Category: $F_{2, 74} = 0.65$, $p = 0.526$, $\eta_p^2 = 0.017$).

The ANOVA conducted on mRT revealed a main effect of Contrast ($F_{2, 74} = 7.26$, $p = 0.001$, $\eta_p^2 = 0.164$). Sequences were categorized faster in the CON condition (553 ± 70 ms) than in the SF-CON (574 ± 70 ms) and SF conditions (565 ± 70 ms; SF-CON vs. CON: $F_{1, 37} = 10.38$, $p = 0.003$, $\eta_p^2 = 0.219$; SF vs. CON: $F_{1, 37} = 6.35$, $p = 0.016$, $\eta_p^2 = 0.147$; SF-CON vs. SF: $F_{1, 37} = 3.17$, $p = 0.083$). There was a main effect of Order ($F_{1, 37} = 66.21$, $p < 0.001$, $\eta_p^2 = 0.641$). CtF sequences were categorized faster than FtC sequences (556 ± 67 ms and 573 ± 67 ms, respectively). There was no interaction between Contrast and Order ($F_{2, 74} = 2.13$, $p = 0.126$, $\eta_p^2 = 0.054$). However, with respect to our hypothesis of the advantage of coarse-to-fine over fine-to-coarse processing, we compared CtF and FtC categorizations for each Contrast condition. Planned comparisons revealed that CtF sequences were categorized faster than FtC sequences in the FS-CON condition (CtF: 563 ± 73 ms; FtC: 586 ± 71 ms; $F_{1, 37} = 25.01$, $p < 0.001$, $\eta_p^2 = 0.403$), in the SF condition (CtF: 555 ± 74 ms; FtC: 574 ± 68 ms; $F_{1, 37} = 15.24$, $p < 0.001$, $\eta_p^2 = 0.292$) and CON condition (CtF: 549 ± 71 ms; FtC: 558 ± 71 ms; $F_{1, 37} = 4.27$, $p = 0.046$, $\eta_p^2 = 0.103$). Furthermore, the interaction between SF-CON/CON and CtF/FtC conditions was significant ($F_{1, 37} = 4.61$, $p = 0.038$, $\eta_p^2 = 0.111$), suggesting that the coarse-to-fine advantage was reduced in the CON relative to the SF-CON condition. On the other hand, the interaction between SF-CON/SF and CtF/FtC conditions was not significant ($F_{1, 37} = 0.33$, $p = 0.568$; $\eta_p^2 = 0.009$). Post hoc comparisons (Tukey's HSD) revealed that for CtF sequences, there was no difference between the SF-CON and CON conditions ($p = 0.084$), between the SF-CON and SF conditions ($p = 0.702$) or between the CON and SF conditions ($p = 0.799$). For FtC sequences, scenes were categorized more rapidly in the CON than in the SF-CON ($p < 0.001$) and SF condition ($p = 0.018$), but there was no difference between the SF-CON and SF conditions ($p = 0.229$). There was no main effect of Category (outdoor: 563 ± 66 ms; indoor: 566 ± 71 ms; $F_{1, 37} = 0.41$, $p = 0.524$, $\eta_p^2 = 0.011$) or interaction between Category and the other manipulated factors (Contrast × Category: $F_{2, 74} = 0.35$, $p = 0.705$, $\eta_p^2 = 0.009$; Order × Category: $F_{1, 37} = 3.16$, $p = 0.084$, $\eta_p^2 = 0.078$; Contrast × Order × Category: $F_{2, 74} = 0.27$, $p = 0.768$, $\eta_p^2 = 0.007$).

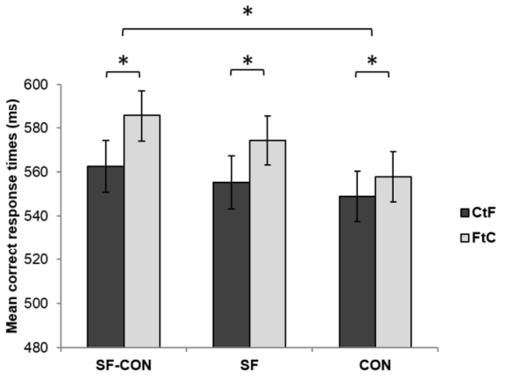


Figure 4: Mean correct reaction times (in milliseconds) to categorize coarse-to-fine (CtF) and fine-to-coarse (FtC) sequences in the SF-CON, SF, and CON conditions. Error bars correspond to standard errors. ***** indicates significant differences

3.3. Discussion

Experiment 2 aimed to determine whether differences in luminance contrast, in spatial frequency content, or both, account for the advantage of the coarse-to-fine over fine-to-coarse categorization observed in behavioral performances. The design used in Experiment 2 allowed us to investigate the coarse-to-fine advantage when both spatial frequency content and luminance contrast values were different (SF-CON condition) and to investigate the effects of spatial frequencies (SF condition) and luminance contrast (CON condition) separately. Results showed that in the SF-CON condition (i.e. when spatial frequency content and luminance contrast values differed between filtered scenes), coarse-to fine sequences were categorized more rapidly than fine-to-coarse sequences, thus replicating results of Experiment 1 (Successive condition). We also observed a coarse-to-fine advantage in the SF condition (i.e. when only spatial frequency differed) and the CON condition (i.e. when only luminance contrast value differed). However, the coarse-to-fine advantage was significantly reduced in the CON condition in comparison to the SF-CON condition. These results suggest that both spatial frequencies and luminance contrast account for a predominant coarse-to-fine (rather than fine-to-coarse) processing, but that the coarse-to-fine advantage observed in the SF-CON condition was mainly driven by differences in spatial frequencies.

4. General Discussion

In the present study, we investigated the influence of the presentation order of spatial frequencies, the accumulation mode of spatial frequency information over time, and differences in luminance contrast between spatial frequencies on rapid scene categorization. Two experiments were conducted to address these issues. The first experiment sought to determine whether rapid scene categorization was favored by the processing of LSF prior to HSF (coarse-to-fine processing), rather than HSF prior to LSF (fine-to-coarse processing), or by the addition of different spatial frequency bands, rather than successive and separate processing of different spatial frequency bands. Our results showed first of all that dynamic scenes depicting coarse-to-fine processing were categorized more rapidly than those depicting fine-to-coarse processing, irrespective of the type of accumulation of spatial frequencies (successive or additive). This result is consistent with a considerable number of behavioural and computational studies which have shown that processing of LSF precedes that of HSF using sine wave gratings and filtered scenes, and that categorization is facilitated when LSF are available before HSF (De Cesarei & Loftus, 2011; Hughes et al., 1996; Kihara & Takeda, 2010; Loftus & Harley, 2005; Mermillod et al., 2005; Musel et al., 2014, 2012; Parker et al., 1992; Schyns & Oliva, 1994). It is also consistent with neurophysiological and neuroimaging evidence (Bar et al., 2006; Bar, 2003; Bullier, 2001; Kveraga, Boshyan, & Bar, 2007; Peyrin et al., 2010) suggesting that LSF activate visual pathways before HSF in order to activate plausible interpretations of the visual input.

In Experiment 1, we tested whether coarse-to-fine processing of scene was favored more by the accumulation of spatial frequency information throughout the

sequence by adding spatial frequency bands (Additive condition). We compared this to the successive and separate presentation of information from different spatial frequency bands (Successive condition). We found that coarse-to-fine sequences were categorized more rapidly than fine-to-coarse sequences for both successive and additive sequences. In a recent study, Musel et al. (2012) used dynamic stimuli similar to our successive sequences, in order to impose a coarse-to-fine or a fine-to-coarse processing of scenes, in both young and elderly participants. Their results showed that young participants categorized coarse-to-fine sequences more rapidly than fine-to-coarse sequences. Our results therefore replicate their findings, and this suggests that these dynamic stimuli are highly suited to the investigation of coarse-to-fine processing. Other studies have used a protocol involving the progressive revelation of scenes, from LSF to HSF or from LSF to HSF, relatively similar to our additive sequences (De Cesarei & Codispoti, 2011; De Cesarei et al., 2013; see also Schettino, Loeys, Bossi, & Pourtois, 2012; Schettino, Loeys, Delplanque, & Pourtois, 2011; Schettino, Loeys, & Pourtois, 2013). In these studies, sequences started with an LSF or an HSF scene, and HSF or LSF information, respectively, was added progressively. A sufficiently long inter-image interval allowed participants – who were asked to decide if they were able to identify the gist of a scene - time to respond. The authors report that the identification rate of scenes increased as spatial frequency information was added, but no differences were observed between sequences starting with either LSF or HSF information. These results suggested that scene identification did not critically depend on the initial processing of LSF, but rather on the addition of spatial frequency information. This result differed from the more rapid reaction times observed in our Experiment 1 for additive sequences starting with LSF compared to those starting with HSF. Our results suggest that order (and the initial processing of LSF) is more predominant than the amount of spatial frequencies in rapid scene categorization. It should be noted that speed of categorization (e.g., response times) was not the central focus of previous studies, in which each revelation step (i.e. each filtered scene containing increasing spatial frequency content) was presented for 1000 ms and separated from the next scene by more than one second. Importantly, exposure duration of stimuli has been shown to influence spatial frequency processing (De Cesarei & Loftus, 2011; Schyns & Oliva, 1994). Very short presentation times favor LSF processing, and longer presentation times favor HSF processing. However when presentation duration goes beyond a certain length, it no longer influences the processing of spatial frequencies and the difference between LSF and HSF processing disappears. In this context, long exposure duration, of for example 1000 ms, may have masked the effects of spatial frequencies in gist identification. The shorter duration of each scene in our sequences (25 ms), and the absence of any inter-image interval, seems to be more appropriate in the investigation of the effects of spatial frequency order during rapid scene categorization[5].

To summarize, results from Experiment 1 indicate that the presentation order of spatial frequency information, rather than the accumulation of spatial frequency information, plays a predominant role in rapid scene categorization. However, since LSF and HSF are characterized by high and low luminance contrast, respectively, differences in contrast rather than in spatial frequencies may account for the advantage of coarse-to-fine over fine-to-coarse categorization. In Experiment 2, we investigated the extent to which spatial frequency order and luminance contrast differences account for the coarse-to-fine advantage. The design used in Experiment 2 allowed us to investigate the coarse-to-fine advantage when both spatial frequency content and luminance contrast values were different (SF-CON condition) and to investigate the effects of spatial frequencies (SF condition) and luminance contrast (CON

condition) separately. Results showed that coarse-to-fine sequences were categorized more rapidly than fine-to-coarse sequences in the SF-CON condition, when both spatial frequency content and luminance contrast values differed between the filtered scenes composing the sequences. A coarse-to-fine advantage was also observed in the SF condition (i.e. when only spatial frequency content differed) and the CON condition (i.e. when only luminance contrast values differed). However, the advantage of coarse-to-fine over fine-to-coarse processing was reduced in the CON condition in comparison to the SF-CON condition. These results therefore suggest that the advantage of a predominant coarse-to-fine processing stems from both spatial frequencies and luminance contrast processing, and to a greater extent from differences in spatial frequencies.

Equalization of luminance contrast in all filtered scenes (i.e. RMS contrast equalization) is a control method used in a growing number of studies. This procedure is thought to allow the examination of spatial frequency processing without any of the potentially confounding effects of luminance contrast differences between spatial frequencies. Several recent studies which used filtered images equalized in terms of luminance contrast as stimuli failed to observe strong effects of spatial frequencies on a behavioural level. For example, Vlamings et al. (2009) used LSF and HSF filtered faces as stimuli in an ERP study, in which contrast equalization between spatial frequencies was either present or absent. On a behavioural level, they found that LSF faces were categorized more rapidly than HSF faces. However, this observed difference was significantly reduced when contrast was equalized between LSF and HSF faces. Similarly, Goffaux et al. (2011) manipulated both the spatial frequency content of faces, and the exposure duration of stimuli. All filtered faces were equalized in terms of luminance contrast. Their behavioural results did not reveal any effect of spatial frequency or any interaction with exposure duration on participants' performance. These results are therefore consistent with the idea that luminance contrast plays an important role in highlighting differences in LSF and HSF processing. It also suggests that the use of luminance contrast equalization across filtered stimuli in studies investigating spatial frequency processing may be a relevant methodological aspect to consider when comparing studies since it appears to influence behavioural performances in tasks involving the processing of spatially frequency filtered stimuli. Although differences in luminance contrast in part account for the advantage of coarse-to-fine over fine-to-coarse processing, our results suggest that differences in spatial frequencies play a more predominant role, therefore cautioning the use of contrast normalization.

Indeed, luminance contrast equalization between spatial frequencies results in a reduction of contrast in LSF, and the enhancement of HSF contrast. It therefore induces severe modifications in the amplitude spectrum properties of stimuli. Many studies attest to the importance of the amplitude spectrum during scene recognition, and psychophysical and computational studies have observed that the distribution of contrast across spatial frequencies and orientations in scenes follows statistical regularities that are exploited to categorize scenes (Field, 1987; Guyader et al., 2004; Oliva & Torralba, 2001; Torralba & Oliva, 2003). These statistical regularities on the amplitude spectrum are also relevant on a neurobiological level. In an fMRI study, Andrews et al. (2010), showed for example, that category-selective regions of the inferotemporal cortex, such as the parahippocampal place area (Epstein & Kanwisher, 1998) and the fusiform face area (Kanwisher, McDermott, & Chun, 1997), responded more strongly to amplitude spectrum properties that were typical of the preferred category (i.e. scenes and faces, respectively). Therefore, any modification of the amplitude spectrum using RMS contrast

equalization may result in an irrelevant analysis of its properties by the visual system (including high-level areas), and may bias visual processing strategies. RMS contrast normalization should, therefore, be used with caution when investigating spatial frequency processing in scenes.

To sum up, the present study investigated how rapid scene categorization is influenced on a behavioral level by the presentation order of spatial frequencies, the accumulation mode of spatial frequency over time, and differences in luminance contrast between spatial frequencies. Firstly, we showed that categorization performance was enhanced when LSF information was available before HSF information, irrespective of the spatial frequency accumulation mode in sequences (i.e. successive or additive). This result supports influential models of visual perception and suggests that coarse-to-fine processing is the predominant, and indeed default strategy in scene analysis (Bar, 2003; Kauffmann et al., 2014; Peyrin et al., 2010; Schyns & Oliva, 1994). We also demonstrated that the advantage of coarse-to-fine processing over fine-to-coarse processing was driven by both spatial frequencies and luminance contrast in scenes, and to a greater extent by differences in spatial frequencies. The present study also cautions against the use of luminance contrast normalization in studies investigating spatial frequency processing. We argue that such manipulation impairs properties in the stimulus which need to be exploited by the visual system to enable recognition, and may bias visual processing strategies.

Acknowledgment

This work was supported by the RECOR ANR Grant (ANR-12-JHS2-0002-01 RECOR). Louise Kauffmann was supported by Région Rhône-Alpes (Cible Grants). We thank Catherine Dal Molin for the English revision of the manuscript.

Footnotes

[1] The use of large visual scenes covering 24 x 18 degrees of visual angle was motivated by several reasons. Firstly, large visual scenes ensure to stimulate both the fovea and the peripheral visual field and avoid a bias for HSF processing. Indeed, a recent study (Musel et al., 2013) showed that HSF information contained in similar large scenes elicited strong activation within the occipital cortex linked to the foveal projection whereas LSF scenes activated more strongly retinotopic areas involved in the processing of the peripheral visual field. Furthermore, it should be noted that we are conducting parallel work on patients with age-related macular degeneration (AMD), for whom small stimuli sizes are not appropriate since these patients present a deficit in processing information in the central visual field. Therefore, large visual scenes were also used in the present study in order to maintain consistency with these works and adapt our paradigms with these patients.

[2] The logarithmic scale used to define central frequencies of the filters was given by $f_k = f_0/X^k$, with f_0 the highest central frequency, X the geometric separation between each filter central frequency, and k the filter number from high to low frequency.

[3] Each filter is built as follows: $F_k = \min(1, F_{k-1} + G_k)$, with F_k the filter for the image k and G_k the band pass Gaussian filter k.

[4] It should be noted that we also created and tested in a pilot experiment the difference between additive (i.e. sequences starting with either LSF or HSF) and reverse sequences (i.e. sequences starting with an almost intact scenes and ending with LSF or HSF, respectively). Results showed that these reverse sequences were categorized faster than additive sequences ($F_{1, 29} = 5.24$ $p < 0.05$). Furthermore, for reverse sequences, there was no difference between sequences ending with LSF and sequences ending with HSF ($F_{1, 29} < 1$). The advantage of reverse sequences on additive sequences could be simply due to the fact that, reverse sequences started with the whole spatial frequency information. Therefore, these sequences were not further used in the present experiment.

[5] We kept dynamic sequences short in our experiments (150 ms) in order to ensure optimal imitation of the rapid spatial frequency-based processing of scenes. It is therefore possible that this short presentation time did not allow us to distinguish a clear difference between the successive and additive conditions. A protocol involving the progressive revelation of scenes by gradually adding HSF and LSF information to LSF and HSF scenes, respectively, and in which a categorical choice is made at each revelation step might be used to work out if the addition of spatial frequency information is more advantageous than the successive processing of information from different spatial frequency bands in scene categorization, since this procedure allows us to determine exactly how much information is needed to enable recognition, even if it does not reflect rapid scene processing. We thus conducted a control experiment on 8 participants (5 females, mean age ±

SD: 20 ± 2 years) in which participants had to perform a categorization task after each image in the sequence. We presented sequences of six filtered scenes, according to two accumulation modes (Successive and Additive) and two spatial frequency orders (CtF and FtC). Each scene composing the sequences was presented individually during 25 ms and immediately followed by a mask during 30 ms and a grey background screen during which the participants could give a response. Thus, after each image, participants were asked to categorize the scene as belonging to two different categories (indoor and outdoor). They were asked to respond as rapidly and accurately as possible. The second image of the sequence appeared once a response was given, and similarly for the next images in the sequence. Results revealed that participants correctly categorized the scene earlier when it was presented in a coarse-to-fine (Mean ± SD: 1.06 ± 0.07 images on 6 images in the sequences) than a fine-to coarse sequence (Mean ± SD: 1.24 ± 0.14 images; $F_{1, 7} = 23.16$, $p < 0.005$). Furthermore, they correctly categorized the scene more rapidly in a coarse-to-fine (Mean ± SD: 526 ± 92 ms) than a fine-to-coarse sequence (Mean ± SD: 648 ± 187 ms; $F_{1, 7} = 11.85$, $p < 0.05$). However, there was no effect of the accumulation mode and no interaction between the accumulation mode and the spatial frequency order for all measures (all Fs < 1).

References

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

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*, 449–454.

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of Optical Society of America*, *19*, 1096–1106.

Boynton, G. M. (2005). Contrast gain in the brain. *Neuron*, *47*, 476–477.

Bullier, J. (2001). Integrated model of visual processing. *Brain research*, *36*, 96–107.

De Cesarei, A., & Codispoti, M. (2011). Scene identification and emotional response: which spatial frequencies are critical? *The Journal of neuroscience*, *31*, 17052–7.

De Cesarei, A., & Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychonomic bulletin & review*, *18*, 840–7.

De Cesarei, A., Mastria, S., & Codispoti, M. (2013). Early spatial frequency processing of natural images: an ERP study. *PloS one*, *8*, e65103.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559.

De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 531–544.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 6–9.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of Optical Society of America*, *4*, 2379–2394.

Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In *Handbook of Perception and Human Performance* (Ginsburg, pp. 1–41). Wiley, New York,.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral cortex*, *21*, 467–76.

Guyader, N., Chauvin, A., Peyrin, C., Hérault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, *327*, 313–318.

Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision research*, *18*, 1579–86.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in neurobiology*, *84*, 405–39.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global Precedence, Spatial Frequency Channels, and the Statistics of Natural Images. *Journal of Cognitive Neuroscience*, *8*, 197–230.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of neuroscience*, *17*, 4302–11.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in integrative neuroscience*, *8*. doi:10.3389/fnint.2014.00037

Kihara, K., & Takeda, Y. (2010). Time course of the integration of spatial frequency-based information in natural scenes. *Vision research*, *50*, 2158–62.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of neuroscience*, *27*, 13232–40.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic bulletin & review*, *12*, 43–65.

Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling. *Brain and cognition*, *57*, 151–7.

Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental brain research.*, *227*, 367–77.

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PloS one*, *7*, e38493.

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine Categorization of Visual Scenes in Scene-selective Cortex. *Journal of cognitive neuroscience*. doi:10.1162/jocn_a_00643

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, *42*, 145–175.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–60.

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*, 2768–80.

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., & Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *NeuroImage*, *28*, 464–73.

Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative ophthalmology*, *11*, 368–77.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*, 1–22.

Schettino, A., Loeys, T., Bossi, M., & Pourtois, G. (2012). Valence-specific modulation in the accumulation of perceptual evidence prior to visual scene recognition. *PloS one*, *7*, e38064.

Schettino, A., Loeys, T., Delplanque, S., & Pourtois, G. (2011). Brain dynamics of upstream perceptual processes leading to visual object recognition: a high density ERP topographic mapping study. *NeuroImage*, *55*, 1227–41.

Schettino, A., Loeys, T., & Pourtois, G. (2013). Multiple synergistic effects of emotion and memory on proactive processes leading to scene recognition. *NeuroImage*, *81C*, 81–95.

Schyns, P. G., & Oliva, A. (1994). Evidence for time- and spatial-scale-dependent scene recognition. *Psychological science*, *5*, 195–201.

Shams, L., & Von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision research*, *42*, 2547–54.

Shapley, R., & Enroth-Cugell, C. (1984). Visual Adaptation and Retinal Gain Controls. *Progress in retinal research*, *3*, 263–346.

Shapley, R. M., & Lam, D. M. K. (Eds.). (1993). *Contrast sensitivity* (Vol. 5). MIT Press.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic & physiological optics*, *12*, 229–32.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, *14*, 391–412.

Vlamings, P. H. J. M., Goffaux, V., & Kemner, C. (2009). Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information ? *Journal of vision*, *9(5)*, 1–13.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., … Gosselin, F. (2010). Does face inversion change spatial frequency tuning? *Journal of experimental psychology. Human perception and performance*, *36*, 122–35

.

2.3. Conclusions

Les résultats des études comportementales présentées dans ce chapitre vont donc à l'appui de nombreux travaux, qui suggéraient qu'une analyse « coarse-to-fine » est privilégiée et avantageuse pour la perception (Parker et al., 1992, 1997) et la catégorisation rapide de scènes (e.g., Musel et al., 2012; Schyns & Oliva, 1994). D'un point de vue méthodologique, ces études ont permis d'évaluer le rôle de facteurs expérimentaux tels que le mode de présentation de l'information en fréquences spatiales au cours des séquences (i.e. Successive ou Additive, **Expérience 1**) ou le contraste associé aux scènes filtrées (**Expérience 2**). En effet, ces deux études comportementales ont permis de montrer (1) que l'accumulation de l'information en fréquences spatiales de manière additive ne facilite pas davantage la catégorisation « coarse-to-fine » des scènes et (2) que l'avantage d'une analyse « coarse-to-fine » est en partie expliqué par une différence de contraste entre les BFS et les HFS mais résulte de façon prédominante d'un effet des fréquences spatiales.

On peut noter que les résultats comportementaux des précédentes études utilisant des visages filtrés dont le contraste RMS est égalisé rapportent également un effet moins important des fréquences spatiales (Goffaux et al., 2011; Vlamings et al., 2009). Par exemple, dans une étude en EEG, Vlamings et al. (2009) ont utilisé des visages filtrés en BFS et en HFS dont le contraste RMS était égalisé ou non. Au niveau comportemental, ces auteurs ont observé que les visages en BFS étaient catégorisés plus rapidement que les visages en HFS. Cependant, cet effet des fréquences spatiales était significativement réduit lorsque le contraste RMS était égalisé entre les BFS et les HFS. L'égalisation du contraste RMS semble donc influencer le traitement des fréquences spatiales. Comme nous l'avons vu précédemment, la répartition asymétrique du contraste en fonction des fréquences spatiales est une propriété intrinsèque des scènes naturelles qui semble être exploitée par le système visuel pour la reconnaissance et la catégorisation de scènes (Guyader et al., 2004; Hérault et al., 1997; Oliva et al., 1999; Torralba & Oliva, 2003). On peut alors se demander si l'égalisation du contraste RMS des scènes filtrées n'implique pas une manipulation contre-nature aux propriétés fonctionnelles du système visuel, susceptible de biaiser les stratégies d'analyse. Nous verrons dans la suite de ce manuscrit que l'égalisation du contraste RMS peut effectivement influencer le traitement des fréquences spatiales, et ce, même dans les régions cérébrales sélectives aux scènes naturelles (Chapitre 4).

En résumé, ces deux expériences montrent qu'une analyse « coarse-to-fine » est plus avantageuse pour la catégorisation rapide de scènes, et elles nous permettent également de

sélectionner les paramètres de stimulation visuelle appropriés pour l'étude des bases cérébrales de l'analyse « coarse-to-fine » des scènes présentée dans le Chapitre 3 qui suit.

RESUME

Dans ce chapitre, nous avons présenté deux expériences comportementales visant à spécifier les mécanismes de traitement des fréquences spatiales lors de l'analyse « coarse-to-fine » des scènes. Pour cela, nous avons utilisé des séquences de scènes filtrées afin de simuler et d'imposer au système visuel l'ordre d'analyse des fréquences spatiales. Dans une première expérience (**Expérience 1**), nous avons examiné l'influence de l'ordre de présentation des fréquences spatiales (i.e. des BSF aux HFS ou des HFS aux BFS) lors de la catégorisation rapide de scènes (extérieur vs. intérieur). Nous avons également examiné si l'accumulation de l'information en fréquences spatiales de façon additive (i.e. addition de l'information en fréquences spatiales d'une image à l'autre dans la séquence) était plus avantageuse pour la catégorisation rapide de scènes qu'une accumulation de façon successive (i.e. traitement de bandes de fréquences spatiales différentes d'une image à l'autre dans la séquence). Nous avons montré qu'une analyse « coarse-to-fine » était toujours plus avantageuse pour une catégorisation rapide et que l'accumulation de l'information en fréquences spatiales de façon additive ou successive n'influençait pas les performances. Ces résultats suggèrent donc un rôle important de l'ordre de traitement des fréquences spatiales pour la catégorisation rapide de scènes, indépendamment de la façon dont elles sont accumulées par le système visuel. Dans une deuxième expérience (**Expérience 2**), nous avons testé dans quelle mesure l'avantage d'une analyse « coarse-to-fine » pouvait être expliqué par une différence de contraste entre les fréquences spatiales. Nous avons utilisé un paradigme permettant de tester séparément l'effet des fréquences spatiales et l'effet du contraste, par rapport à l'effet combiné des fréquences spatiales et du contraste. Nous avons montré que si la différence de contraste entre les fréquences spatiales expliquerait en partie l'avantage d'une analyse « coarse-to-fine », cet avantage serait dû de façon prédominante au contenu en fréquences spatiales des séquences. Ces résultats nous permettent donc de préciser les paramètres de stimulation visuelle pouvant influencer le traitement des fréquences spatiales et de sélectionner des stimuli appropriés pour l'étude des bases cérébrales de l'analyse « coarse-to-fine » des scènes.

CHAPITRE 3 - BASES CEREBRALES DE L'ANALYSE « COARSE-TO-FINE »

Ce chapitre sera consacré à l'étude des bases cérébrales de l'analyse « coarse-to-fine » lors de la perception et la catégorisation de scènes. Nous présenterons tout d'abord quelques travaux portant sur le traitement des fréquences spatiales au sein du cortex visuel qui indiquent que le traitement des fréquences spatiales se ferait en lien avec l'organisation rétinotopique du cortex visuel, mais aussi de façon latéralisée entre les hémisphères. Ces travaux seront peu détaillés ici de par leur intérêt moindre pour la suite de ce travail, mais ont fait l'objet d'une revue de littérature publiée dans la revue *Frontiers in Integrative Neuroscience* (Kauffmann, Ramanoël, & Peyrin, 2014; cf.ANNEXE I).

Nous présenterons ensuite les travaux récents qui se sont intéressés aux bases cérébrales d'une analyse « coarse to fine » des scènes. Dans la continuité de ces travaux, nous avons mené deux études en IRMf, la première ayant pour objectif de préciser les bases cérébrales de l'analyse « coarse-to-fine » ainsi que la connectivité au sein de ce réseau (**Expérience 3**), et la seconde visant à examiner la façon dont le traitement des BFS peut influencer le traitement des HFS (**Expérience 4**).

1. Organisation rétinotopique et latéralisée du traitement des fréquences spatiales au sein du cortex visuel

Sur la base de données issues de la neuropsychologie (Dos Santos, Andrade, & Fernandez Calvo, 2013; Lamb, Robertson, & Knight, 1990; Peyrin, Chokron, et al., 2006; Robertson & Lamb, 1991) et d'études comportementales utilisant des formes hiérarchiques (Sergent, 1982), des réseaux sinusoïdaux (Kitterle, Christman, & Hellige, 1990; Kitterle, Hellige, & Christman, 1992; Kitterle & Selig, 1991; Reinvang, Magnussen, & Greenlee, 2002) ou des scènes filtrées en fréquences spatiales comme stimuli (Peyrin, Chauvin, Chokron, & Marendaz, 2003; Peyrin, Mermillod, Chokron, & Marendaz, 2006; Peyrin, Chokron, et al., 2006; voir aussi pour une revue Grabowska & Nowicka, 1996), plusieurs travaux en neuroimagerie ont mis en évidence une spécialisation hémisphérique pour le traitement des fréquences spatiales au niveau du cortex visuel (Fink et al., 1996; Fink,

Marshall, Halligan, & Dolan, 2000; Fink, Marshall, et al., 1997; Iidaka, Yamashita, Kashikura, & Yonekura, 2004; Peyrin, Baciu, Segebarth, & Marendaz, 2004). Le cortex occipital droit serait prédominant pour le traitement des BFS, tandis que le cortex occipital gauche serait prédominant pour le traitement des HFS. En effet, Peyrin et al. (2004) ont par exemple mis en évidence, en comparant directement l'activité d'un hémisphère à l'activité de l'hémisphère controlatéral (méthode de comparaison directe inter-hémisphérique), une activité plus importante du gyrus occipital moyen droit que gauche lors du traitement de scènes en BFS, et une activité plus importante du gyrus occipital moyen gauche que droit lors du traitement de scènes en HFS. Ces résultats ont donc fourni des arguments en faveur d'une spécialisation hémisphérique pour le traitement des fréquences spatiales lors de la catégorisation de scènes, et ce, dès le cortex occipital.

Cependant, d'autres travaux rapportent que le traitement des fréquences spatiales ne se ferait pas de manière latéralisée, mais suivrait l'organisation rétinotopique du cortex visuel. En effet, rappelons que des données obtenues chez des patients cérébrolésés (Holmes, 1918; Horton & Hoyt, 1991), chez le macaque (Daniel & Whitteridge, 1961; Gattass, Gross, & Sandell, 1981; Gattass, Sousa, & Gross, 1988), ainsi que chez l'homme sain (Engel et al., 1994, 1997), indiquent une correspondance entre les points du champ visuel et leur projection au niveau du cortex visuel. La partie centrale (fovéale) du champ visuel se projette sur la partie latérale postérieure du cortex visuel, tandis que la partie périphérique du champ visuel se projette sur la partie médiale antérieure du cortex visuel, en avant de la scissure calcarine. Par ailleurs, nous avons vu que les photorécepteurs et les cellules ganglionnaires ne sont pas répartis de façon homogène sur la rétine (Curcio & Allen, 1990; Curcio et al., 1990; Osterberg, 1935). La densité de cônes et de cellules ganglionnaires de type « midget », qui sont à l'origine de la voie parvocellulaire (véhiculant principalement les HFS), est plus importante dans la partie fovéale de la rétine, tandis que la densité de bâtonnets et de cellules ganglionnaires de type « parasol », qui sont à l'origine de la voie magnocellulaire (véhiculant principalement les BFS), est plus importante en périphérie de la rétine. Ainsi, l'information en HFS devrait être projetée dans les aires dédiées au traitement du champ visuel fovéal, tandis que l'information en BFS devrait être projetée dans les aires dédiées au traitement du champ visuel périphérique.

C'est en effet ce qui ressort d'études en neurophysiologie chez le chat (Everson et al., 1998; Issa, Trepel, & Stryker, 2000), le primate non-humain (De Valois, Albrecht, et al., 1982; Foster, Gaska, Nagler, & Pollen, 1985; Gegenfurtner, Kiper, & Levitt, 1997; Xu,

Anderson, & Casagrande, 2007), ainsi que chez l'humain (Henriksson, Nurminen, & Vanni, 2008; Sasaki et al., 2001; Singh, Smith, & Greenlee, 2000), qui ont cartographié la représentation des fréquences spatiales au sein des aires rétinotopiques du cortex visuel. Par exemple, Sasaki et al. (2001) ont mesuré l'activité cérébrale induite par la vision passive de réseaux sinusoïdaux, dont la fréquence spatiale variait entre 0,05 cpd (BFS) et 2 cpd (HFS) et couvrant une large partie du champ visuel (48 × 36 degrés d'angle visuel). Ils ont ensuite mis en correspondance ces activations avec une carte rétinotopique d'excentricité, acquise en stimulant différentes excentricités du champ visuel à l'aide d'un damier en expansion. Les résultats (cf. Figure 16) ont montré que les activations induites par les réseaux en BFS correspondaient à la représentation du champ visuel périphérique (i.e. dans la partie médiale antérieure du cortex visuel) et que les activations induites par les réseaux en HFS correspondaient à la représentation du champ visuel central (i.e. dans la partie latérale postérieure du cortex visuel).

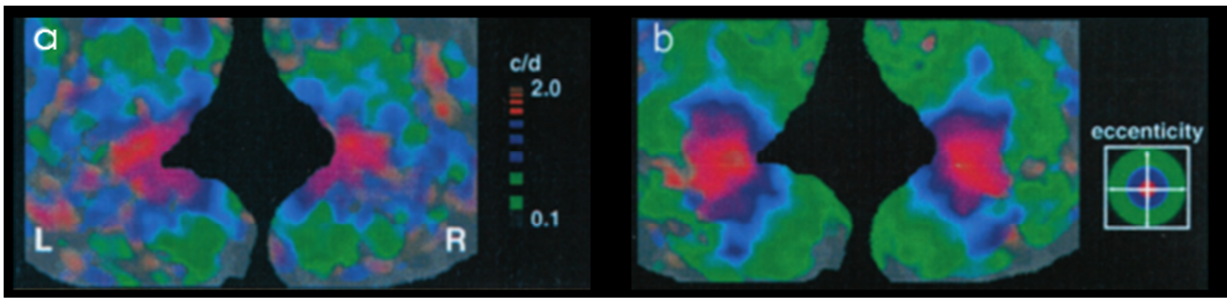


Figure 16 : Cartographie de la représentation des fréquences spatiales et de l'excentricité dans le champ visuel obtenue par Sasaki et al. (2001). (a) Cartographie de la représentation des fréquences spatiales (BFS en vert et HFS en rouge) chez un participant. (b) Cartographie rétinotopique de l'excentricité dans le champ visuel (fovéa en rouge et périphérie en vert) chez le même participant. Adapté de Sasaki et al. (2001).

On peut noter que cette étude n'a pas révélé d'asymétrie cérébrale au niveau du cortex occipital pour le traitement des fréquences spatiales, telle que mise en évidence par Peyrin et al. (2004). Cependant, un point méthodologique important pourrait expliquer ces divergences de résultats. En effet, Sasaki et al. (2001) ont comparé l'activité induite par les BFS à l'activité induite par les HFS (comparaison inter-conditions). Avec ce type de comparaison, Peyrin et al. (2004) n'observaient pas non plus d'asymétries cérébrales pour le traitement des BFS et des HFS. Seule la comparaison directe inter-hémisphérique leur a permis de mettre en évidence une activité plus importante du cortex occipital droit pour le traitement de BFS et du cortex occipital gauche pour le traitement des HFS. Notons également qu'en comparant les activations induites par les BFS à celles induites par les HFS, Peyrin et al. (2004) n'ont toutefois pas observé d'organisation rétinotopique du traitement des fréquences spatiales.

Cependant, ces auteurs utilisaient des stimuli de petite taille, couvrant seulement 4 degrés d'angle visuel (*versus* 48 × 36 degrés d'angle visuel dans l'étude de Sasaki et al., 2001). Ces stimuli stimulant principalement le champ visuel central n'étaient donc pas appropriés pour l'étude de la représentation des fréquences spatiales en fonction de l'excentricité dans le champ visuel.

Dans une étude récente en IRMf, Musel et al. (2013) sont parvenus à mettre en évidence la coexistence d'une organisation rétinotopique, mais également latéralisée, du traitement des fréquences spatiales lors de la catégorisation de scènes (cf. Figure 17). Ces auteurs ont utilisé de grandes images de scènes couvrant 24 × 18 degrés d'angle visuel et filtrées en BFS et en HFS. Les participants avaient pour tâche de catégoriser les scènes comme étant des scènes d'extérieur ou d'intérieur. En comparant les activations induites par les BFS aux activations induites par les HFS (comparaison inter-conditions), les auteurs ont observé que les BFS activaient plus fortement la partie médiale antérieure du cortex visuel, correspondant à la représentation du champ visuel périphérique, tandis que les HFS activaient plus fortement la partie latérale postérieure du cortex visuel, correspondant à la représentation du champ visuel central. Par ailleurs, la projection de ces activations sur les cartes rétinotopiques acquise chez trois participants a révélé que les BFS étaient représentées sur la partie antérieure de V1, correspondant à la représentation du champ visuel supérieure à 3 degrés d'excentricité, tandis que les HFS étaient représentées sur les parties postérieures et ventrales de V2, V3 et V4, correspondant à la représentation du champ visuel inférieure à 3 degrés d'excentricité Ces résultats indiquent donc une organisation rétinotopique du traitement des fréquences spatiales lors de la catégorisation de scènes. Les auteurs ont, par la suite, comparé les activations entre les deux hémisphères (comparaison inter-hémisphérique) pour les BFS et HFS séparément. Ces analyses ont révélé une activation plus importante du gyrus occipital moyen droit que gauche lors de la catégorisation de scènes en BFS, et une activation plus importante du gyrus temporal moyen gauche que droit lors de la catégorisation des scènes en HFS, en accord avec la latéralisation pour le traitement des fréquences spatiales précédemment rapportée (Iidaka et al., 2004; Peyrin et al., 2004).

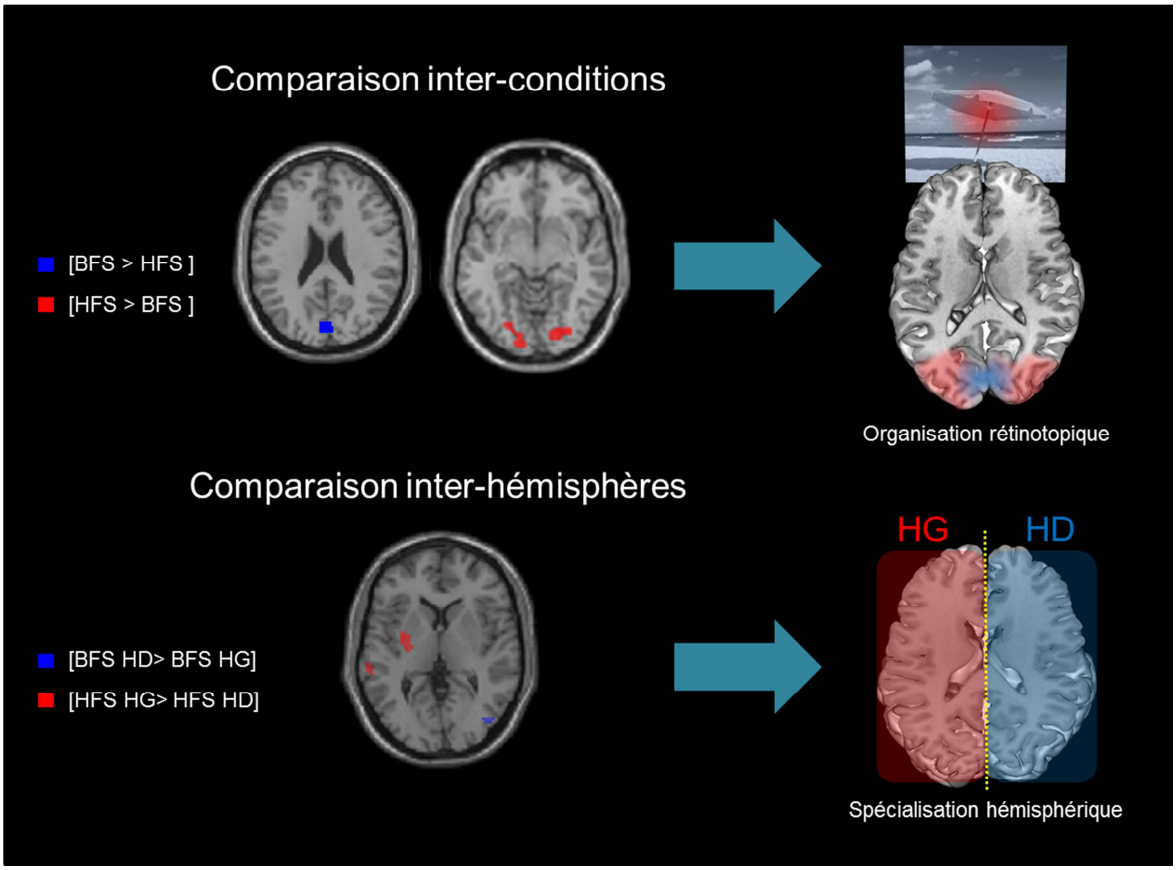


Figure 17 : Résultats obtenus par Musel et al. (2013) en fonction de la méthode d'analyse utilisée. La comparaison entre les deux conditions de fréquences spatiales révèle une organisation rétinotopique du traitement des fréquences spatiales tandis que la comparaison inter-hémisphérique (HD : hémisphère droit ; HG : hémisphère gauche) révèle une spécialisation hémisphérique pour le traitement des fréquences spatiales. Adapté de Musel et al. (2013).

Les résultats de cette étude ont donc permis de mettre en évidence pour la première fois la coexistence d'une organisation rétinotopique, mais également d'une spécialisation hémisphérique, pour le traitement des fréquences spatiales, et ce, sur les mêmes données. Cette étude souligne donc le rôle majeur de la méthode d'analyse (i.e. comparaison inter-conditions ou inter-hémisphérique) pour mettre en évidence différents types de mécanismes de traitement des fréquences spatiales (i.e. rétinotopique ou latéralisé). Par ailleurs, ces résultats suggèrent que les stimuli de grande taille, couvrant une large proportion du champ visuel (i.e. 24 × 18°) sont appropriés pour étudier le traitement des fréquences spatiales, puisqu'ils permettent de stimuler à la fois la fovéa et la périphérie du champ visuel.

Dans une étude IRMf menée en collaboration avec Stephen Ramanoël dans le cadre de sa thèse (Ramanoël, Kauffmann, Cousin, Dojat, & Peyrin, Sous presse; cf.ANNEXE II), nous nous sommes demandé si le vieillissement normal s'accompagnait de changements majeurs du traitement des fréquences spatiales au sein du cortex occipital. Des participants jeunes, de moins de 30 ans, et des participants âgés, de plus de 60 ans, avaient pour tâche de catégoriser

des scènes en BFS et HFS (extérieur *versus* intérieur). Au niveau comportemental, nous avons observé un déficit du traitement des HFS chez les participants âgés, par rapport aux participants jeunes. Au niveau cérébral, les participants jeunes présentaient une organisation rétinotopique du traitement des fréquences spatiales similaire à celle précédemment observée par Musel et al. (2013). La catégorisation des scènes en BFS activait plus fortement la partie médiale antérieure du cortex occipital, en lien avec la représentation du champ visuel périphérique, tandis que la catégorisation des scènes en HFS activait plus fortement la partie latérale postérieure du cortex occipital, en lien avec la représentation du champ visuel central. Les participants âgés présentaient une activation occipitale pour la catégorisation des scènes en BFS (comparée aux scènes en HFS), mais pas pour la catégorisation des scènes en HFS (comparée aux scènes en BFS). La comparaison entre les groupes montrait clairement une plus forte activation du cortex occipital pour les participants jeunes qu'âgés lors de la catégorisation des scènes en HFS. En contrepartie, les participants âgés activaient plus fortement des régions temporales, pariétales et frontales quelles que soient les fréquences spatiales à traiter. Cette étude montre clairement un déficit du traitement des HFS chez les personnes âgées qui s'accompagne d'une hypoactivité du cortex occipital. Ces résultats ne sont pas sans conséquence pour la caractérisation des troubles visuels au cours du vieillissement pathologique. Nous reviendrons plus en détails dans le Chapitre 5, consacré à la Discussion générale, sur l'interprétation et les implications cliniques de ces différences en fonction de l'âge.

Dans l'ensemble, ces travaux suggèrent donc le rôle fondamental du cortex occipital dans le traitement des fréquences spatiales, et ce, à tous les moments de la vie. D'autres travaux rapportent qu'au sein d'un large réseau cortical comprenant également des régions frontales et temporales, le cortex occipital jouerait également un rôle primordial dans l'intégration des BFS et HFS lors de l'analyse « coarse-to-fine » de scènes. Dans la partie qui suit nous présenterons les modèles neurocognitifs qui formalisent la perception visuelle sur la base d'une analyse « coarse-to-fine » des fréquences spatiales, et nous décrirons systématiquement le réseau cortical sous-jacent à cette analyse.

2. Bases cérébrales de l'analyse « coarse-to-fine »

2.1. Modèles neurocognitifs de la reconnaissance visuelle

Le modèle par itération proposé par Bullier (2001) fournit des bases neurophysiologiques venant conforter l'hypothèse d'un traitement des fréquences spatiales selon une analyse « coarse-to-fine ». Rappelons que selon ce modèle, une première vague d'informations, véhiculées rapidement par la voie magnocellulaire (M) viendrait activer le cortex visuel primaire, qui transmettrait à son tour l'information le long des voies dorsale et ventrale, jusque dans des régions pariétales, temporales, et frontales de haut niveau. L'information serait ensuite rétro-injectée vers les aires visuelles de bas niveau (telles que V1 et V2) et permettrait de guider le traitement ultérieur de l'information véhiculée plus lentement par la voie parvocellulaire (P). Ce modèle souligne donc le rôle d'influences d'aires de haut niveau sur les traitements effectués au sein de régions de plus bas niveau (i.e. influences « top-down »). Au vu des propriétés de sélectivité des cellules des voies M et P, les BFS seraient majoritairement véhiculées par la voie M, tandis que les HFS seraient véhiculées majoritairement par la voie P. Le modèle par itération nous amène donc à penser que l'information en BFS pourrait être rapidement traitée au sein de régions de haut niveau, et qu'elle pourrait alors influencer le traitement ultérieur de l'information en HFS au niveau du cortex visuel primaire.

Dans un modèle proposé spécifiquement pour la reconnaissance d'objets, Bar (2003) propose un mécanisme de facilitation de cette reconnaissance, via des influences « top-down », explicitement basé sur un traitement fréquentiel de l'information visuelle. Ce modèle dit « proactif » postule que lors de la perception visuelle d'un objet, l'information en BFS serait rapidement projetée du cortex visuel vers le cortex préfrontal. Cela permettrait d'activer des interprétations possibles du stimulus visuel, sur la base d'une information grossière, et de générer des prédictions quant à la nature de l'objet. Ces prédictions seraient ensuite envoyées vers le cortex inférotemporal, connu pour être impliqué lors des dernières étapes de la reconnaissance et de l'identification d'objets (Tanaka et al., 1991; Tanaka, 1993, 1996), où elles activeraient les représentations d'objets correspondantes. Celles-ci pourraient alors être intégrées au traitement ascendant ou « bottom-up » de l'information acheminée en parallèle du cortex visuel au cortex inférotemporal le long de la voie ventrale (cf. Figure 18). Ainsi, le traitement rapide des BFS et les prédictions générées dans le cortex préfrontal permettraient de limiter les interprétations possibles du stimulus visuel et de contraindre l'analyse plus détaillée de l'information, facilitant ainsi la reconnaissance.

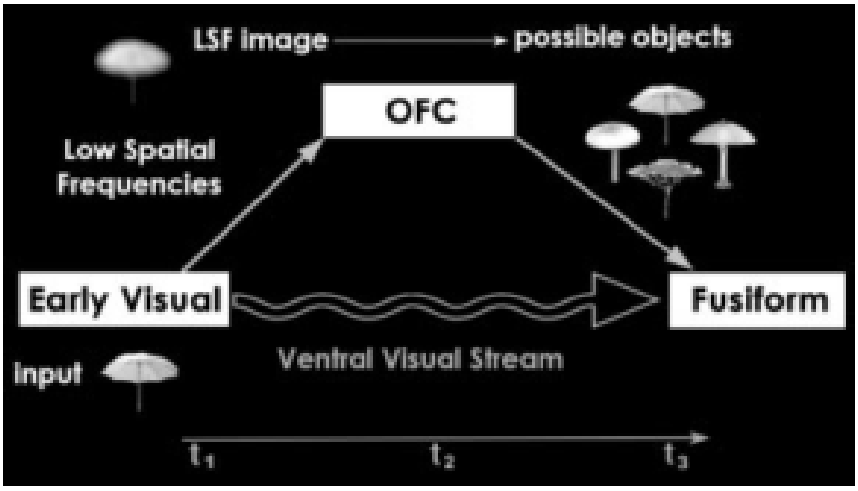


Figure 18 : Illustration du mécanisme de facilitation de la reconnaissance d'objet proposé par Bar (2003). L'information en BFS du stimulus visuel est rapidement projetée du cortex visuel vers le cortex orbitofrontal (OFC). En parallèle, l'information visuelle (BFS et HFS) se propage plus lentement le long de la voie ventrale, jusqu'au gyrus fusiforme dans le cortex inférotemporal. Au niveau du cortex orbitofrontal, l'information en BFS permet d'activer des interprétations possibles du stimulus visuel, qui seront ensuite projetées vers le cortex inférotemporal et intégrées à l'analyse « bottom-up » de l'information visuelle, afin de faciliter sa reconnaissance. Adapté de Bar et al. (2006).

Ce modèle se distingue donc du modèle de Bullier (2001) par la région source des influences « top-down » (ici le cortex préfrontal), ainsi que par la région cible de ces influences, et point d'intégration de l'information en BFS et en HFS (ici le cortex inférotemporal). Selon Bar (2003), l'implication du cortex préfrontal, et plus particulièrement sa partie orbitale (cortex orbitofrontal), lors du traitement rapide des BFS et de l'initiation d'influences « top-down » apparaît plausible dans la mesure où (1) cette région est impliquée dans la reconnaissance visuelle (e.g., Freedman, Riesenhuber, Poggio, & Miller, 2001, 2003), (2) elle reçoit des projections magnocellulaires (Barbas, 1995), (3) elle est directement connectée à des aires visuelles telles que V2 et V4 (Barbas, 1995, 2000) et (4) elle exerce une influence « top-down » sur des régions du cortex inférotemporal, par exemple lors de la récupération d'information en mémoire (e.g., Miyashita & Hayashi, 2000). De plus, la partie orbitale du cortex préfrontal serait tout particulièrement impliquée dans le traitement d'informations visuelles, les prédictions et la génération d'attentes relatives, par exemple, au contenu d'un stimulus visuel ou à la survenue d'une stimulation tactile (Carlsson, Petrovic, Skare, Petersson, & Ingvar, 2000; Fink, Halligan, et al., 1997; Petrides, Alivisatos, & Frey, 2002; Thorpe, Rolls, & Maddison, 1983). Enfin, le cortex préfrontal est fortement connecté au cortex inférotemporal (Cavada, Compañy, Tejedor, Cruz-Rizzolo, & Reinoso-Suárez, 2000).

Ces liens de connectivité pourraient donc fournir un substrat neuro-anatomique au mécanisme d'influences « top-down » proposé dans ce modèle.

Ce modèle prédit donc au niveau neurobiologique une activation précoce du cortex préfrontal, associée spécifiquement au traitement d'une information en BFS, et préalable à l'implication du cortex inférotemporal pour reconnaître un objet. Bar et al. (2006) ont directement testé cette hypothèse dans une étude utilisant la magnétoencéphalographie (MEG) et l'IRMf. Dans une première expérience en MEG, les participants devaient identifier des dessins d'objets présentés pendant 26 ms. Etant donné la courte durée de présentation, les objets étaient difficilement identifiables à leur première présentation, mais la reconnaissance devenait possible avec la répétition des stimuli. Les résultats MEG ont montré que la reconnaissance des objets (relativement aux essais dans lesquels les objets n'étaient pas reconnus) était associée à une activité plus importante du cortex orbitofrontal environ 130 ms après la présentation des stimuli, puis du cortex inférotemporal (gyrus fusiforme) environ 180 ms après la présentation des stimuli (cf. Figure 19a). Ces résultats suggèrent donc que le cortex orbitofrontal interviendrait 50 ms avant le cortex inférotemporal, et que son implication serait alors prédictive de la reconnaissance d'objet. Par ailleurs, les auteurs ont mis en évidence une forte synchronie entre l'activité des régions occipitales et du cortex orbitofrontal d'une part, dans les 100 premières millisecondes suivant la présentation des stimuli, et entre le cortex orbitofrontal et le gyrus fusiforme d'autre part, 130 ms après la présentation des stimuli, suggérant une connectivité fonctionnelle forte entre les régions concernées à ces différentes périodes. Dans une deuxième expérience en IRMf, des images d'objets filtrés en BFS, en HFS et non filtrés étaient présentées pendant 750 ms, et les participants avaient pour tâche de les catégoriser comme étant soit des objets courants, soit des sculptures abstraites. Les auteurs ont montré que les objets non filtrés et filtrés en BFS activaient davantage le cortex orbitofrontal que les objets filtrés en HFS (cf. Figure 19b). Les résultats de cette deuxième expérience suggèrent cette fois-ci que l'activité du cortex orbitofrontal, préalable à la reconnaissance, serait principalement liée au traitement des BFS d'un objet.

Cette étude a donc permis de fournir des arguments neurobiologiques en faveur du modèle proactif proposé par Bar (2003) en démontrant l'implication précoce du cortex orbitofrontal, spécifique au traitement des BFS d'un objet, avant l'implication des aires temporales permettant la reconnaissance finale de l'objet.

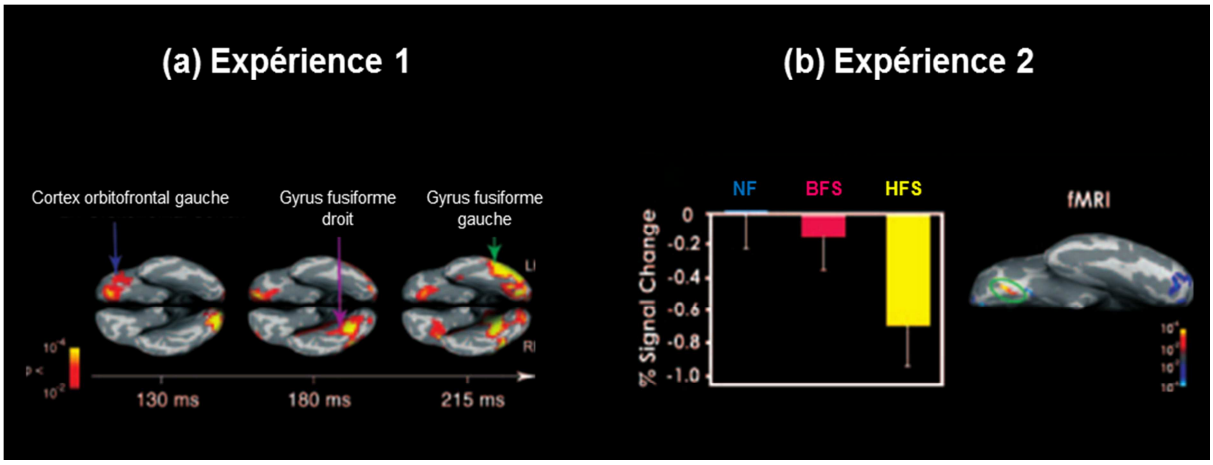


Figure 19 : Résumé des résultats obtenus par Bar et al. (2006). (a) Les résultats de l'Expérience 1 révèlent que la reconnaissance d'objets est associée à une activité plus importante du cortex orbitofrontal 130 ms après la présentation des stimuli, puis des gyri fusiformes droit et gauche 180 et 215 ms après présentation des stimuli, respectivement. (b) Les résultats de l'Expérience 2 (pourcentage de changement de signal dans le cortex orbitofrontal en fonction du contenu en fréquences spatiales de stimuli) montrent que les objets filtrés en BFS et non filtrés (NF) activent plus fortement le cortex orbitofrontal que les objets filtrés en HFS. Adapté de Bar et al. (2006).

Par la suite, Kveraga et al. (2007) ont cherché à comprendre comment le cortex orbitofrontal pouvait être impliqué de façon aussi précoce lors de la perception d'objets, afin d'en faciliter la reconnaissance. Ces auteurs ont fait l'hypothèse que l'implication rapide du cortex orbitofrontal se ferait par le biais d'afférences magnocellulaires, caractérisées par un temps de conduction rapide, permettant de déclencher la facilitation « top–down » vers le cortex inférotemporal. Pour tester cette hypothèse, les auteurs ont mené une étude en IRMf dans laquelle ils utilisaient comme stimuli des dessins d'objets dont les propriétés étaient manipulées pour activer préférentiellement la voie M ou la voie P. Les stimuli biaisés pour activer la voie M (« M-biased ») étaient des stimuli achromatiques et au faible contraste de luminance, tandis que les stimuli biaisés pour activer la voie P (« P-biased ») étaient chromatiques (rouge et vert) et isoluminants. Les participants devaient juger si l'objet présenté était plus grand ou plus petit qu'une boite à chaussures. Dans un premier temps, les auteurs ont comparé les activations induites par les différents types de stimuli (i.e. « M- et P-biased ») et ont observé que le traitement des stimuli « M-biased » était associé à une activité plus importante du cortex orbitofrontal, tandis que le traitement des stimuli « P-biased » était associé à une activité plus importante du cortex inférotemporal, et plus particulièrement du gyrus fusiforme (cf. Figure 20a). Ces résultats semblent donc aller dans le sens de l'hypothèse des auteurs selon laquelle le cortex orbitofrontal recevrait principalement des afférences magnocellulaires lors de la reconnaissance d'objets. Dans un deuxième temps, les auteurs ont effectué une analyse de connectivité sur ces données (« Dynamic causal modelling », DCM).

Avec cette méthode, les auteurs ont examiné, d'une part, la connectivité effective (i.e. influence causale qu'une région exerce sur l'activité d'une autre région) entre le gyrus occipital moyen (région d'entrée du stimulus visuel), le cortex orbitofrontal et le gyrus fusiforme. D'autre part, les auteurs se sont également intéressés à la façon dont la connectivité effective pouvait être modulée en fonction des conditions expérimentales (i.e. catégorisation des stimuli « M-biased » ou « P-biased »). Les résultats ont tout d'abord montré des liens de connectivité effective du gyrus occipital moyen vers le cortex orbitofrontal et le gyrus fusiforme, et de façon très intéressante en ce qui concerne les hypothèses des auteurs, un lien de connectivité effective du cortex orbitofrontal vers le gyrus fusiforme. De plus, les auteurs ont observé que le traitement des stimuli « M-biased » avait pour effet d'augmenter la connectivité entre le gyrus occipital moyen et le cortex orbitofrontal, mais également entre le cortex orbitofrontal et le gyrus fusiforme, tandis que le traitement des stimuli « P-biased » avait pour effet d'augmenter la connectivité entre le gyrus occipital moyen et le gyrus fusiforme (cf. Figure 20b).

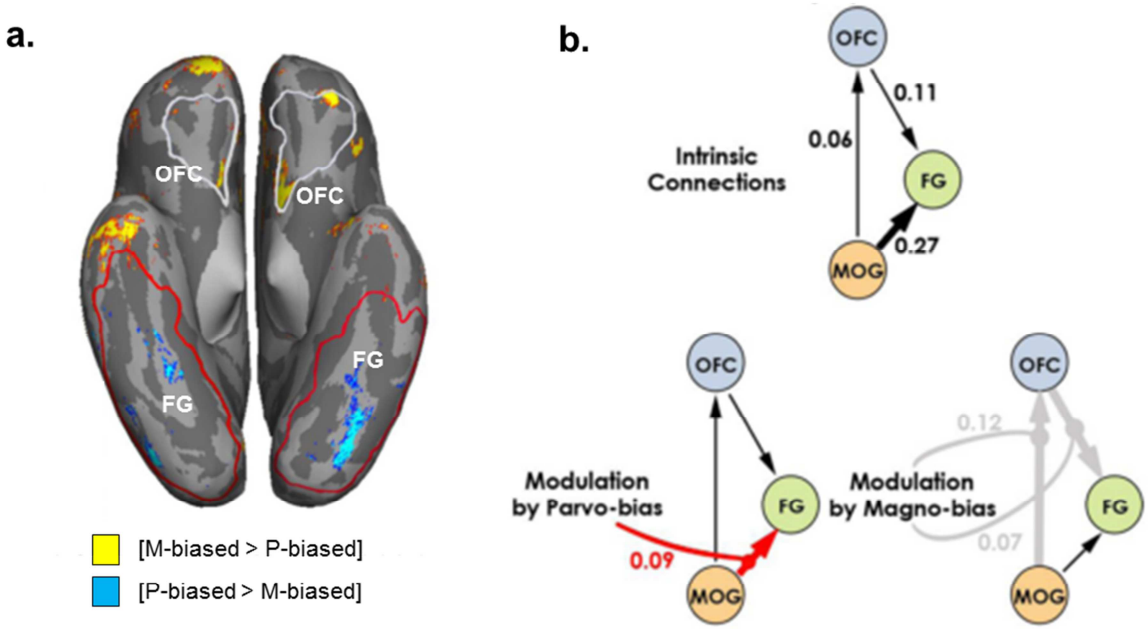


Figure 20 : Résumé des résultats obtenus par Kveraga et al. (2007). (a) Activations induites par les stimuli M-biased (en jaune) et P-biased (en bleu : le cortex orbitofrontal (OFC) et le gyrus fusiforme (FG). (b) Résultats de l'analyse de connectivité. En haut : connectivité effective (i.e. influence qu'une région exerce sur l'activité d'une autre, indépendamment des conditions expérimentales) entre le gyrus occipital moyen (MOG), le cortex orbitofrontal (OFC) et le gyrus fusiforme (FG). Les chiffres correspondent aux paramètres de connectivité effective (exprimés en Hertz). En bas : effet de la condition P-biased (en rouge) sur la connectivité entre le MOG et le FG et de la condition M-biased (en gris) sur la connectivité entre le MOG et l'OFC et entre l'OFC et le FG. Les chiffres correspondent à l'augmentation de la connectivité (en Hertz), par rapport aux paramètres de connectivité effective.

Dans l'ensemble, les résultats de cette étude fournissent donc des arguments supplémentaires en faveur du modèle proactif de Bar (2003 ; 2007 ; Bar et al., 2006), notamment en montrant que le cortex orbitofrontal serait susceptible d'influencer l'activité du cortex inférotemporal (au niveau du gyrus fusiforme). L'implication précoce du cortex orbitofrontal serait permise grâce à des projections magnocellulaires en provenance du gyrus occipital moyen, ce qui permettrait de déclencher des influences « top-down » vers le cortex inférotemporal.

Les données de la neurophysiologie du système visuel du primate non-humain (Bullier, 2001) et les travaux en neuroimagerie réalisés chez l'homme (Bar et al., 2006; Kveraga et al., 2007) présentés ci-dessus fournissent donc des pistes intéressantes pour mieux cerner les bases cérébrales de l'analyse « coarse-to-fine » des scènes. On peut cependant noter que, bien que les études en neuroimagerie postulent une analyse de l'information visuelle selon un mode de traitement « coarse-to-fine », aucune d'entre elles n'a explicitement manipulé l'ordre de présentation de l'information en fréquences spatiales afin de simuler et d'imposer effectivement un traitement « coarse-to-fine » de l'information visuelle. De plus, ces études portent spécifiquement sur bases cérébrales de la reconnaissance d'objets et il n'est donc pas certain que les mêmes substrats corticaux soient à l'œuvre lors de la catégorisation et la reconnaissance de scènes.

Dans une étude combinant l'enregistrement de potentiels évoqués et l'IRMf chez les mêmes participants, Peyrin et al. (2010) ont directement étudié les bases cérébrales et le décours temporel des activations dans les régions impliquées lors de l'analyse « coarse-to-fine » des scènes. Afin de simuler et d'imposer au système visuel une analyse « coarse-to-fine » ou « fine-to-coarse », les auteurs ont présenté des séquences de deux scènes filtrées, la première filtrée en BFS et la seconde en HFS (séquence « coarse-to-fine », CtF) ou inversement (séquence « fine-to-coarse », FtC). Les scènes composant les séquences étaient présentées 100 ms chacune et étaient séparées par un intervalle de 400 ms. Les participants avaient pour tâche de déterminer si les deux scènes de la séquence appartenaient à la même catégorie (intérieurs, plages ou villes). L'analyse des résultats IRMf a tout d'abord montré que les séquences CtF (relativement aux séquences FtC) activaient plus fortement le gyrus frontal inférieur gauche, les « frontal eye fields », le gyrus temporal supérieur et moyen droit et gauche, le lobule pariétal inférieur droit, ainsi que le gyrus occipital moyen droit. Les auteurs ont ensuite examiné le décours temporel de ces activations grâce à des enregistrements EEG

et une localisation de la source de la configuration topographique des champs électriques sur le scalp (ou carte topographique) pour les séquences CtF et FtC. Les résultats ont révélé deux cartes topographiques spécifiques à l'analyse des séquences CtF dont la source (relativement à l'analyse des séquences FtC) était à chaque fois localisée dans le gyrus frontal inférieur : la première fois lors du traitement de la première scène en BFS (entre 150 et 194 ms après l'apparition de la scène) et la deuxième fois lors de l'apparition de la deuxième scène en HFS (entre 12 et 96 ms après son apparition). De façon intéressante, les auteurs ont également observé une carte topographique spécifique au traitement de la scène en HFS seulement lorsque celle-ci était précédée par une scène en BFS (cas de la séquence CtF) et dont la source était localisée dans le cortex occipital (cf. Figure 21). Sur la base des travaux de Bar (2003 ; 2007 ; Bar et al., 2006), les auteurs ont postulé que, lors du traitement des séquences CtF, le cortex préfrontal et les régions temporales seraient rapidement impliquées dans l'analyse sémantique et la génération de prédictions concernant la catégorie de la première scène en BFS. Cela déclencherait des influences « top-down » permettant de guider le traitement ultérieur de la scène en HFS par le cortex occipital, et ce, en accord avec le mécanisme de rétro-injection proposé par Bullier (2001).

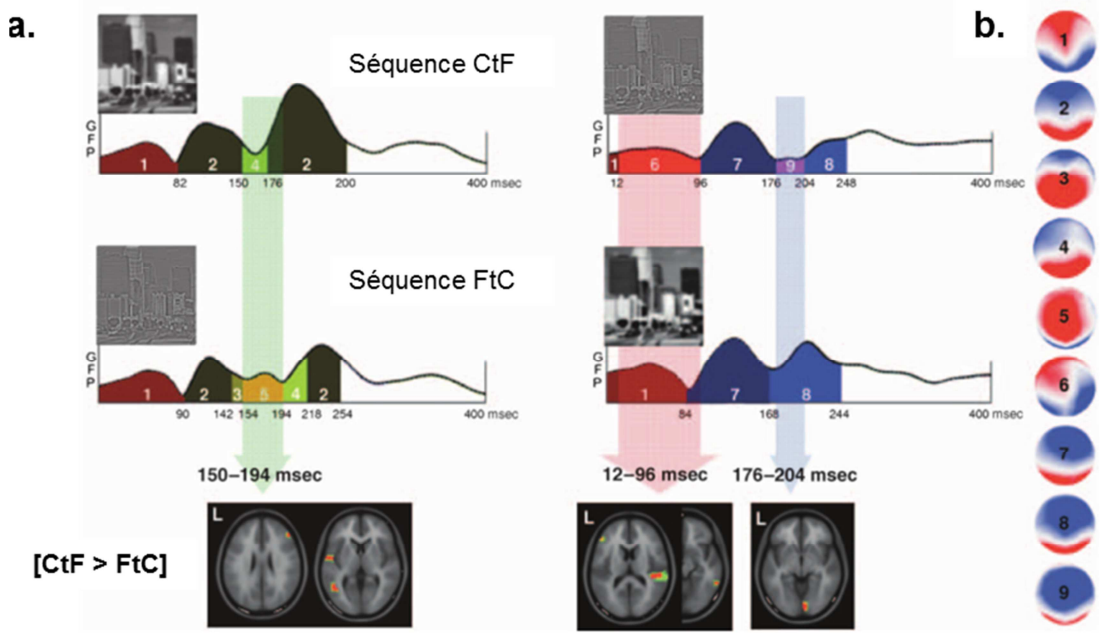


Figure 21 : Résumé des résultats obtenus par Peyrin et al. (2010). (a) Résultats des potentiels évoqués pour les séquences CtF et FtC. (b) Neuf cartes topographiques correspondant à la succession des différentes configurations du champ électrique pour chaque type de séquences. Les flèches colorées indiquent les périodes pour lesquelles l'analyse révèle des différences de topographie entre les deux types de séquences. Les cartes d'activations correspondent à la différence d'activation des sources correspondant à ces périodes. Les séquences CtF (relativement aux séquences FtC) induisent des potentiels évoqués de sources frontales et temporo-pariétales entre 150 et 194 ms après apparition de la première scène de la séquence, et entre 12 et 96 ms après apparition de la deuxième scène, et de source occipital entre 176 et 204 ms après apparition de la deuxième scène.

Les résultats de cette étude vont donc encore une fois à l'appui d'un mécanisme de facilitation « top-down » initié par le cortex préfrontal et permettent de généraliser ce mécanisme à la catégorisation « coarse-to-fine » des scènes. Selon les auteurs, ces résultats soulignent également le rôle intégratif du cortex occipital dans l'analyse « coarse-to-fine ». Plus précisément, le cortex occipital intègrerait les influences de régions cérébrales supérieures au traitement de l'information visuelle en provenance des corps genouillés latéraux. Malheureusement, l'hypothèse d'une influence « top-down » entre les régions cérébrales impliquées dans l'analyse « coarse–to-fine » (i.e. du cortex préfrontal au cortex occipital) n'a pas été directement testée dans cette étude.

Dans la continuité des travaux de Peyrin et al., (2010), nous avons donc mené une étude en IRMf (**Expérience 3**) dans le but (1) d'identifier les régions cérébrales impliquées dans l'analyse « coarse-to-fine » des scènes et (2) de préciser les interactions entre ces régions en utilisant une analyse de connectivité effective (DCM).

2.2. EXPERIENCE 3 : Connectivité effective au sein du réseau cérébral impliqué dans l'analyse « coarse-to-fine » de scènes visuelles

L'**Expérience 3** est présentée en détail dans l'article ci-après, publié dans la revue *Brain and Cognition* (Kauffmann, Chauvin, Pichat, & Peyrin, 2015). Cette étude a eu pour objectif de préciser les bases cérébrales de l'analyse « coarse-to-fine » des scènes, puis de déterminer les interactions entre les régions identifiées afin de tester l'hypothèse avancée par Peyrin et al. (2010) selon laquelle l'augmentation de l'activité du cortex occipital lors du traitement des HFS résulterait d'influences « top-down » en provenance du cortex préfrontal. En nous inspirant des travaux de Kveraga, nous avons utilisé une méthode d'analyse de connectivité effective des données IRMf, le « Dynamic causal modelling » (DCM ; Friston, Harrison, & Penny, 2003). Cette méthode permet d'inférer, dans un cadre Bayésien, l'influence causale qu'une région exerce sur une autre indépendamment du contexte expérimental (i.e. connectivité effective ou endogène), et la façon dont ces liens de connectivité peuvent être modulés en fonction des conditions expérimentales (i.e. modulations).

Afin d'imposer une analyse « coarse-to-fine » des scènes, nous avons utilisé les séquences dynamiques adaptées des **Expérience 1 et 2**. Les séquences étaient composées de six versions filtrées d'une même scène, assemblées des BFS aux HFS (séquences CtF) ou des

HFS aux BFS (séquences FtC). Pour cette étude, nous avons choisi des séquences « successives ». Contrairement aux séquences « additives », les images des séquences « successives » sont construites en utilisant des filtres passe-bande de largeur constante. Au cours des séquences CtF et FtC, les mêmes images sont présentées et seul leur ordre de présentation varie. Ces séquences ont donc l'avantage de contenir exactement la même information en fréquences spatiales. Les images des séquences « additives » sont construites en utilisant des filtres de plus en plus larges afin d'ajouter progressivement de l'information en HFS pour les séquences CtF ou de l'information en BFS pour les séquences FtC. Les séquences « additives » CtF et FtC présentent donc des images différentes. L'utilisation de séquences « successives » dans cette étude nous permettait ainsi d'éviter que des différences d'activations IRMf entre les séquences CtF et FtC puissent être expliquées en termes de différences de signal entre les deux séquences, plutôt qu'en termes d'ordre relatif de traitement des fréquences spatiales. Il est à noter également que le contraste RMS des scènes filtrées était égalisé. En effet, bien que ces stimuli paraissent moins écologiques et qu'ils aient pour effet de réduire les différences observées entre la catégorisation des séquences CtF et FtC au niveau comportemental (cf. **Expérience 2**), ils semblent néanmoins plus adaptés dans le contexte d'études en IRMf, car le signal BOLD, utilisé comme mesure d'activation cérébrale, est très sensible au contraste (Boynton, Engel, Glover, & Heeger, 1996; Boynton, 2005; Dakin, Hess, Ledgeway, & Achtman, 2002; Dean, 1981; Goodyear & Menon, 1998; Olman, Ugurbil, Schrater, & Kersten, 2004; Rieger, Gegenfurtner, Koechy, Heinze, & Grueschow, 2013). L'égalisation du contraste RMS entre les scènes filtrées nous permettait donc d'éviter que les activations observées puissent être dues à des différences de contraste entre les scènes composant les séquences.

Les séquences étaient présentées 150 ms et les participants avaient pour tâche de catégoriser les scènes comme des scènes d'intérieur ou d'extérieur. La comparaison des activations induites par la catégorisation des séquences CtF à celle des séquences FtC (contraste [CtF > FtC]) a révélé une activité plus importante et de façon bilatérale (1) du gyrus frontal inférieur dans sa partie orbitale, (2) du cuneus, un gyrus de la face interne du lobe occipital, et (3) des gyri fusiforme et parahippocampique (cf. Figure 22). Ces régions avaient été également mises en évidence par Bar et al. (2006) et Kveraga et al. (2007) dans des tâches de reconnaissance d'objets, ainsi que par Peyrin et al. (2010) dans une tâche de catégorisation de scènes. Nous avons ensuite conduit une analyse de connectivité effective afin de préciser les interactions entre le cuneus (C), le gyrus frontal inférieur (IFG) et le gyrus fusiforme (FG).

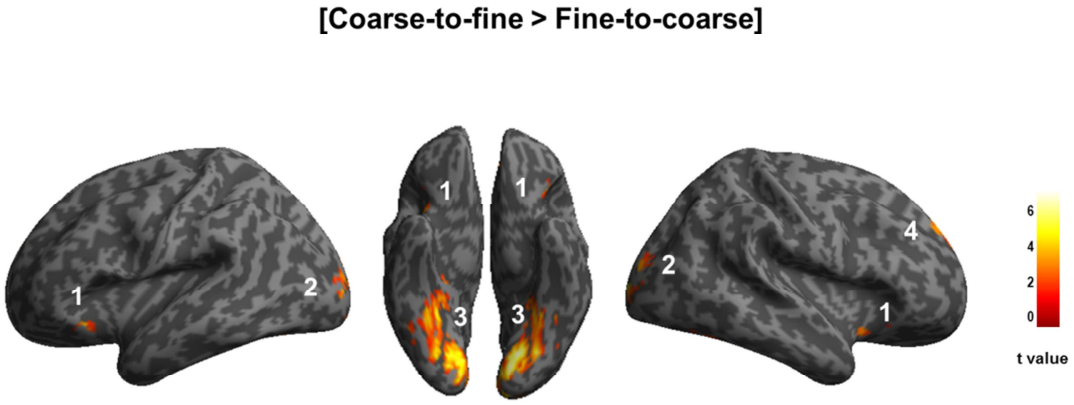


Figure 22 : Activations cérébrales induites par la catégorisation des scènes selon une séquence « coarse-to-fine » relativement à une séquence « fine-to-coarse ». (1) partie orbitale du gyrus frontal inférieur, (2) cuneus, (3) gyri fusiforme et parahippocampique, (4) gyrus frontal supérieur droit.

Les résultats de l'analyse de connectivité ont montré que la catégorisation « coarse-to-fine » des scènes avait pour effet d'augmenter la connectivité du cuneus vers le gyrus frontal inférieur, et du gyrus frontal inférieur vers le gyrus fusiforme, répliquant ainsi les résultats de Kveraga et al. (2007). De façon importante pour l'interprétation des résultats de Peyrin et al. (2010), nous avons également observé que le traitement des séquences CtF avait pour effet d'augmenter la connectivité du gyrus frontal inférieur vers le cuneus (cf. Figure 23). Ces résultats suggèrent donc que l'activité plus importante du cuneus lors du traitement des séquences CtF résulterait principalement d'influences « top-down » en provenance des aires frontales. A noter que cette analyse de connectivité a été réalisée sur des régions de l'hémisphère droit, afin de nous placer dans les mêmes conditions que Kveraga et al. (2007) dont l'analyse de connectivité portait également sur les régions de l'hémisphère droit. De plus, ce choix était motivé par le fait que l'hémisphère droit serait prédominant pour l'analyse « coarse-to-fine » des scènes (Peyrin et al., 2005). Par souci d'exhaustivité, nous avons également réalisé cette analyse sur les régions de l'hémisphère gauche. Les résultats n'ont révélé aucun effet de l'analyse « coarse-to-fine » sur la connectivité du gyrus frontal inférieur vers le gyrus fusiforme ou le cuneus. Les résultats de cette étude sont donc principalement applicables aux régions de l'hémisphère droit.

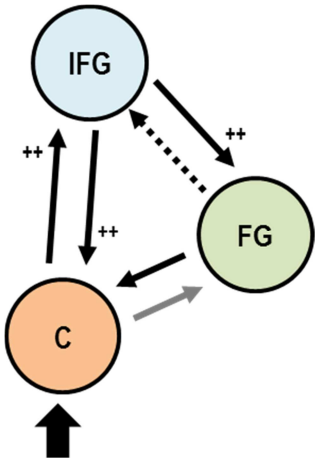


Figure 23 : Synthèse illustrée des résultats de l'analyse de connectivité. Les flèches en gras indiquent les liens de connectivité modulés par les séquences CtF (en noir) et FtC (en gris). La flèche en pointillé correspond à un lien de connectivité effective (i.e. non modulé par les conditions expérimentales). La flèche noire en bas correspond à l'input (stimulation visuelle) au niveau du cuneus. Les symboles « ++ » indiquent que la modulation par les séquences CtF a pour effet d'augmenter la connectivité entre les régions concernées.

Ces résultats viennent donc à l'appui du modèle proactif de Bar (Bar, 2003, 2007; Bar et al., 2006; Kveraga et al., 2007) qui postule un mécanisme de facilitation « top-down » pour la reconnaissance, exercé par le cortex préfrontal sur le cortex inférotemporal. Ils permettent aussi de préciser ce mécanisme dans le contexte de la catégorisation « coarse-to-fine » de scènes et viennent conforter l'idée du rôle intégratif du cortex occipital tel que précédemment proposé par Bullier (2001), puis Peyrin et al. (2010). Le cortex occipital ne serait pas simplement une aire sensorielle codant les afférences rétino-thalamiques, mais intègrerait également les influences d'aires frontales de haut niveau.

Article publié dans la revue *Brain & Cognition* (2015)

Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A Dynamic causal modelling study.

Louise Kauffmann[1,2*], Alan Chauvin[1,2], Cédric Pichat [1,2] & Carole Peyrin[1,2]

[1] Univ. Grenoble Alpes, LPNC, F-38040 Grenoble

[2] CNRS, LPNC UMR 5105, F-38040 Grenoble

Abstract

According to current models of visual perception scenes are processed in terms of spatial frequencies following a predominantly coarse-to-fine processing sequence. Low spatial frequencies (LSF) reach high-order areas rapidly in order to activate plausible interpretations of the visual input. This triggers top-down facilitation that guides subsequent processing of high spatial frequencies (HSF) in lower-level areas such as the inferotemporal and occipital cortices. However, dynamic interactions underlying top-down influences on the occipital cortex have never been systematically investigated. The present fMRI study aimed to further explore the neural bases and effective connectivity underlying coarse-to-fine processing of scenes, particularly the role of the occipital cortex. We used sequences of six filtered scenes as stimuli depicting coarse-to-fine or fine-to-coarse processing of scenes. Participants performed a categorization task on these stimuli (indoor vs. outdoor). Firstly, we showed that coarse-to-fine (compared to fine-to-coarse) sequences elicited stronger activation in the inferior frontal gyrus (in the orbitofrontal cortex), the inferotemporal cortex (in the fusiform and parahippocampal gyri), and the occipital cortex (in the cuneus). Dynamic causal modeling (DCM) was then used to infer effective connectivity between these regions. DCM results revealed that coarse-to-fine processing resulted in increased connectivity from the occipital cortex to the inferior frontal gyrus and from the inferior frontal gyrus to the inferotemporal cortex. Critically, we also observed an increase in connectivity strength from the inferior frontal gyrus to the occipital cortex, suggesting that top-down influences from frontal areas may guide processing of incoming signals. The present results support current models of visual perception and refine them by emphasizing the role of the occipital cortex as a cortical site for feedback projections in the neural network underlying coarse-to-fine processing of scenes.

Keywords: Inferior frontal gyrus; Inferotemporal cortex; Occipital cortex; Orbitofrontal cortex; Top-down facilitation; Spatial frequencies.

Introduction

Natural scenes are complex stimuli that are processed and categorized by the visual system in a remarkably fast and reliable fashion. Supported by convergent data on the functional neuroanatomy of visual pathways (Van Essen & Deyoe, 1995), neurophysiological recordings in primates (De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Hupé et al., 2001; Poggio, 1972; Shams & von der Malsburg, 2002; for a review, see Bullier, 2001), and psychophysical results in humans (Ginsburg, 1986; Hughes et al., 1996; Parker et al., 1992; Schyns & Oliva, 1994), several models of visual perception (Bar, 2003, 2007; Bullier, 2001; Hegdé, 2008; Kauffmann, Ramanoël, & Peyrin, 2014; Schyns & Oliva, 1994) suggest that visual analysis begins with the parallel extraction of different visual elementary features at different spatial frequencies following a predominantly coarse-to-fine default processing sequence. Low spatial frequencies (LSF) in scenes, conveyed by fast magnocellular pathways, provide coarse information about the stimulus (e.g., the global shape and structure of a scene), whereas high spatial frequencies (HSF) in scenes, conveyed more slowly by the parvocellular pathways, provide finer information about the stimulus (e.g., the edges and borders of an object in the scene).

One influential model (Bullier, 2001) proposed an integrated model of visual processing based on neurophysiological recordings in primates. According to this, information (such as LSF in scenes) conveyed rapidly by magnocellular pathways reaches the visual cortex first and then proceeds to high-order areas (parietal, frontal, and temporal areas) to generate a first-pass analysis of visual scenes. Results from these first-pass computations might be then retroinjected through feedback signals into lower level areas (e.g., primary visual cortex, V1) in time for the arrival of the visual information (such as HSF in scenes) conveyed more slowly by the parvocellular pathway. This retroinjected information is used to guide further processing of parvocellular information in high-order areas of the ventral cortical stream that ultimately mediates recognition (inferotemporal areas). The primary visual cortex may therefore act as an "active blackboard," integrating computations made by higher-order cortical areas. However, this model did not explicitly address the role of spatial frequencies in visual processing. Bar et al. (2006) investigated the neural correlates and time-course of spatial frequency processing during recognition of objects in humans using drawings of objects as stimuli. Using MEG, they showed firstly that recognition of non-filtered objects elicited activation in the orbitofrontal cortex before activation of the fusiform gyrus in the inferotemporal cortex. They also demonstrated strong synchrony between the orbitofrontal and temporal cortices during object recognition, suggesting important functional interactions between these regions. Using fMRI, they went on to investigate whether activity in the orbitofrontal cortex was driven by a particular spatial frequency band. They presented drawings of objects filtered in either LSF or HSF, and observed activation of the orbitofrontal cortex for LSF stimuli, but not for HSF stimuli. Overall, their results suggest that the early activation of orbitofrontal cortex observed in MEG was in fact driven by the LSF content. The authors hypothesized a top-down facilitation mechanism, postulating that LSF information is projected rapidly, possibly via the magnocellular pathway of the dorsal cortical stream, from early visual areas to the orbitofrontal cortex, where it activates plausible interpretations of the visual input. Results of these computations are, according to their hypothesis, then projected to the fusiform gyrus, in the inferotemporal cortex, where object recognition is achieved. Kveraga, Boshyan, and Bar (2007) subsequently used dynamic causal modeling (DCM) in an fMRI study to investigate the interaction between the cortical structures involved in object recognition. They used

drawings of objects as stimuli that were either preferentially processed by the magnocellular pathway (achromatic and low-luminance contrast drawings) or the parvocellular pathway (chromatically defined and isoluminant drawings). The study showed that magnocellular-biased stimuli (compared to parvocellular-biased stimuli) elicited strong activation in the orbitofrontal cortex, whereas parvocellular-biased stimuli induced stronger activation in the fusiform gyrus. They also showed that the processing of magnocellular-biased stimuli resulted in an increase in connectivity strength from the occipital cortex to the orbitofrontal cortex and from the orbitofrontal cortex to the fusiform gyrus, while the processing of parvocellular-biased stimuli resulted in increasing connectivity strength from the occipital cortex to the fusiform cortex. These results provided additional support for the existence of a top-down facilitation mechanism (Bar et al., 2006) that may be triggered by magnocellular information, such as LSF, projected early and rapidly to the orbitofrontal cortex. Overall, data from neurophysiological recordings in primates (Bullier, 2001) and object recognition in humans (Bar et al., 2006; Bar, 2007; Kveraga et al., 2007) have offered useful insights into the neural bases of a predominantly coarse-to-fine default processing of spatial frequencies. Critically, although these studies assumed that a coarse-to-fine processing occurs during visual perception, the order of spatial frequency presentation was not manipulated.

Peyrin et al. (2010) combined fMRI and ERP in the same participants and directly investigated the neural bases of coarse-to-fine processing of scenes. They used sequences of two spatial frequency-filtered scenes in rapid succession as stimuli, with either a coarse-to-fine sequence (LSF scene followed by an HSF scene), or a fine-to-coarse sequence (HSF scene followed by an LSF scene) in order to experimentally "mimic" the sequential processing of spatial frequencies and impose a "coarse-to-fine" versus "fine-to-coarse" processing of scenes. Participants had to decide whether the two scenes belonged to the same category. FMRI results showed firstly that coarse-to-fine sequences elicited greater activation of the prefrontal cortex (in the inferior frontal gyrus), the temporo-parietal areas, and the occipital gyrus compared to fine-to-coarse sequences. ERPs were used to provide the time course of activation in these regions. Critically, the inferior frontal gyrus was activated twice during the coarse-to-fine sequences: during processing of the first image in LSF, and then at the onset of the second image in HSF. Furthermore, specific activation of the occipital cortex for coarse-to-fine sequences was observed during processing of HSF scenes. Importantly, activation of the occipital cortex was only observed when HSF scenes were preceded by LSF scenes. These results therefore added support to Bar et al. (2006; 2007), and suggest that LSF information may activate the prefrontal cortex rapidly, in order to generate predictions about the visual category of the scene, and trigger top-down influences that subsequently constrain processing of the HSF scene. Furthermore, based on the integrated model of visual processing (Bullier, 2001), activation in the occipital cortex was interpreted as corresponding to the cortical site on which higher order areas can exert modulatory influences through feedback signals to guide further processing of parvocellular information (HSF) into higher order areas of the ventral visual stream (fusiform gyrus). However, Peyrin et al. (2010) did not investigate the interaction between the cortical structures and the causal influence (effective connectivity) of the high-order areas (either belonging to the dorsal or ventral streams) on the occipital cortex.

The present fMRI study aimed to further explore the neural bases of coarse-to-fine scene analysis and to determine their interactions, paying particular attention to the role of the occipital cortex. As in Peyrin et al. (2010), we explicitly manipulated the order of spatial frequency presentation. We

used sequences, composed of six filtered versions of a scene assembled from LSF to HSF, allowing us to impose a coarse-to-fine processing of scenes. We also used reverse fine-to-coarse sequences, in which the same scenes were assembled from HSF to LSF, as control stimuli. Participants performed a categorization task on these stimuli (indoor vs. outdoor). Stimuli were adapted from previous studies (Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012) which have shown more rapid categorization of coarse-to-fine than fine-to-coarse sequences. These studies provided new arguments in favor of a predominantly coarse-to-fine categorization of natural scenes, and suggested that such sequences are could be appropriate for the investigation of the neural substrates of coarse-to-fine processing.

We began by identifying the cerebral regions specifically involved in coarse-to-fine processing, compared to fine-to-coarse processing. Based on previously-mentioned studies (Bar et al., 2006; Kveraga et al., 2007; Peyrin et al., 2010), we expected that coarse-to-fine sequences would elicit greater activation in three cortical hubs: the occipital cortex, the prefrontal cortex, and the inferotemporal cortex. We then used DCM (Friston, Harrison, & Penny, 2003) to further explore the dynamic interactions between these regions. DCM is a generic Bayesian framework used to infer directed connectivity between a pre-defined set of cortical regions, based on fMRI time series from these regions. More specifically, DCM enables to estimate and make inference about how one neural system influences another (i.e. effective connectivity), and how it can be affected by the experimental context. To our knowledge, only one study has used DCM to investigate the effective connectivity between these three cortical hubs in the theoretical context of a top-down facilitation mechanism for visual recognition (Kveraga et al., 2007). Their DCM analysis did not, however, integrate potential feedback connections from high order areas to the occipital cortex. In the present study, our DCM analysis was driven by results obtained by Kveraga et al. (2007). In addition, we explicitly considered potential feedback connections to the occipital cortex during coarse-to-fine processing.

2. Materials and Methods

Different analyses of the present data set were reported earlier (Musel et al. , 2014)[1]. Data have been fully reanalyzed in order to investigate the effective connectivity between the regions underlying coarse-to-fine processing using dynamic causal modeling.

2.1. Participants

Fourteen right-handed participants (7 males; 23 years ± 2) with no history of neurological disorders were included in this experiment. Participants had normal visual acuity or visual acuity corrected-to-normal using contact lenses. All participants gave their informed written consent before participating in the study, which was in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.2. Stimuli and Procedure

Stimuli were built from 40 black and white photographs (256-level grey-scales, 1024 x 768 pixels) of man-made scenes classified into two distinct categories (20 indoor scenes and 20 outdoor scenes) with a visual angle of 24 × 18 degrees. Exemplar from both categories were chosen to ensure similarity of amplitude spectrum and to prevent categorization from being based on this type of visual cue (Guyader et al., 2004), and to avoid contrast energy differences between categories that could interfere with the sequence of spatial frequency processing. In both categories, images had the same distribution of energy in spatial frequencies and dominant orientations (as shown by the mean amplitude spectrum of non-filtered natural scenes in each category; Figure 1). Furthermore, outdoor and indoor categories were equivalent in terms of visual cluttering

(Subband Entropy measures; see Rosenholtz, Li, & Nakano, 2007). The mean subband entropy was equivalent for outdoors and indoors (2.95±0.16 and 2.95±0.14, respectively; $F_{1,38} < 1$).

Stimuli were created using the MATLAB image processing toolbox (Mathworks Inc., Sherborn, MA, USA). Each scene was filtered by six band-pass filters with different central spatial frequencies: 24, 34, 49, 71, 101, 144 cycles/image and a standard deviation of 25.6 cycles/image (or 1, 1.42, 2.04, 2.96, 4.21, 6 cycles/degree and a standard deviation of 1.07 cycles/degree). The central frequencies of filters followed a logarithmic scale. This enabled us to obtain a better sample of the amplitude spectrum of natural scene, in which the energy decreases as frequency increases ("1/f" shape; Field, 1987) and a better sampling of low spatial frequencies with more filters centered on low spatial frequencies (for a similar approach, see Willenbockel et al., 2010). The cut off frequencies at 67% of the height of each Gaussian were, therefore, [7 47] [11 58] [26 72] [47 93] [78 124] [121 167] cycles/image (i.e. [0.29 1.96] [0.46 2.42] [1.08 3] [1.96 3.88] [3.25 5.17] [5.04 6.97] cycles/degree). Filtered scenes were then normalized to obtain a mean luminance 0.5 (for luminance values of between 0 and 1; i.e. mean luminance of 128 on a 256 grey-level scale) with a standard deviation of 0.1 (25.5 on a 256 grey-level scale). This ensured that all filtered images had equivalent contrast level and differed only by their spatial frequency content. For each image, the resulting six filtered scenes were then assembled in order to create short movies, from LSF to HSF (CtF sequence) or from HSF to LSF (FtC sequence). We used a backward mask built with a 1/f noise to prevent retinal persistence of the stimuli. The mask also had a mean luminance of 128 on a grey-level scale with a standard deviation of 25.5, in order to match the mean luminance and RMS contrast of the filtered scenes composing the sequences. Each movie lasted 150 ms (i.e. six filtered versions of the same scene presented for 25 ms each). Stimuli were displayed using E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, USA) and back-projected on a translucent screen positioned at the rear of the magnet using a video projector with a refresh rate of 80 Hz. Participants viewed this screen at a distance of about 222 cm via a mirror fixed on the head coil. We used a backward mask, built with 1/f noise, to prevent retinal persistence of the scene. It should be noted that stimuli and task were identical to that of Musel et al. (2014)[1].

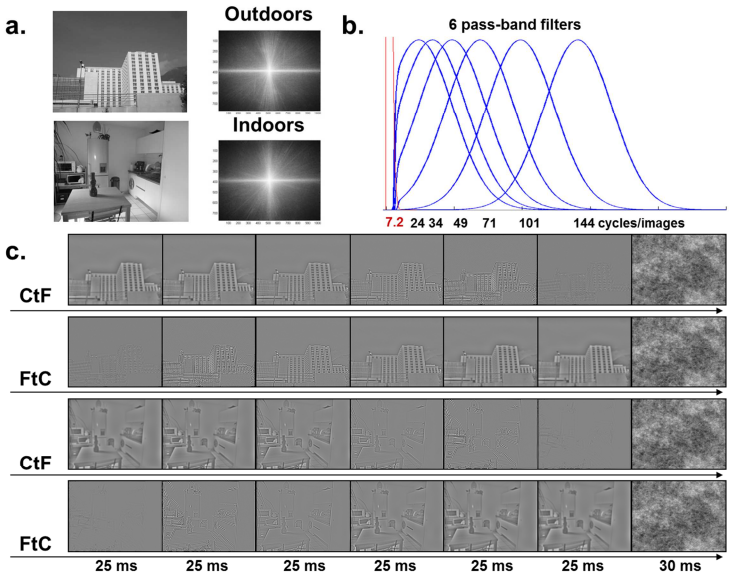


Figure 1. : (a) Example of scenes belonging to different categories (outdoors and indoors). Mean amplitude spectra of each category. On each amplitude spectrum, low spatial frequencies are close to the center, and high spatial frequencies are on the periphery. Vertical orientations are represented on the x-axis and horizontal orientations on the y-axis. (b) Example of six spatial frequency filtered images of scenes depicting (c) the Coarse-to-Fine (CtF) and Fine-to-Coarse (FtC) sequences. **It should be noted that the perception of spatial frequencies could be affected by the picture reduction of scenes for illustrative purpose. The size of images in the Figure is about 10 times smaller than the actual picture size on the screen used for the experiment. The images presented in the Figure have thus been slightly modified in order to make them more visible in the Figure.**

Each sequence stimulus was displayed for 150 ms followed by a mask for 30 ms and a fixation dot in the center of the screen. The average interval between the onsets of two successive stimuli was 2.5 s. A block-design paradigm was used with CtF and FtC movies and the whole experiment consisted of four functional runs. A block was composed of 10 trials (five indoor scenes and five outdoor scenes). There were 12 blocks within a run: four blocks of CtF sequences and four blocks of FtC sequences, interspersed with four 25-sec blocks with a fixation dot on the center of the screen (Fixation condition). It resulted in 80 scenes in each run. The 40 scene images (20 indoors and 20 outdoors) were presented twice within a run, once in a CtF sequence and once in a FtC sequence. The four runs included the same stimuli, but the order of blocks within each run and the order of trials within each block were randomized. It should be noted that a block-design paradigm did not allow us to analyze individual response to trials as an event-related paradigm allowed, and it did not allow us to investigate the neural correlates of priming effect in our study. However, a block-design paradigm increases the statistical power (Friston, Zarahn, Josephs, Henson, & Dale, 1999) and the relatively large BOLD signal change relative to baseline (Glover, 1999). The robustness of activations is particularly important for the present study in which we used two experimental conditions involving the exact same images for the same total exposure duration but differing only by their relative temporal order. Additionally, we ensured at the end of the experiment that participants did not realize that the order of spatial frequency content within a sequence was manipulated. Therefore, participants could not have attended a particular spatial frequency band within a sequence to perform the task.

Participants had to give a categorical answer on sequences ("indoors" or

"outdoors") by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. They were instructed to fixate the center of the screen (fixation dot) during the whole run and to respond as quickly and as accurately as possible by pressing one of two response buttons. Half the participants had to answer "indoor" with the forefinger and "outdoor" with the middle finger, while the other half had to answer 'indoor" with the middle finger and "outdoor" with the forefinger. Response accuracy (ACC) and reaction times (RT, in milliseconds) were recorded.

2.3. fMRI Acquisition

Experiments were performed on a whole-body 3T Philips scanner (Achieva 3.0T TX Philips – Grenoble MRI facility IRMaGE, France). For all functional scans, the manufacturer-provided gradient echo/T2* weighted EPI method was used. Thirty-nine adjacent axial slices parallel to the bicommissural plane were acquired in interleaved mode. Slice thickness was 3.5 mm. The inplane voxel size was 2.7×2.7 mm (216×216 mm field of view acquired with a 80×80 pixel data matrix; reconstructed with zero filling to 128×128 pixels). The main sequence parameters were: TR = 2.5 s, TE = 30 ms, flip angle = 77°. To correct images for geometric distortions induced by local B0-inhomogeneity, a B0 field map was obtained from two gradient echo data sets acquired with a standard 3D FLASH sequence (ΔTE = 9.1 ms). The field map was subsequently used during data processing. Finally, a T1-weighted high-resolution three dimensional anatomical volume was acquired, by using a 3D T1 TFE sequence (field of view = 256×224×176 mm; resolution: 1.333×1.69×1.375 mm; acquisition matrix: 192×132×128 pixels; reconstruction matrix: 288×288×128 pixels).

2.4. FMRI data analysis

Data analysis was performed using the general linear model (Friston 1999) for block designs in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK, http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB. Individual scans were realigned, time corrected, normalized to the Montreal Neurological Institute (MNI) space and spatially smoothed by an 8-mm FWHM (full width at half maximum) Gaussian kernel. Time series for each voxel were high-pass filtered (1/128 Hz cutoff) to remove low-frequency noise and signal drifts. Three conditions of interest (CtF, FtC, Fixation) were modeled as three regressors constructed as box-car functions convolved with a canonical hemodynamic response function. RTs for each trial and movement parameters derived from realignment correction were entered as additional regressors of no interest in the design matrix, to account for RT- and motion-related variance, respectively. Analyses were performed at the individual subject level to examine the contrast between the two conditions ([CtF > FtC] and [FtC > CtF]). These contrast images were then entered into second-level random effect analyses to test for within group effects (one-sample *t* tests). We identified the cerebral regions specifically activated by the CtF and FtC conditions by conducting a whole brain analysis. It should be noted that the CtF and FtC experimental conditions involved the exact same images for the same total exposure duration (150 ms) and differed only by their relative temporal order. These conditions were thus likely to induce subtle differences and a less restrictive threshold of $p < 0.001$ uncorrected for multiple comparisons ($T > 7.97$, $k = 5$) was therefore used to identify the activations that differed between CtF and FtC sequences ([CtF > FtC] and [FtC > CtF] contrasts; see Peyrin et al., 2010 for a similar rationale). A transformation of MNI into Talairach and Tournoux (1988) coordinates was performed using the MNi2TAL function (created by Matthew Brett, available at www.mrc-cbu.cam.ac.uk/Imaging), in order to facilitate comparison with other studies.

2.5. Dynamic Causal modeling analysis

DCM is a generic Bayesian framework used to infer directed connectivity between a pre-defined set of cortical regions, based on fMRI time series from these regions (Friston et al., 2003; Penny, Stephan, Mechelli, & Friston, 2004). More specifically, DCM enables to estimate and make inference about how one neural system influences another, and how it can be affected by the experimental context. This connectivity analysis method is a validated tool which has been widely used in the past ten years, and is suited for fMRI studies using blocked designs (e.g., Kawabata Duncan et al., 2013; Torrisi, Lieberman, Bookheimer, & Altshuler, 2013; Vossel, Weidner, Driver, Friston, & Fink, 2012). We used the contrast [CtF > FtC] to define regions that were more strongly activated during coarse-to-fine compared to fine-to-coarse processing. These regions were used as regions of interest (ROI) from which we extracted time series for each run. We then defined alternative models with these regions and a set of connections, in order to test our specific hypotheses. The different parameters of these models are estimated at the neuronal level using a hemodynamic forward model, in order to compare the predicted functional responses (i.e. predicted with the defined models) to the measured responses (i.e. extracted time series). For each specified model, three sets of parameters are estimated: (i) driving input parameters that reflect how cortical regions respond to external stimuli (e.g., sensorial stimulation), (ii) endogenous connection parameters that reflect the baseline effective connectivity between regions (i.e. what influence one region exerts over another), and (iii) modulatory parameters that quantify how effective connectivity is influenced by the experimental context (i.e. experimental conditions). These parameters are expressed in Hertz (Hz). For endogenous connections, they reflect how rate of change in activity of one target region is influenced by increase in activity in the source region. Positive (negative) values indicate that as rate of change in activity in one region increases, activity in the target region increases (decreases). For modulatory connections, estimated parameter values are simply added to endogenous parameters and reflect context-induced change in connectivity.

The model evidence, i.e. the probability of observing the measured data given a particular model, was then computed. The model evidence is approximated by the negative variational Free-energy (Penny, 2012; Stephan, Penny, Daunizeau, Moran, & Friston, 2009). This approximation is supposed to point to the optimal compromise between the accuracy and the complexity of a given model. Alternative models were thus specified and compared based on their model evidence using the Bayesian model selection (BMS) procedure to identify which model(s) best predict(s) the data (i.e. likelihood that a given model generated the observed data). The relative superiority of one model, compared to all other models, is given by the exceedance probability, i.e. the probability that this model is more likely than any other model, given the data.

We then applied the Bayesian model averaging (BMA) procedure to obtain a single model with new connectivity parameters which correspond to the average connectivity of all models, but where each parameter used to calculate the average connectivity (for endogenous and modulatory parameters) is weighted by its model evidence. Therefore, models with high evidence contribute more than models with lower evidence. Finally, inference on parameters was made, to assess the significance of endogenous and modulatory parameters (complete details on DCM methodology can be found in Seghier, Zeidman, Neufeld, Leff, & Price, 2010; Stephan, Moran, Den Ouden, Daunizeau, & Friston, 2010; Stephan et al., 2009).

3. Results

3.1. Behavioral Results

Two 2 × 2 variance analyses (ANOVA) with Sequences (CtF and FtC) and Categories (outdoor and indoor) as

within-subjects factors were conducted on mean error rates (mER) and mean correct reaction times (mRTs). The ANOVA conducted on mER showed no effect of Sequence (Mean ± SD; CtF: 4.02 ± 4.91 %; FtC: 6.03 ± 8.84 %; $F_{1,13} = 1.69$, $p = 0.22$), but revealed a main effect of Categories ($F_{1,13} = 7.48$, $p < 0.05$). Participants made more errors when categorizing indoor (5.89 ± 7.67 %) than outdoor scenes (4.42 ± 6.63 %). No interaction between Sequence and Categories was observed ($F_{1,13} < 1$). The ANOVA conducted on mRTs revealed that participants categorized CtF sequences faster than FtC sequences (Mean ± SD; CtF: 642 ± 136 ms; FtC: 658 ± 138 ms; $F_{1,13} = 5.31$, $p < 0.05$). There was no effect of categories (outdoor: 653 ± 133 ms; indoor: 657 ± 142 ms; $F_{1,13} < 1$) and no interaction between Sequence and Categories ($F_{1,13} < 1$).

3.2. fMRI Results

The whole-brain analysis allowed us to identify the cerebral regions specifically involved in the processing of spatial frequency sequences (CtF and FtC) which are summarized in Table 1 and Figure 2. CtF sequences elicited stronger activation than FtC sequences ([CtF > FtC] contrast) in occipito-temporal areas involving bilaterally the cuneus, the lingual gyrus and the fusiform/parahippocampal gyrus, and in frontal areas such as the bilateral inferior frontal gyrus (pars orbitalis) and the right superior frontal gyrus, at a statistical threshold of $p < 0.001$ uncorrected for multiple comparisons. No significant activation cluster was observed with the reverse contrast ([FtC > CtF]).

Table 1: Cerebral regions activated by contrasting coarse-to-fine and fine-to-coarse sequences. For each cluster, the region showing the maximum t value is listed first, followed by the other regions in the cluster [in brackets]. Talairach coordinates (x, y, z) are indicated. Abbreviations: R = right hemisphere; L = left hemisphere; BA = Brodmann area; k = number of voxels in the cluster.

Contrast	Region	Side	BA	k	x	y	z	t
[CtF > FtC]	cuneus	R	17/18	614	12	-90	17	7.78
	[lingual gyrus]	L	18		-18	-71	-9	6.79
	[lingual gyrus]	R	18		12	-77	1	6.22
	[cuneus]	L	18		-9	-90	17	5.15
	[fusiform / parahippocampal gyrus]	L	19		-30	-48	-7	4.89
	fusiform gyrus	R	19	64	30	-45	-7	5.58
	inferior frontal gyrus	R	38	15	39	10	-13	5.19
	inferior frontal gyrus	L	47	6	-36	16	-10	4.76
	superior frontal gyrus	R	9	17	15	50	32	4.73
[FtC > CtF]	*No significant cluster*							

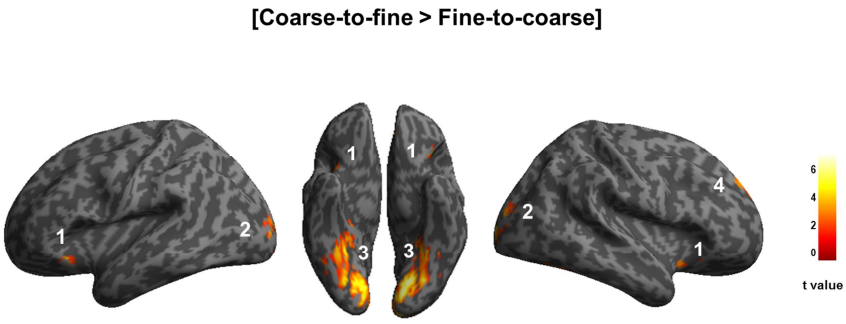


Figure 2: Cerebral regions activated by contrasting coarse-to-fine to fine-to-coarse sequences: (1) cuneus, (2) fusiform and parahippocampal gyri, (3) inferior frontal gyrus, and (4) right superior frontal gyrus.

3.3. DCM results

3.3.1. Time-series extraction

Our DCM analysis of effective connectivity focused on three cortical regions previously identified in whole brain analysis as being more strongly activated by coarse-to-fine categorization ([CtF > FtC] contrast): the cuneus (C), the fusiform gyrus (FG), and the inferior frontal gyrus (IFG). Similar areas were modeled in the DCM analysis by Kveraga et al (2007). Time series were concatenated across the four runs. For each participant, and each region of interest, we first created a sphere (16 mm radius) centered on the peak of activation at group level. We localized subject-specific maxima inside this sphere. We then extracted principle eigenvariates (adjusted for the participant's effect of interest) from 6-mm spheres centered on subject-specific maxima given that these subject-specific maxima would fall below an uncorrected threshold of p = .05. It should be noted that although our results revealed bilateral activations for the three ROIs at the group level, individual time series were only extracted from regions in one hemisphere, in order to maintain our models as simple as possible, because DCM is less effective as the number of regions in the models increases (Seghier et al., 2010; Stephan et al., 2010; see also for example Mechelli, Price, Noppeney, & Friston, 2003; Parker Jones et al., 2013 for similar rationale). We thus chose to extract time series from regions in the right hemisphere, as done in Kveraga et al. (2007)'s DCM analysis, and because the right hemisphere has been shown to be predominantly involved in coarse-to-fine processing (Peyrin et al., 2005)[2].

3.3.2. Construction of model space

We began by defining a base model composed of three hubs (C, FG, and IFG) from which alternative models were derived. CtF and FtC conditions were grouped into a single driving input (i.e. external visual stimulation) that entered the system at C. We specified bidirectional endogenous connections between all regions. Importantly, in DCM analysis, effective connectivity between regions does not necessarily imply anatomical connections. It can also result from polysynaptic connections between regions that do not all need to be included in the models. We then specified modulated connections. In order to test our specific hypotheses**,** we restricted the parameters of our models as much as possible, and three modulated connections were specified a priori based on DCM results from Kveraga et al. (2007) as follows: CtF modulated connectivity from C to IFG and from IFG to FG. FtC modulated connectivity from C to FG. We then elaborated this base model in accordance with our research objectives. The aim of the present study was to identify the dynamic interaction that best accounted for enhanced

activity in C during coarse-to-fine compared to fine-to-coarse processing. Use of DCM makes it possible to ascertain whether increase in activity in one region (i) reflects the consequence, via endogenous connectivity, of a context-related increase in activity in a related area, (ii) stems from the modulation of afferent connection(s) by the experimental context (i.e. experimental conditions), or (iii) reflects a modulation in its self-connectivity (i.e. changes in intra-regional inhibition; Stephan et al., 2010). Our base model was thus elaborated to distinguish between these three explanations, and it resulted in five alternative models. In Model 1, CtF modulated no afferent connection to C (i.e., base model). In Model 2, CtF modulated the feedback connection from IFG. In Model 3, CtF modulated the feedback connection from FG. In model 4, CtF modulated the feedback connection from both IFG and FG. In Model 5, CtF modulated C self-connectivity. Figure 3 shows the five model variants. These five models were specified, estimated, and compared for each participant.

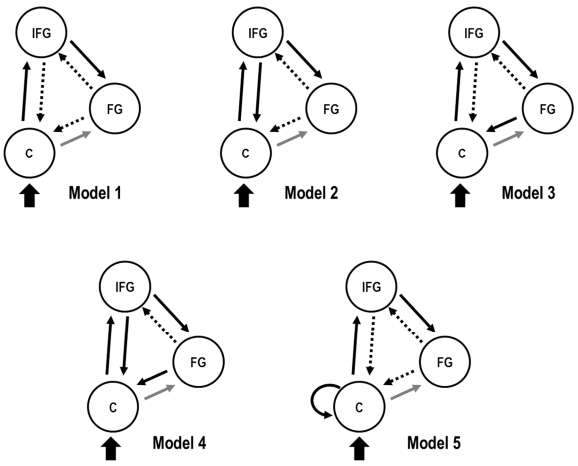


Figure 3: Illustration of the model space. Dotted arrows represent non-modulated endogenous connections, black arrows represent connections modulated by CtF, and grey arrows represent the connection modulated by FtC. External visual input entered C and is represented by a solid black arrow.

3.3.3. Model comparison

Model comparison was implemented using random-effects BMS in SPM8, using the DCM tool (DCM10 version) to compute exceedance and posterior probabilities (i.e. likelihood of a model given the data) at group level (Stephan et al., 2009). It should be noted that exceedance probabilities are conditional to the size of model space, and all exceedance probabilities sum to 1 over all models tested. An exceedance probability of 0.9 for a particular model therefore indicates that this model is 90% more likely to be better than any other model tested given the data. BMS revealed that Model 4 outperformed all other models with an exceedance probability of 0.77. This model is characterized by both IFG to C and FG to C connectivity modulated by CtF processing. The second best model was Model 5, with an exceedance probability of 0.14. The remaining models had exceedance probabilities below 0.05 (see Figure 4a).

Finally, BMA was applied over the model space to obtain a single model with average connectivity parameters, which are weighted by the evidence of their own models (i.e. by the probability of observing the measured data given the model) for endogenous and modulatory parameters. As a reminder, endogenous connection parameters reflect the baseline effective connectivity between regions (i.e. what influence one region exerts over another), and modulatory parameters quantify how effective connectivity is influenced by the experimental context (i.e. experimental conditions). We performed inference on parameters within the Bayesian framework[3], in line with the previous inference on model space (see Desseilles et al., 2011; Richardson, Seghier, Leff, Thomas, & Price, 2011; Seghier, Josse, Leff, & Price, 2011 for a similar procedure). BMA generates posterior distributions of model parameters with 10,000 Gibbs samples. The significance of each connectivity parameter can be assessed by the proportion of samples from the posterior distribution that differ from zero. Significant effects are reported at a

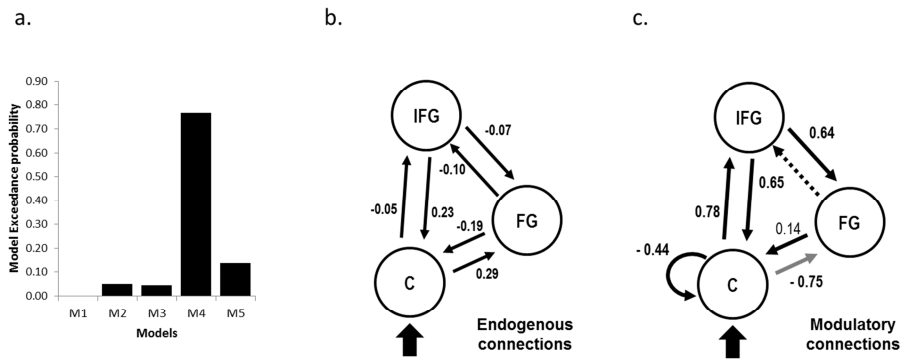


Figure 4: (a) Exceedance probabilities of the five models tested. (b) BMA parameter analysis results for endogenous connections, expressed in Hz. (c) BMA parameter analysis results from modulatory connections (expressed in Hz). Solid black arrows indicate connections modulated by CtF and the grey arrow indicates connections modulated by FtC. The dotted arrow represents non-modulated connections. Parameters significant at P > 0.95 are indicated in bold.

posterior probability threshold of p = 0.95 (i.e. probability of 95 % that a given parameter is different from zero).

BMA was applied to these five models to obtain a single model with average connectivity parameters (weighted by the evidence of their models) for endogenous and modulatory parameters. Results are summarized in Table 3 and Figure 4b and 4c. All endogenous connections were significant, with a posterior probability of p > 0.95 of being different from zero. The endogenous (i.e. baseline) connectivity parameters from C to FG (0.29 Hz) and from IFG to C (0.23 Hz) were positive, suggesting that an increase in C activity results in an increase in FG activity and an increase activity in IFG results in an increase in C activity. Conversely, connectivity parameters from C to IFG (-0.05 Hz), FG to C (-0.19 Hz), FG to IFG (-0.10), and IFG to FG (-0.07 Hz) were negative, suggesting that an increase in activity in all the source regions results in a decrease in activity in the target regions.

Regarding now modulatory parameters related to the CtF condition, we observed significant positive modulations for connectivity from C to IFG (0.78 Hz), IFG to C (0.65 Hz), and IFG to FG (0.65 Hz), and a significant negative modulation in C self-connectivity (-0.44 Hz). This result suggests that coarse-to-fine processing influenced the effective connectivity between these regions. The modulation of the FG to C connectivity was not significant (p = 0.88). Regarding modulatory parameters induced by the FtC condition, we observed significant negative modulations from C to FG (0.75 Hz) suggesting an influence of fine-to-coarse processing on C to FG connectivity. Therefore, once modulatory parameters are added up to endogenous parameters, the results suggest that the processing of CtF sequences results in an increase in the positive connectivity from IFG to C, an increase in the connectivity from C to IFG and IFG to FG that became positive, and in an increase in C negative self-connectivity. The processing of FtC sequences resulted in an increased in the negative connectivity from C to FG.

Table 3: Mean and standard deviation of endogenous and modulatory parameters for all connections across all subjects and models and their posterior probabilities.

Connection		**Mean**	***Standard deviation***	**Posterior Probability**
Endogenous parameters				
from	*to*			
C	FG	0.29	*0.13*	**1.00**
C	IFG	-0.05	*0.18*	**0.98**
FG	C	-0.19	*0.29*	**1.00**
FG	IFG	-0.10	*0.18*	**1.00**
IFG	C	0.23	*0.18*	**1.00**
IFG	FG	-0.07	*0.19*	**0.97**
Modulatory parameters (CtF)				
from	to			
C	C	-0.44	*0.74*	**0.99**
C	IFG	0.78	*0.71*	**1.00**
FG	C	0.14	*1.77*	0.88
IFG	C	0.65	*0.97*	**0.99**
IFG	FG	0.65	2.20	**0.99**
Modulatory parameters (FtC)				
from	to			
C	FG	-0.75	*0.90*	**1.00**

5. Discussion

The present fMRI study aimed to investigate the neural bases of the coarse-to-fine processing of scenes and to further explore their interaction. In order to do this, we used dynamic scenes depicting either coarse-to-fine or fine-to-coarse processing of scenes as stimuli. This allowed us to experimentally "mimic" the sequential processing of spatial frequencies and impose a "coarse-to-fine" versus "fine-to-coarse" processing sequence of scenes. Participants were asked to perform a categorization task on these stimuli (indoor vs. outdoor). Our behavioral results showed that coarse-to-fine sequences were categorized more rapidly than fine-to-coarse sequences. This result is consistent with a considerable number of behavioural and computational studies that have shown temporal precedence of LSF over HSF processing using sine wave gratings and filtered scenes as stimuli (De Cesarei & Loftus, 2011; Hughes et al., 1996; Kihara & Takeda, 2010; Loftus & Harley, 2005; Mermillod, Guyader, & Chauvin, 2005; Musel et al., 2012; Parker et al., 1992; Schyns & Oliva, 1994). Our behavioral results therefore support models of visual perception which suggest that coarse-to-fine analysis is a predominant and advantageous strategy in rapid scene recognition (Hegdé, 2008; Kauffmann et al., 2014; Schyns & Oliva, 1994).

At the neurobiological level, the contrast between coarse-to-fine and fine-to-coarse sequences [CtF > FtC] allowed us to examine the effect of the spatial frequency processing order, and to identify the cerebral regions more strongly activated when LSF are processed before HSF than in the reverse order. Our fMRI results revealed that coarse-to-fine sequences (compared to fine-to-coarse sequences) elicited stronger bilateral activation in three cortical hubs: the occipital

cortex (C), the inferior frontal gyrus (IFG), and the fusiform and parahippocampal gyri (FG). The involvement of the occipital cortex in visual perception and in spatial frequency processing has already been well-established using sinusoidal gratings (Henriksson et al., 2008; Sasaki et al., 2001) and scenes (Musel et al., 2013). Activation in the left inferior frontal gyrus (peak coordinates x, y, z*:* -36, 16, -10) almost overlaps with the orbitofrontal activation reported by Bar et al. (2006; x, y, z*:* -36, 23, -14) and Kveraga et al. (2007; x, y, z: -30, 28, -11). The orbitofrontal cortex is a polymodal region involved in multiple processes such as reward association learning, decision making, cognitive control and emotion processing. It also responds to visual information (Rolls, Browning, Inoue, & Hernadi, 2005; Rolls & Baylis, 1994; Thorpe, Rolls, & Maddison, 1983). Importantly, the orbitofrontal cortex is also known to be involved in LSF-based categorical inference (Bar et al., 2006; Chaumon, Kveraga, Barrett, & Bar, 2014; Fintzi & Mahon, 2013). It has already been established that the fusiform and parahippocampal gyri are involved in face, object and scene recognition (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998; Grill-Spector, 2003; Haxby et al., 2001; Tanaka, 1996). These regions were also recently shown to be sensitive to low-level properties of stimuli such as spatial frequencies (Goffaux et al., 2010; Kauffmann, Ramanoël, Guyader, Chauvin, & Peyrin, 2015; Musel et al., 2014; Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011). These activations are therefore consistent with Bar and colleagues' model of visual perception (Bar, 2003; 2007; Bar et al., 2006) which postulate that visual analysis begins with the parallel extraction of different elementary attributes at different spatial frequencies, and that, at neurobiological level, LSF content of an image is rapidly projected through magnocellular pathways from the occipital cortex to the orbitofrontal cortex, where it activates plausible interpretations of the visual input. The initial LSF-based computations from the orbitofrontal cortex trigger top-down influences which guide subsequent analysis of HSF information, conveyed more slowly from early visual areas, via parvocellular pathways, to recognition-related inferotemporal areas (e.g., the fusiform gyrus). In that context, greater activation of the IFG during coarse-to-fine relative to fine-to-coarse processing of scenes may reflect its rapid engagement in processing LSF content of the sequences in order to generate predictions about the scene category. During the processing of fine-to-coarse sequences, involvement of this region may have been of a lesser degree, since LSF-based predictions could not be performed rapidly to facilitate scene categorization. Regarding the fusiform and parahippocampal gyri, greater activation of these regions may be explained by the fact that during coarse-to-fine processing, these regions were involved in the bottom-up processing of scene information conveyed through the ventral pathway, but also in integrating top-down influences from the IFG to these processes. This would have not been the case during processing of fine-to-coarse sequences, resulting in less activation of this region. This top-down facilitation hypothesis was directly tested in a DCM study by Kveraga et al. (2007). They showed that magnocellular-biased stimuli increased conduction of information from the occipital cortex to the orbitofrontal cortex and from the orbitofrontal cortex to the fusiform gyrus. Based on their results, we first constrained the models in our DCM analysis by modulating connections from C to IFG and from IFG to FG by our CtF experimental condition. Analysis of parameters revealed that these modulations were significant and indicated that the processing of CtF sequences influenced connectivity between these regions (relative to the baseline connectivity), such that connectivity strength from C to IFG and from IFG to FG was increased in the CtF condition. These results are consistent with previous DCM results obtained by Kveraga

et al. (2007) and therefore add support to the existence of the top-down facilitation mechanism proposed by Bar (2003; 2007; Bar et al., 2006; Kveraga, 2007).

Critically, the aim of the present study was to identify the effective connectivity that best account for enhanced activity in the occipital cortex during coarse-to-fine processing. In particular, activation of the occipital cortex during coarse-to-fine categorization of scenes has been previously observed and interpreted as corresponding to the cortical site in which high-order areas can exert modulatory influences through feedback signals to guide further processing of parvocellular information (Bullier, 2001; Peyrin et al., 2010). Our fMRI results showed stronger activation in the cuneus (C), a gyrus of the occipital cortex, during coarse-to-fine than fine-to-coarse categorization of scenes. This activation could neither be attributed to luminance contrast differences since all images in sequences had an equal level of contrast, nor to signal differences since all sequences contained exactly the same images for the same total exposure duration and differed only in their relative order of presentation. However, it should be noted that a mask (i.e. 1/f noise) followed the sequences. This mask contained both LSF and HSF information. Importantly, the primary visual cortex is known to respond maximally to LSF information than HSF (Movshon, Thompson, & Tolhurst, 1978). Therefore, it is possible that during the processing of coarse-to-fine sequences, the visual cortex was maximally activated twice, the first time during the processing of the LSF scene at the beginning of the sequence, and the second time later, during the presentation of the mask. Conversely, during the processing of fine-to-coarse sequences, these two peak responses would have followed each other in time (i.e. at the end of the sequence and just after during the presentation of the mask), producing only one maximal response. Greater activation of the cuneus for coarse-to-fine than fine-to-coarse sequences could therefore be explained in part by such mechanism. However, Peyrin et al. (2010) used sequences of two filtered scenes going from low-to-high spatial frequencies (coarse-to-fine) or from high-to-low spatial frequencies (fine-to-coarse). As in the present study, both sequences contained the exact same images, for the same total exposure duration and only the presentation order varied between them. Importantly, these authors did not mask these sequences. Yet, fMRI results revealed greater activation of the primary visual cortex in response to coarse-to-fine than fine-to-coarse sequences, similarly to the present results. These results, which cannot be explained by the presence of a mask, therefore suggest that greater activation of the primary visual cortex during CtF relative to FtC processing of scenes in the present study may not be exclusively due to the fact that the visual cortex responded maximally twice during CtF sequences and only once during FtC sequences. We used the BMS procedure implemented in our DCM analysis to ascertain whether stronger activity in the cuneus during coarse-to-fine processing (i) reflects the consequence, via endogenous connectivity, of context-related increase in activity in related areas such as IFG and FG (Model 1), (ii) results from a modulation by coarse-to-fine processing of feedback connections from IFG (Model 2), from FG (Model 3), or from both regions (Model 4), or (iii) reflects a modulation in its self-connectivity (Model 5). Results from BMS revealed that Model 4, in which stronger activation of C resulted from a modulation of its feedback connections from both IFG and FG, was the most plausible model. Analysis on parameters revealed that an increased rate in IFG activity resulted in an increased rate in C activity, and this coupling was greater in the CtF condition. We also observed that an increased in FG activity resulted in a decrease in C activity. This coupling tended to be canceled when considering modulations by the CtF condition, even if this modulatory parameter was not significant. It should however be noted that the probability that this

connection was modulated by CtF was high (88%). It is important to note that since the DCM analysis was performed on regions of the right hemisphere, as did Kveraga et al. (2007), the inferred effective connectivity in our study may primarily apply to the right hemisphere. Besides, to our knowledge, there is no evidence of a symmetric connectivity network between the right and the left hemisphere and previous studies have shown right hemisphere predominance for coarse-to-fine analysis (Musel et al., 2013; Peyrin et al., 2005, 2004, 2003).

Overall, these results provide new evidence supporting the notion of top-down influences from high-level areas to the occipital cortex (Bullier, 2001; Peyrin et al., 2010). Although results from BMS suggested that CtF processing influences connectivity from both frontal and inferotemporal regions to the occipital cortex, analysis on parameters indicates an increase in the connectivity from IFG to C only, suggesting that top-down influences on C during coarse-to-fine processing originate mainly in frontal regions. Top-down influences from high-level areas on the occipital cortex have been demonstrated in a large number of studies, which have shown, for example, modulation of activity in the visual cortex by attention or learning (Martínez et al., 1999; Sasaki et al., 2001; Saygin & Sereno, 2008; Silver, Ress, & Heeger, 2007; Tootell, Hadjikhani, Mendola, Marrett, & Dale, 1998). Top down influence may also originate in frontal, parietal, or temporal areas Top down influence may also originate in frontal, parietal, or temporal areas (e.g., Koivisto, Railo, Revonsuo, Vanni, & Salminen-Vaparanta, 2011; Ruff et al., 2006, 2009). In a recent review, Rauss, Schwartz, and Pourtois (2011) integrated these data into a predictive coding framework. According to these authors, top-down influences mostly reflect predictions, based on learned contingencies in the visual environment and stored knowledge, used to process incoming signals in the visual cortex. Our results therefore fall well within that theoretical framework and support current views of the occipital cortex as an integrative area that is involved in complex functions which are more than merely sensorial (see also Lee, Mumford, Romero, & Lamme, 1998).

These results obviously raise the question of the functional and anatomical connectivity that could subserve interactions between these regions. Again, it is important to emphasize that in DCM analysis, effective connectivity between regions does not necessarily imply anatomical connections. It can also result from polysynaptic connections between regions that are not modeled in the specified models. Firstly, in the case of anatomical connections between the occipital cortex and high-order areas, there is evidence of anatomical connections between the occipital cortex and high-order areas such as the fusiform and parahippocampal gyri (along the inferior longitudinal fasciculus; Catani, Jones, Donato, & Ffytche, 2003; Catani & Thiebaut de Schotten, 2008; Powell et al., 2004). These connections are known to mediate recognition of objects, faces, and scenes (Ishai, Ungerleider, Martrin, Schouten, & Haxby, 1999; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013). As far as connections with the orbitofrontal cortex are concerned, studies on primates and tractography studies in humans (Barbas, 2000; Catani & Thiebaut de Schotten, 2008; Kringelbach & Rolls, 2004; Wakana, Jiang, Zijl, & Mori, 2004) have shown direct ventral anatomical connections between the occipital cortex and the orbitofrontal cortex (along the inferior fronto-occipital fasciculus). Other anatomical projections occur in the dorsal cortex, between the occipital cortex and the parietal cortex, and between the parietal cortex and the orbitofrontal cortex (Cavada et al., 2000; Kravitz, Saleem, Baker, & Mishkin, 2011). Finally, functional connectivity between the occipital and orbitofrontal cortices is supported by neuroimaging studies using DCM analyses that revealed effective connectivity from the occipital areas to the prefrontal cortex during, for example, facial

affect (Dima, Stephan, Roiser, Friston, & Frangou, 2011; Torrisi, Lieberman, Bookheimer, & Altshuler, 2013) and language processing (Yvert, Perrone-Bertolotti, Baciu, & David, 2012).

Regarding anatomical connections between the orbitofrontal cortex and the regions on the inferotemporal cortex, studies on the macaque brain indicate that the orbitofrontal cortex has strong and reciprocal anatomical connections with medial regions of the temporal cortex including parahippocampal areas (Cavada et al., 2000; Ghashghaei & Barbas, 2002). In humans, studies using diffusion tensor MRI have shown anatomical connectivity between the parahippocampal gyrus and both the inferior temporal cortex and orbitofrontal cortex (Powell et al., 2004). Finally, tractography studies in humans have revealed anatomical connections between the orbitofrontal and inferotemporal cortices (Catani & Thiebaut de Schotten, 2008). Important functional interactions have been demonstrated between these regions during object recognition (Bar et al., 2006) and contextual associations (Chaumon et al., 2014).

To conclude, we have demonstrated the involvement of three cortical hubs in the coarse-to-fine categorization of scenes: the inferior frontal gyrus, the fusiform and parahippocampal gyri, and the occipital cortex. Furthermore, effective connectivity analysis performed on these fMRI data revealed that greater activity in the occipital cortex during coarse-to-fine processing is likely to be driven by top-down influences originating in both frontal and inferotemporal areas. Our results suggest that these influences guide the processing of incoming signals in the occipital cortex and support current models of visual perception (Bar, 2003; 2007; Bullier, 2001; Kauffmann et al., 2014; Peyrin et al., 2010). They also enable them to be refined, by emphasizing the role of the occipital cortex as a cortical site for feedback projections in the neural network underlying coarse-to-fine processing of scenes.

Footnotes

[1]Different analyses of the present data set were reported earlier in in order to investigate the spatial frequency processing in scene-selective regions (Musel et al., 2014). In fact, all participants endorsed a dynamic scene experiment followed by a separate functional localizer in order to localize scene selective regions, and a region of interest (ROI) analysis was performed. Results of the localizer experiment and of the ROI analysis were not reported here since they took place in a different theoretical context (i.e. spatial frequency processing within category-specific regions) and were not relevant to the present goals, but can be found in Musel et al. (2014). Data have been fully reanalyzed in order to investigate the effective connectivity between the regions underlying coarse-to-fine processing using dynamic causal modeling, in the theoretical context of proactive models of visual perception. We believe that it would have been confusing to report both goals and thus both analyses in the same paper, since there are directly relevant to different experimental goals and would have distracted from each theme of inquiry. We also strongly believe that it would have unnecessarily lengthened the papers but also made them less focused.

[2] We also ran the DCM analysis on regions of the left hemisphere. However, since the left IFG could not be identified with an uncorrected threshold of $p < 0.05$ in all participants, time series in this region were extracted without thresholding. Results from BMS revealed that model 5 (i.e. modulation of C self-connectivity) was the best model with an exceedance probability of 0.68, followed by Model 4 (modulation by CtF of feedback connectivity from IFG and FG) with an exceedance probability of 0.14 and Model 2 (modulation of IFG to C connectivity) with an exceedance probability of 0.10. Model 1 and 3 had an exceedance probability below 0.10. We then ran the BMA procedure to perform inference on parameters. Regarding endogenous connectivity, only the connectivity parameter of the C to FG connection was significant (0.38 Hz). Regarding modulated connections, none of the modulatory parameters was significant. The discrepancy between these results and those obtained in the right hemisphere could be due to the fact that the right hemisphere has been shown to be predominantly involved in coarse-to-fine processing of scenes, relative to the left hemisphere (Peyrin et al., 2005).

[3] A different approach could have been to submit estimated parameters from each subject to classical frequentist t-tests. In using this particular approach, our results differed slightly from those we obtained in making inferences based on the Bayesian theoretical framework. Modulation by CtF of the IFG to FG connection was not significant ($t_{13} = 1.09$, $p = 0.29$). It should be noted that during the BMA procedure,

the contribution of all connections is considered to generate posterior densities used to infer significance of each connection parameter, whereas inference using t-tests is based only on the individual contribution of connection parameters. This could explain why results from BMA and from t-tests on parameter estimated may not always concord.

Acknowledgment

This work was supported by the RECOR ANR Grant (ANR-12-JHS2-0002-01 RECOR) and a subvention from "Grenoble Pôle Cognition". Louise Kauffmann was supported by Région Rhône-Alpes (Cible Grants). Grenoble MRI facility IRMaGe was partly funded by the French program "Investissement d'Avenir" run by the "Agence Nationale pour la Recherche": Grant "Infrastructure d'Avenir en Biologie Santé" (ANR-11-INBS-0006). We thank Catherine Dal Molin for the English revision of the manuscript.

References

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

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*, 280–9.

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*, 449–454.

Barbas, H. (2000). Connections underlying the synthesis of cognition , memory , and emotion in primate prefrontal cortices. *Brain Research Bulletin*, *52*, 319–330.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research*, *36*, 96–107.

Catani, M., Jones, D. K., Donato, R., & Ffytche, D. H. (2003). Occipito-temporal connections in the human brain. *Brain : A Journal of Neurology*, *126*, 2093–107.

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, *44*, 1105–32.

Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J., & Reinoso-Suárez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cerebral Cortex*, *10*, 220–42.

Chaumon, M., Kveraga, K., Barrett, L. F., & Bar, M. (2014). Visual Predictions in the Orbitofrontal Cortex Rely on Associative Content. *Cerebral Cortex*, 1–9.

De Cesarei, A., & Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychonomic Bulletin & Review*, *18*, 840–7.

De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559.

De Valois, R., Yund, W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 531–544.

Desseilles, M., Schwartz, S., Dang-Vu, T. T., Sterpenich, V., Ansseau, M., Maquet, P., & Phillips, C. (2011). Depression alters "top-down" visual attention: a dynamic causal modeling comparison between depressed and healthy subjects. *NeuroImage*, *54*, 1662–8.

Dima, D., Stephan, K. E., Roiser, J. P., Friston, K. J., & Frangou, S. (2011). Effective connectivity during processing of facial affect: evidence for multiple parallel pathways. *The Journal of Neuroscience*, *31*, 14378–85.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The Parahippocampal Place Area : Recognition , Navigation , or Encoding ? *Neuron*, *23*, 115–125.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 6–9.

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.*, *4*, 2379–94.

Fintzi, A. R., & Mahon, B. Z. (2013). A bimodal tuning curve for spatial frequency across left and right human orbital frontal cortex during object recognition. *Cerebral Cortex*, *24*, 1311–8.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.

Friston, K., Zarahn, E., Josephs, O., Henson, R. N., & Dale, a M. (1999). Stochastic designs in event-related fMRI. *NeuroImage*, *10*, 607–19.

Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience*, *115*, 1261–79.

Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance.* (Ginsburg, ., pp. 1–41). Oxford, England: John Wiley & Sons.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*, 416–29.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2010). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex, In pub ahe*, doi:10.1093/cercor/bhq112.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*, 159–166.

Guyader, N., Chauvin, A., Peyrin, C., Hérault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, *327*, 313–318.

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–30.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*, 405–39.

Henriksson, L., Nurminen, L., & Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *Journal of Vision*, *8*, 1–13.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, *8*, 197– 230.

Hupé, J., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., Bullier, J., … E, J. H. U. P. (2001). Feedback Connections Act on the Early Part of the Responses in Monkey Visual Cortex. *Journal of Physiology*, *85*, 134–145.

Ishai, A., Ungerleider, L., Martrin, A., Schouten, J., & Haxby, J. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*, 9379–9384.

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial Frequency Processing in Scene-Selective Cortical Regions. *NeuroImage*, *112*, 86–95.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 1–14.

Kihara, K., & Takeda, Y. (2010). Time course of the integration of spatial frequency-based information in natural scenes. *Vision Research*, *50*, 2158–62.

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *The Journal of Neuroscience*, *31*, 2488–92.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews. Neuroscience*, *12*, 217–30.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*, 26–49.

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, *72*, 341–72.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of topdown facilitation in recognition. *The Journal of Neuroscience*, *27*, 13232–40.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. a. (1998). The role of the primary visual

cortex in higher level vision. *Vision Research*, *38*, 2429–54.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review*, *12*, 43–65.

Martínez, a, Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., … Hillyard, S. a. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*, 364–9.

Mechelli, A., Price, C. J., Noppeney, U., & Friston, K. J. (2003). A dynamic causal modeling study on category effects: bottom-up or top-down mediation? *Journal of Cognitive Neuroscience*, *15*, 925–34.

Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling. *Brain and Cognition*, *57*, 151–7.

Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Bas, J. Le, & Peyrin, C. (2013). Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. *Journal of Cognitive Neuroscience*, *25*, 1315–1331.

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PloS One*, *7*, e38493.

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of Cognitive Neuroscience*, *26*, 1–11.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–60.

Parker Jones, O., Seghier, M. L., Kawabata Duncan, K. J., Leff, A. P., Green, D. W., & Price, C. J. (2013). Auditory-motor interactions for the production of native and non-native speech. *The Journal of Neuroscience*, *33*, 2376–87.

Penny, W. D. (2012). Comparing dynamic causal models using AIC, BIC and free energy. *NeuroImage*, *59*, 319–30.

Penny, W. D., Stephan, K. E., Mechelli, a, & Friston, K. J. (2004). Comparing dynamic causal models. *NeuroImage*, *22*, 1157–72.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *NeuroImage*, *23*, 698–707.

Peyrin, C., Chauvin, A., Chokron, S., & Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain and Cognition*, *53*, 278–282.

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*, 2768–80.

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., & Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *NeuroImage*, *28*, 464–73.

Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative Ophthalmology*, *11*, 368–77.

Powell, H. W. R., Guye, M., Parker, G. J. M., Symms, M. R., Boulby, P., Koepp, M. J., … Duncan, J. S. (2004). Noninvasive in vivo demonstration of the connections of the human parahippocampal gyrus. *NeuroImage*, *22*, 740–7.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology*, *9*, e1000608.

Rauss, K., Schwartz, S., & Pourtois, G. (2011). Top-down effects on early visual processing in humans: a predictive coding framework. *Neuroscience and Biobehavioral Reviews*, *35*, 1237–53.

Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S. C., & Price, C. J. (2011). Multiple routes from occipital to temporal cortices during reading. *The Journal of Neuroscience*, *31*, 8239–47.

Rolls, E., Browning, A., Inoue, K., & Hernadi, I. (2005). Novel visual stimuli activate a population of neurons in the primate orbitofrontal cortex. *Neurobiology of Learning and Memory*, *84*, 111–23.

Rolls, T., & Baylis, L. (1994). Gustatory , Olfactory , Orbitofrontal Cortex and Visual Convergence within the Primate. *The Journal of Neuroscience*, *14*, 5437–5452.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*, 1–22.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., … Driver, J. (2006). Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. *Current Biology*, *16*, 1479–88.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Weiskopf, N., & Driver, J. (2009). Hemispheric differences in frontal and parietal influences on human occipital cortex: direct confirmation with concurrent TMS-fMRI. *Journal of Cognitive Neuroscience*, *21*, 1146–61.

Sasaki, Y., Hadjikhani, N., Fischl, B., Liu, a K., Marrett, S., Dale, a M., … Marret, S. (2001). Local and global attention are mapped retinotopically in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 2077–82.

Saygin, A. P., & Sereno, M. I. (2008). Retinotopy and attention in human occipital, temporal, parietal, and frontal cortex. *Cerebral Cortex*, *18*, 2158–68.

Schyns, P. G., & Oliva, A. (1994). Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*, 195–201.

Seghier, M. L., Josse, G., Leff, A. P., & Price, C. J. (2011). Lateralization is predicted by reduced coupling from the left to right prefrontal cortex during semantic decisions on written words. *Cerebral Cortex*, *21*, 1519–31.

Seghier, M. L., Zeidman, P., Neufeld, N. H., Leff, A. P., & Price, C. J. (2010). Identifying abnormal connectivity in patients using dynamic causal modeling of FMRI responses. *Frontiers in Systems Neuroscience*, *4*, 1–14.

Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, *42*, 2547–54.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural Correlates of Sustained Spatial Attention in Human Early Visual Cortex. *Journal of Neurophysiology*, *97*, 229–237.

Stephan, K. E., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, *49*, 3099–109.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*, 1004–17.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–39.

Thorpe, S. J., Rolls, E. T., & Maddison, S. (1983). The Orbitofrontal Cortex : Neuronal Activity in the Behaving Monkey. *Experimental Brain Research*, *49*, 93–115.

Tootell, R. B., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, a M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in Cognitive Sciences*, *2*, 174–83.

Torrisi, S. J., Lieberman, M. D., Bookheimer, S. Y., & Altshuler, L. L. (2013). Advancing understanding of affect labeling with dynamic causal modeling. *NeuroImage*, *82*, 481–8.

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.

Wakana, S., Jiang, H., Zijl, P. C. M. Van, & Mori, S. (2004). Fiber tract-based Atlas of human white matter anatomy. *Radiology*, *230*, 77–87.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., … Gosselin, F. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology. Human Perception and Performance*, *36*, 122–35.

Yvert, G., Perrone-Bertolotti, M., Baciu, M., & David, O. (2012). Dynamic causal modeling of spatiotemporal integration of phonological and semantic processes: an electroencephalographic study. *The Journal of Neuroscience*, *32*, 4297–306.

Les résultats de l'**Expérience 3** sont donc compatibles avec un mécanisme de facilitation « top-down » lors de la catégorisation de scènes. L'information en BFS véhiculée rapidement au cortex orbitofrontal permettrait de générer des prédictions sur la nature possible du stimulus visuel et viendraient guider le traitement des HFS lors de la catégorisation de scènes (Bar, 2003; Bar et al., 2006; Bullier, 2001; Kveraga et al., 2007; Peyrin et al., 2010). En ce sens, l'analyse de l'information visuelle consisterait à mettre en relation en permanence les entrées sensorielles avec des « templates » formés sur la base d'expériences précédentes et sur l'extraction de caractéristiques élémentaires du stimulus. Cela permettrait de limiter l'éventail des interprétations possibles et de faciliter le traitement ultérieur des caractéristiques qui coïncident avec les « templates ». Nous verrons en détail en Discussion générale des propositions concernant les mécanismes par lesquels ces prédictions influenceraient le traitement visuel ultérieur. Mais on remarque dans l'ensemble que la façon dont ces deux types d'informations sont intégrés lors de la catégorisation de scènes reste méconnue car elle a très peu été étudiée.

Dans une étude récente, Mu et Li (2013) ont examiné le décours temporel de l'intégration des BFS et HFS en enregistrant les potentiels évoqués de participants effectuant une tâche de catégorisation de scènes. Ces auteurs ont présenté des scènes hybrides aux participants, qui avaient pour consigne d'ignorer l'information en BFS et de les catégoriser (artificiel *versus* naturel) sur la base de l'information en HFS seulement. Selon l'hypothèse d'une catégorisation de scènes selon un mode de traitement par défaut « coarse-to-fine », les auteurs s'attendaient à ce que l'information en BFS des hybrides soit rapidement traitée (bien qu'elle ne soit pas pertinente pour la tâche), et à ce qu'elle influence le traitement de l'information en HFS. Afin de tester cela, les auteurs ont manipulé la similarité sémantique entre les deux scènes composant les images hybrides (cf. Figure 24). Celles-ci étaient soit sémantiquement similaires (e.g., BFS d'une scène naturelle superposées aux HFS d'une autre scène naturelle), soit sémantiquement dissimilaires (e.g., BFS d'une scène naturelle superposées aux HFS d'une scène artificielle). Dans le cas où l'information sémantique en BFS et en HFS étaient non-congruentes, le traitement rapide des BFS devait alors conduire à faire des prédictions erronées, et donc interférer avec la catégorisation des HFS (i.e. interférence sémantique). Les auteurs ont également testé si l'interférence sémantique des BFS sur les HFS pouvait être modulée par la similarité physique (i.e. recouvrement spatial) entre les deux scènes composant les hybrides. Celles-ci étaient donc soit physiquement similaires, soit physiquement dissimilaires. Les auteurs s'attendaient à ce que l'effet

d'interférence sémantique soit plus important lorsque les scènes composant les hybrides étaient physiquement similaires.

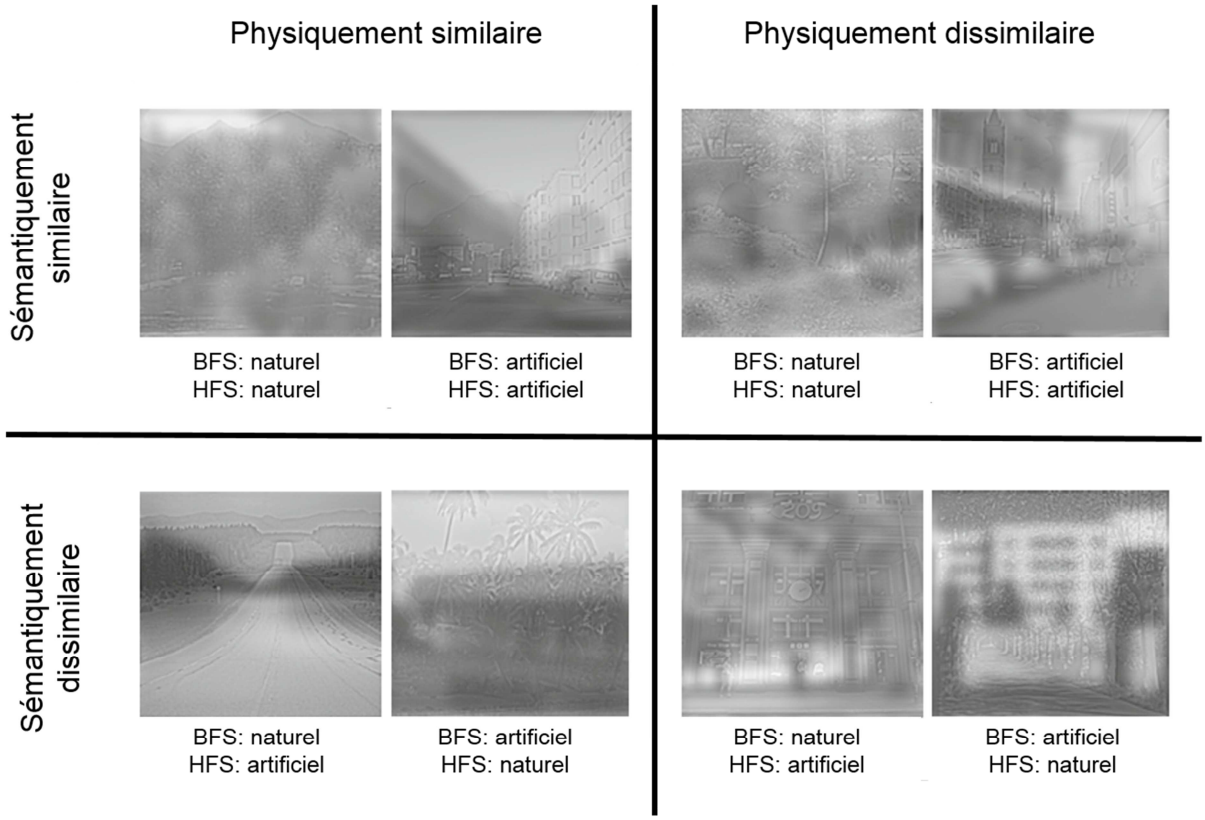


Figure 24 : Exemple de scènes hybrides utilisées dans l'étude de Mu et Li (2013). Les scènes en BFS et en HFS composant les hybrides étaient sémantiquement et physiquement similaires (en haut à gauche), sémantiquement similaires et physiquement dissimilaires (en haut à droite), sémantiquement dissimilaires et physiquement similaires (en bas à gauche) ou sémantiquement en physiquement dissimilaires (en bas à droite). Adapté de Mu et Li (2013).

Comme attendu, les résultats comportementaux ont révélé un taux d'erreur plus important lorsque les deux scènes composant les hybrides étaient sémantiquement dissimilaires que sémantiquement similaires. Cet effet d'interférence sémantique était d'autant plus important que les deux scènes étaient physiquement similaires. Ces résultats suggèrent que l'information en BFS, bien que non pertinente pour la tâche, avait pourtant été traitée au niveau sémantique et interférait avec la catégorisation de l'information en HFS, surtout si les deux scènes partageaient des propriétés physiques similaires. Selon les auteurs, dans ce dernier cas, le recouvrement spatial entre les deux scènes ne permettait pas de les différencier sur la base de leurs propriétés physiques. Les auteurs ont utilisé cet effet d'interférence sémantique comme la signature de l'intégration au niveau sémantique des deux informations en fréquences spatiales (BFS et HFS) contenues dans les scènes hybrides.

Au niveau neurobiologique, l'analyse des potentiels évoqués a révélé que l'effet d'interférence sémantique était associé à une composante négative frontale (N1) plus importante, 122 ms après apparition du stimulus, suivie de composantes positives pariétale et occipitale (P2 et P3), respectivement 247 et 344 ms après apparition du stimulus. Les auteurs ont interprété ces résultats comme reflétant une intégration très précoce des informations sémantiques en BFS et en HFS au niveau des régions frontales, et ce, avant l'implication d'aires pariétales et occipitales. Cependant, l'implication de ces régions reste hypothétique dans la mesure où les auteurs n'ont pas identifié la source de leurs potentiels évoqués. Dans la continuité des travaux de Mu et Li (2013), nous avons donc mené une étude en IRMf (**Expérience 4**) afin de préciser les bases cérébrales de cet effet d'interférence sémantique.

2.3. EXPERIENCE 4 : Etude de l'influence du traitement des BFS sur le traitement des HFS via l'effet d'interférence sémantique

L'**Expérience 4** est présentée en détail dans l'article ci-après, actuellement sous presse dans la revue *Journal of Cognitive Neuroscience*. Cette étude a eu pour but d'examiner si le traitement rapide des BFS dans les scènes est susceptible de moduler, voire d'interférer avec le traitement des HFS, et de préciser grâce à la technique de l'IRMf les bases cérébrales de l'interférence sémantique des BFS sur les HFS. Nous avons pour cela utilisé un paradigme expérimental similaire à celui de Mu et Li (2013). Nous avons donc présenté aux participants des scènes hybrides dont l'information en BFS et en HFS étaient sémantiquement similaire ou dissimilaire, et physiquement similaire ou dissimilaire. Les participants avaient pour consigne d'ignorer l'information en BFS dans les hybrides et de les catégoriser sur la base de l'information en HFS (naturel *versus* artificiel).

Comme attendu, nous avons observé au niveau comportemental des temps de réponse plus longs et un taux d'erreur plus important pour catégoriser les hybrides de la condition sémantiquement dissimilaire que sémantiquement similaire, indiquant que l'information contenue dans les BFS avait été traitée au niveau sémantique et interférait avec la catégorisation de la scène en HFS. Cependant, contrairement aux résultats de Mu et Li (2013), nous n'avons pas observé d'interaction entre la similarité sémantique et la similarité physique des scènes composant les images hybrides.

Au niveau neurobiologique, les résultats ont montré que l'effet d'interférence sémantique était associé à une activité plus importante et bilatérale du gyrus frontal inférieur, du lobule pariétal supérieur et des gyri fusiforme et parahippocampique. Comme pour les

résultats comportementaux, nous n'avons pas observé d'interaction entre la similarité sémantique et la similarité physique des scènes composant les images hybrides. Nous avons également réalisé une analyse de connectivité fonctionnelle (« psycho-physiological interactions », PPI) qui a révélé que l'effet d'interférence sémantique était également associé à une connectivité fonctionnelle plus importante entre le gyrus fusiforme et le gyrus frontal inférieur.

L'effet d'interférence sémantique observé dans notre étude est cohérent avec les résultats de Mu et Li (2013) et semble également compatible avec le modèle proactif de Bar (2003 ; 2007 ; Bar et al., 2006). En effet, dans le cadre de ce modèle, et sur la base du décours temporel des activations rapportées dans l'étude de Mu et Li (2013), nous avons proposé que lors de la catégorisation de scènes l'information en BFS serait rapidement traitée au niveau sémantique par le gyrus frontal inférieur, afin d'activer des interprétations possibles de la catégorie de la scène. Ces prédictions, réalisées sur la base d'une information grossière en BFS, seraient à l'origine d'influences « top-down » vers les régions inférotemporales de la reconnaissance (i.e. gyrus fusiforme et parahippocampique). Ces prédictions pourraient alors interférer avec la catégorisation de l'information en HFS dans le cas où les deux informations seraient sémantiquement dissimilaires. Ce processus de catégorisation impliquerait également le cortex pariétal supérieur afin d'orienter l'allocation de l'attention vers le contenu en fréquences spatiales pertinent pour la tâche.

Notons cependant que le gyrus frontal inférieur est connu pour être également impliqué lors de tâches nécessitant la résolution d'un conflit relatif à la réponse à donner (e.g., tâches de Stroop ou réponses de type « go/no-go » (Nee, Wager, & Jonides, 2007) comme c'était le cas lors de l'interférence sémantique dans notre étude. Plusieurs travaux ont en effet suggéré que cette région puisse intervenir afin de contrôler et minimiser l'influence d'informations interférentes, non pertinentes pour une tâche donnée (Chong, Williams, Cunnington, & Mattingley, 2008; Nelson, Reuter-Lorenz, Persson, Sylvester, & Jonides, 2009; Wais, Rubens, Boccanfuso, & Gazzaley, 2010). En ce sens, l'activation plus importante du gyrus frontal inférieur en cas d'interférence sémantique dans notre étude pourrait également refléter l'implication de cette région dans le pilotage et le contrôle cognitif du traitement des informations pertinentes pour la tâche.

Article sous presse dans la revue *Journal of Cognitive Neuroscience*

The neural bases of the semantic interference of spatial frequency-based information in scenes

Louise Kauffmann[1, 2*], Jessica Bourgin[1,2], Nathalie Guyader[3] and Carole Peyrin[1, 2]

[1] Univ. Grenoble Alpes, LPNC, F-38040 Grenoble

[2] CNRS, LPNC UMR 5105, F-38040 Grenoble

[3] Univ. Grenoble Alpes, GIPSA-lab, F-38402 Grenoble

Abstract

Current models of visual perception suggest that during scene categorization, low spatial frequencies (LSF) are processed rapidly and activate plausible interpretations of visual input. This coarse analysis would then be used to guide subsequent processing of high spatial frequencies (HSF). The present fMRI study examined how processing of LSF may influence that of HSF by investigating the neural bases of the semantic interference effect. We used hybrid scenes as stimuli by combining LSF and HSF from two different scenes and participants had to categorize the HSF scene. Categorization was impaired when LSF and HSF scenes were semantically dissimilar, suggesting that the LSF scene was processed automatically and interfered with categorization of the HSF scene. FMRI results revealed that this semantic interference effect was associated with increased activation in the inferior frontal gyrus, the superior parietal lobules, and the fusiform and parahippocampal gyri. Furthermore, a connectivity analysis (psycho-physiological interaction) revealed that the semantic interference effect resulted in increasing connectivity between the right fusiform and the right inferior frontal gyri. Results support influential models suggesting that during scene categorization, LSF information is processed rapidly in the prefrontal cortex and activates plausible interpretations of the scene category. These coarse predictions would then initiate top-down influences on recognition-related areas of the inferotemporal cortex, and these could interfere with the categorization of HSF information in case of semantic dissimilarity to LSF.

Keywords: Scene categorization; Low spatial frequencies, High spatial frequencies; semantic interference effect; fMRI.

1. Introduction

The human visual system is particularly efficient in processing and categorizing complex stimuli such as scenes (Thorpe et al., 1996). Convergent data on the functional neuroanatomy of visual pathways (Van Essen & Deyoe, 1995), neurophysiological recordings in primates (De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Hupé et al., 2001; Poggio, 1972; Shams & von der Malsburg, 2002; for a review, see Bullier, 2001), and psychophysical results in humans (Ginsburg, 1986; Hughes et al., 1996; Parker et al., 1992; Schyns & Oliva, 1994), support current models of visual perception (Bar, 2003; Bullier, 2001; Hegdé, 2008; Kauffmann et al., 2014; Schyns & Oliva, 1994), which suggest that visual analysis starts with the parallel extraction of different visual elementary features at different spatial frequencies. Low spatial frequencies (LSF) in scenes, conveyed by fast magnocellular pathways, provide coarse information about a visual stimulus (e.g., the global shape and structure of a scene), whereas high spatial frequencies (HSF) in scenes, conveyed more slowly by the parvocellular pathways, provide finer information about the stimulus (e.g., the edges and borders of an object in the scene).

Based on the neurophysiological properties of the magnocellular and parvocellular pathways, a number of studies using sinusoidal gratings (Ginsburg, 1986; Hughes et al., 1996) or scenes filtered in spatial frequency (Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Musel et al., 2012; Parker et al., 1992; Schyns & Oliva, 1994) as stimuli have shown temporal precedence of LSF over HSF processing in the time course of visual processing. Schyns and Oliva (1994), for example, used hybrid images made of two superimposed scenes belonging to different categories and containing different spatial frequency bands (e.g., a city scene in LSF superimposed with a highway scene in HSF). When exposure duration of hybrid images was very short (30 ms), perception of the hybrids was dominated by LSF information. However, when exposure duration was longer (150 ms), perception of the hybrids was dominated by HSF information. This suggests that scene perception follows a predominantly coarse-to-fine processing sequence, with a precedence of LSF processing over HSF processing.

Bar and colleagues (Bar, 2003, 2007; Bar et al., 2006) proposed a proactive model for the cerebral mechanisms which may subserve visual recognition. According to this model, LSF information reaches the prefrontal cortex rapidly and activates plausible semantic interpretations relating to the stimulus, based on its coarse physical properties. The result of these predictions is then projected, via feedback connections, into the inferotemporal cortex (such as the fusiform and parahippocampal areas) in order to guide and facilitate subsequent processing of the finer HSF information conveyed more slowly through the ventral visual stream into the inferotemporal cortex. However, only a few studies have directly investigated the influence of LSF processing on subsequent HSF processing, or the way in which LSF and HSF information integrate during scene categorization.

Mu and Li (2013), for example, investigated the neural correlates and time course of the integration of LSF and HSF information using ERP recordings during scene categorization. The authors presented hybrid scenes to participants who had to ignore the LSF component and perform a categorization task (man-made vs. natural) based only on the HSF component. Based on the assumption that a default coarse-to-fine processing occurs during scene categorization, the authors expected that the LSF component of the hybrid would be processed automatically and would influence the processing of HSF even if it was not relevant to performance of the task. For this purpose, they manipulated the semantic similarity between the LSF and HSF components of the hybrids. The two scenes of which the hybrids were made up were either semantically similar (e.g., LSF of a

natural scene superimposed with HSF of another natural scene) or semantically dissimilar (e.g., LSF of a natural scene superimposed with HSF of a man-made scene). Participants made more categorization errors when hybrids were semantically dissimilar than when they were similar. This semantic interference effect revealed that the semantic information contained in the task-irrelevant LSF component was processed automatically, and that it interfered with the categorization of the semantic information contained in HSF in the semantically dissimilar condition. In addition, the authors tested whether semantic interference could be modulated by the physical properties in scenes by manipulating the physical similarity (i.e. spatial overlap) between LSF and HSF. The two scenes composing the hybrid had either similar or dissimilar physical properties. Behavioral results showed that the semantic interference effect was greater when the two scenes composing the hybrids were physically similar. This result suggested that the spatial overlap of LSF on HSF caused even greater impairment of HSF categorization, in the absence of any semantic congruence between the two scenes. On a neurobiological level, results from ERP recordings revealed that semantic interference was associated with a negative frontal component (N1) 122 ms after stimulus onset, and that this component was greater when hybrids were physically similar than when they were physically dissimilar. The early frontal component associated with the semantic interference effect was followed by positive parietal and occipital components (P2 and P3) 247 and 344 ms after stimulus onset, respectively. In the context of the proactive model of visual perception (Bar, 2003; 2007; Bar et al., 2006), these results suggested that when the LSF and HSF components of the hybrids were semantically dissimilar, coarse processing of LSF physical properties led to erroneous semantic predictions, resulting in greater bias in HSF categorization, and a more pronounced interference effect in frontal areas. Overall, these results suggested that processing of the semantic properties of LSF and HSF may occur in the early stages of visual processing in frontal areas, where it initiates top-down facilitation (Bar et al., 2006).

The aim of the present study was to further examine, within the theoretical framework of coarse-to-fine scene categorization, how LSF in scenes may modulate and interfere with the processing of HSF by investigating, using fMRI, the neural bases of the semantic interference effect. As in the study by Mu and Li (2013), participants had to categorize hybrid scenes (man-made vs. natural) based on their HSF component and ignore the LSF component. We manipulated the semantic and physical similarity between LSF and HSF scenes composing the hybrids. The two scenes composing the hybrids were thus either (1) semantically and physically similar, (2) semantically similar and physically dissimilar, (3) semantically dissimilar and physically similar, or (4) semantically and physically dissimilar. Based on Mu and Li's (2013) results, we expected that categorization of the HSF component in the hybrids would be impaired when the LSF component was semantically dissimilar. Furthermore, we expected the semantic interference effect to be stronger when physical properties of the LSF and HSF components were similar than when they were dissimilar. At the cerebral level, we expected that the semantic interference effect would be associated mainly with activation of the prefrontal cortex (Bar et al., 2006). Furthermore, according to Bar's proactive model of visual recognition, the prefrontal cortex would influence processing of information in the fusiform gyrus. We therefore additionally performed a functional connectivity analysis (psycho-physiological interaction, PPI) to examine whether connectivity between the fusiform gyrus and the prefrontal cortex was influenced by the semantic similarity between LSF and HSF scenes in hybrids.

2. Material and Method

2.1. Participants

Fourteen right-handed participants (2 males; Mean age ± SD: 24 ± 7 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, which was in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans approved by the local ethics committee.

2.2. Stimuli and procedure

Stimuli were constructed from 160 black and white photographs (256-level grey-scales, 256 x 256 pixels) taken among the 2687 scenes from the Labelme database (Oliva & Torralba, 2001). Scenes were classified into two distinct categories (80 natural scenes - e.g., forests, coasts, and open country-side; and 80 man-made scenes - e.g., highways, streets and buildings) and substended a visual angle of 6 x 6 degrees.

Scenes were selected in order to obtain 40 scenes from each category with similar physical properties, and 40 scenes from each category with dissimilar physical properties. Physical similarity of scenes was assessed using two criteria: the correlation between images based on their pixel intensity values (Image correlation) and the correlation between amplitude spectra of images (AS correlation, i.e. correlation between pixel intensity values of amplitude spectra, i.e. the module of the Fourier transform of images).

When considering the whole database (i.e. 2687 scenes) the mean Image correlation between all pairs of scenes in the database was 0.23 (SD = 0.26; ranging from -0.93 to 0.99) and the mean AS correlation was 0.52 (SD = 0.11; ranging from -0.08 to 0.98). Two scenes were considered physically similar if their Image correlation coefficient and AS-correlation coefficient were superior to 0.6 (only 1.59% of images pairs matched both criteria). Two scenes were considered physically dissimilar if their Image 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). It should be noted that overall, AS correlation coefficients were higher than Image correlation coefficients because of the similar 1/f shaped amplitude spectra, characteristic of natural scenes (Field, 1987) Mean correlation coefficients for physically similar (PS) and physically dissimilar (PD) conditions are shown in Table 1.

Table 1: Mean (±*SD*) Image and Amplitude spectra (AS) correlation between the scenes composing the hybrids for each semantic and physical similarity condition.

	Semantically similar		**Semantically dissimilar**	
	Mean	±*SD*	Mean	±*SD*
Physically similar				
Image correlation	0.75	± *0.08*	0.73	± *0.06*
AS correlation	0.67	± *0.03*	0.66	± *0.04*
Physically dissimilar				
Image correlation	0.01	± *0.00*	0.01	± *0.00*
AS correlation	0.43	± *0.06*	0.44	± *0.05*

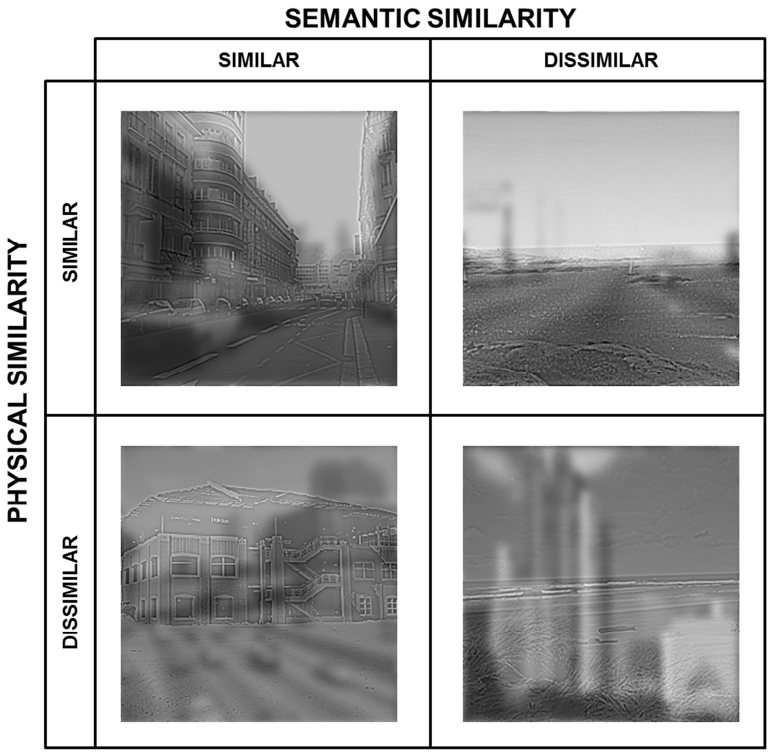


Figure 1: Example of hybrid stimuli from the two semantic similarity conditions and the two physical similarity conditions. The two scenes of which the hybrids were composed could be either (1) semantically and physically similar (e.g., LSF of a man-made scene superimposed with HSF of another man-made scene with similar physical properties), (2) semantically dissimilar and physically similar (e.g., LSF of a man-made scene superimposed with HSF of a natural scene with similar physical properties), (3) semantically similar and physically dissimilar (e.g., LSF of a man-made scene superimposed with HSF of another man-made scene with dissimilar physical properties), or (4) semantically and physically dissimilar (e.g., LSF of a man-made scene superimposed with HSF of a natural scene with dissimilar physical properties).

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 frequency content above 2 cycles per degree of visual angle (cpd, i.e. 12 cycles per image) in LSF scenes 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 luminance values between 0 and 1 (i.e. mean luminance of 128 on a gray-level scale). We then created hybrid stimuli 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 (SS-PS, SS-PD, SD-PS, and SD-PD, see Figure 1) and as a result 160 hybrid stimuli (40 per hybrid condition). The HSF component of half of the hybrids in each experimental condition was that of a natural scene. The HSF component in the other half of the hybrids was that of a man-made scene.

Stimuli were displayed using E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, USA) and back-projected onto a translucent screen positioned at the rear of the magnet. Participants viewed this screen at a distance of about 222 cm via a mirror attached to the head coil. We used a backward mask, built with 1/f noise, to prevent retinal persistence of the scenes.

An event-related design paradigm was used. The experiment consisted of four functional runs. Each functional run lasted about five minutes and was composed of 95 events, including 80 hybrid events (20 per hybrid condition) and 15 null events (i.e. a black fixation dot displayed against a gray background). The order of hybrid conditions and null events was pseudo-randomized based on an optimization algorithm (Friston, Zarahn, Josephs, Henson, & Dale, 1999). Each stimulus was displayed for 100 ms, followed by the mask (1/f noise) for 30 ms, and a fixation dot in the center of the screen. The interval between the onsets of two successive stimuli was 2.5 seconds. Participants had to give a categorical answer based on the HSF components of the hybrids ("natural" or "man-made") by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. They were instructed to fixate on the center of the screen (fixation dot) during the entire run and to respond as quickly and as accurately as possible by pressing one of two response keys. Half of the participants had to answer "natural" with their forefinger and "man-made" with the middle finger, while the second half of the participants had to answer "natural" with their middle finger and "man-made" with the forefinger. Response accuracy and reaction times (in milliseconds) were recorded.

2.3. fMRI Acquisition

Experiments were performed using a whole-body 3T Philips scanner (Achieva 3.0T TX Philips - Philips Medical Systems, Best, NL) with a 32-channel head coil at the Grenoble MRI facility IRMaGe in France. For all functional scans, the manufacturer-provided gradient-echo/T2* weighted EPI method was used. Forty-four adjacent axial slices parallel to the bi-commissural plane were acquired in sequential mode. Slice thickness was 3 mm. The in-plane voxel size was 2.5×2.5 mm (220×220 mm field of view acquired with a 88×85 pixel data matrix; reconstructed with zero filling to 96×96 pixels). The main sequence parameters were: TR = 2.5 s, TE = 30 ms, flip angle = 80°. Finally, a T1-weighted high-resolution three-dimensional anatomical volume was acquired, by using a 3D T1 TFE sequence (field of view = 256×224×175 mm; resolution: 1.333×1.690×1.375 mm; acquisition matrix: 192×115×128 pixels; reconstruction matrix: 288×288×128 pixels).

2.4. Whole-brain analysis

Data analysis was performed using the general linear model (Friston et al., 1995) for event-related designs in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK, http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB 7 (Mathworks Inc., Sherborn, MA, USA). Individual scans were realigned, normalized to the MNI space and spatially smoothed by an 8-mm FWHM (Full Width at Half Maximum) Gaussian kernel. Times-series for each voxel were high-pass filtered (1/128 Hz cutoff) to remove low-frequency noise and signal drift.

FMRI signal was analyzed using single-participant general linear model. For each participant, four conditions of interest (SS-PS, SS-PD; SD-PS, and SD-PD) were modeled as four regressors convolved with a canonical hemodynamic response function. Movement parameters derived from realignment corrections (three translations and three rotations), response accuracy and response times were also entered into the design matrix as additional factors of no interest, to account for head motion-, accuracy-, and reaction time-related variance, respectively. Analyses were performed at the individual subject level to examine the contrasts between hybrid

conditions. These contrast images were then entered into second-level random effect analyses to test for within group effects (one-sample *t* tests). The significance of activations was assessed with a statistical threshold of $p < 0.001$, uncorrected for multiple comparisons at the peak-level ($T > 3.85$, $k = 15$), and a threshold of $p < 0.05$, FWE corrected at the cluster-level.

2.5. Functional connectivity analysis

Based on Bar's proactive model of visual recognition (Bar, 2003; Bar et al., 2006) suggesting that LSF-based computations performed in the prefrontal cortex influence processing of HSF information in the fusiform gyrus, we hypothesized that connectivity between these regions would be influenced by the semantic similarity between LSF and HSF scenes in hybrids. To address this question, we performed a PPI analysis. This approach allows to determine whether the connectivity between a seed region and all other voxels in the brain is affected by a psychological factor, such as the experimental conditions (K J Friston et al., 1997).

PPI analysis was performed using the standard procedure on SPM8. First, for each participant, seed regions were identified with a contrast in which all hybrids conditions were compared to the baseline. We used this contrast instead of the [SD > SS] contrast, because performing PPI analysis with a seed activated in the contrast of interest can lead to regressor collinearity and a lack of power. Indeed, in this latter case, the physiological and psychological variables would be correlated and would therefore also correlate with the psychophysiological interaction. Separate PPI analyses were performed for each seed region. In the present study, the seed regions were the right and left fusiform gyrus. The choice of these seed regions was motivated by the fact that (1) they were more easily identifiable at the individual level than the inferior frontal gyrus which could not be identified in each participants at the statistical threshold used ($p < 0.05$) and (2) they showed less variability in individual activation peaks coordinates than the inferior frontal gyrus, which peak coordinates varied more between participants. The physiological activity of each seed region was computed by extracting time series (i.e. principal component adjusted for effects of interest and BOLD-deconvolved to get an estimate of the neural response; Gitelman, Penny, Ashburner, & Friston, 2003) of all voxels within 4-mm spheres centered on subject-specific maxima (identified at a threshold of $p < 0.05$, uncorrected) closest to the group activation peak (identified at a threshold of $p < 0.001$, uncorrected) of the left and right fusiform gyri (Mean Talairach coordinates *(± SD)*: Left fusiform gyrus, x = -26 *(± 5)*, y = -42 *(± 5)*, z = -10 *(± 3)*; Right fusiform gyrus, x = -27 *(± 4)*, y = -45 *(± 6)*, z = -13 *(± 4)*. The semantic interference effect (i.e. [SD > SS] contrast) was used to compute the psychological factor hypothesized to influence connectivity. Finally, the PPI variable representing the interaction between the physiological and psychological factors was computed. For each participant, a new first-level general linear model was created for each PPI analysis, including the PPI, physiological, and psychological variables as regressors of interest convolved with the canonical hemodynamic response function. RT and accuracy were also modeled as regressors of no interest (O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012). Analyses were performed at the individual level to examine the effect of the PPI term. These contrast images were then entered into second-level random effect analyses to test for within group effects (one-sample *t* tests). This produced statistical parametric maps revealing voxels with significant positive connectivity with the seed regions, due to the semantic dissimilarity between LSF and HSD scenes in hybrids. The significance of activations was assessed with a statistical threshold of $p < 0.001$, uncorrected for multiple comparisons at the peak-level ($T > 3.85$, $k = 5$).

For the whole brain and PPI analyses, a transformation of MNI into

Talairach and Tournoux (1988) coordinates was performed using the MNI2TAL function (created by Matthew Brett, available at www.mrc-cbu.cam.ac.uk/Imaging), in order to facilitate comparison with other studies.

3. Results

3.1. Behavioral results

Two 2 × 2 analyses of variance (ANOVAs) with Semantic similarity (SS and SD) and Physical similarity (PS and PD) as within-subjects factors were conducted on mean error rates (mER) and mean correct reaction times (mRT, see Figure 2).

The ANOVA conducted on mER revealed a main effect of Semantic similarity ($F_{1, 13}$ = 51.52, p < 0.001, η^2_p = 0.798). Participants made more errors when categorizing the hybrids in the SD condition (Mean ± SE: 34.69 ± 3.14%) than in the SS condition (9.73 ± 1.46%), suggesting that the semantic information of the LSF component interfered with the processing of the semantic information of HSF component. There was no main effect of Physical similarity (PS: 22.59 ± 2.01%; PD: 21.83 ± 1.73%; $F_{1, 13}$ < 1, η^2_p = 0.019). The expected interaction between Semantic similarity and Physical similarity was not significant ($F_{1, 13}$ = 2.59, p = 0.13, η^2_p = 0.166).

The ANOVA conducted on mRT revealed a main effect of Semantic similarity ($F_{1, 13}$ = 23.17, p < 0.001, η^2_p = 0.641). Participants categorized the hybrids more rapidly in the SS condition (795 ± 29.46 ms) than in the SD condition (868 ± 33.15 ms). As for the analysis on mER, this semantic interference effect indicated that the processing of the LSF component interfered with the processing of the HSF component at the semantic level. There was no effect of Physical similarity (PS: 840 ± 33.50 ms; PD: 822 ± 27.95 ms; $F_{1, 13}$ = 3.30, p = 0.09, η^2_p = 0.202). The expected interaction between Semantic similarity and Physical similarity was not significant ($F_{1, 13}$ < 1 η^2_p = 0.002).

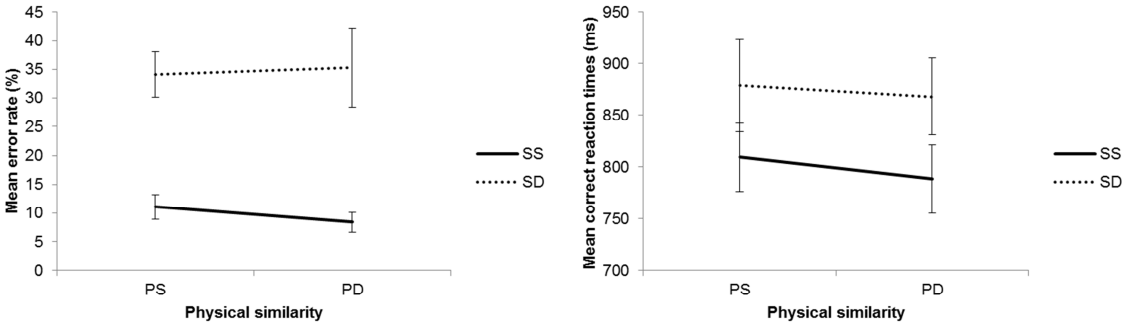


Figure 2: Mean error rates (%) and mean correct reaction times (milliseconds) for the categorization of hybrids, according to semantic similarity (semantically similar – SS, and semantically dissimilar – SD) and physical similarity (physically similar – PS, and physically dissimilar – PD) between the two scenes composing the hybrids. Error bars indicate standard error.

3.2. fMRI results

Whole-brain results

We examined the regions specifically involved by the semantic interference effect of LSF on HSF processing, by contrasting activations elicited by hybrids in the SD condition to activations elicited by hybrids in the SS condition ([SD > SS] contrast, cf. Figure 3). Results revealed that the semantic interference effect involved more strongly temporal, parietal, and frontal areas (cf. Table 2). Temporal activations corresponded to the bilateral fusiform gyri (BA 20/37) and parahippocampal gyri (BA 37), and the right inferior temporal gyrus (BA 37). Occipito-parietal activations involved bilaterally the superior parietal lobules (BA 7) and the superior occipital gyrus (i.e. occipito-parietal junction) extending to the middle occipital gyrus in the right hemisphere. Finally, frontal activation involved the left medial and superior frontal gyrus (BA 8) including the supplementary motor area (BA 6), and the bilateral inferior frontal gyri (AB 45/46) extending to the orbitofrontal cortex (BA 47) in the right hemisphere. The reverse contrast ([SS > SD] revealed stronger activation of the bilateral superior frontal gyri (BA 10), the left posterior cingular gyrus (BA 31), and the left middle temporal gyrus (BA 39).

Although it was not of main interest of the study, we also tested the effect of physical similarity between scenes in hybrids by contrasting activations between the physically similar and dissimilar conditions. The [PD > PS] contrast revealed that physical dissimilarity between scenes in hybrids resulted in greater activation of occipito-temporal areas, including the bilateral lingual gyrus (BA 17/18), the bilateral cuneus (BA 19), the right inferior occipital gyrus (BA 18/19), the left middle occipital gyrus (BA 19/20), and the right superior temporal gyrus (BA 42). The reverse contrast ([PS > PD]) was not significant. Finally, the interaction between the semantic and physical similarity of scenes in hybrids was not significant.

Functional connectivity results

PPI for the semantic interference effect revealed stronger functional connectivity between the right parahippocampal gyrus and the right inferior frontal gyrus (BA 44, k = 19, peak coordinates: x = 39, y = 16, z = 11; t = 5.21; BA 45, k = 10, peak coordinates: x = 39, y = 28, z = 16; t = 5.05; Figure 3b) in the SD condition than in the SS condition. Importantly, this activation cluster overlapped with the activation of the inferior frontal gyrus observed with the [SD > SS] contrast, in the whole brain analysis (cf. Figure 3c). The PPI analysis using the left fusiform gyrus as seed region did not reveal any significant activation.

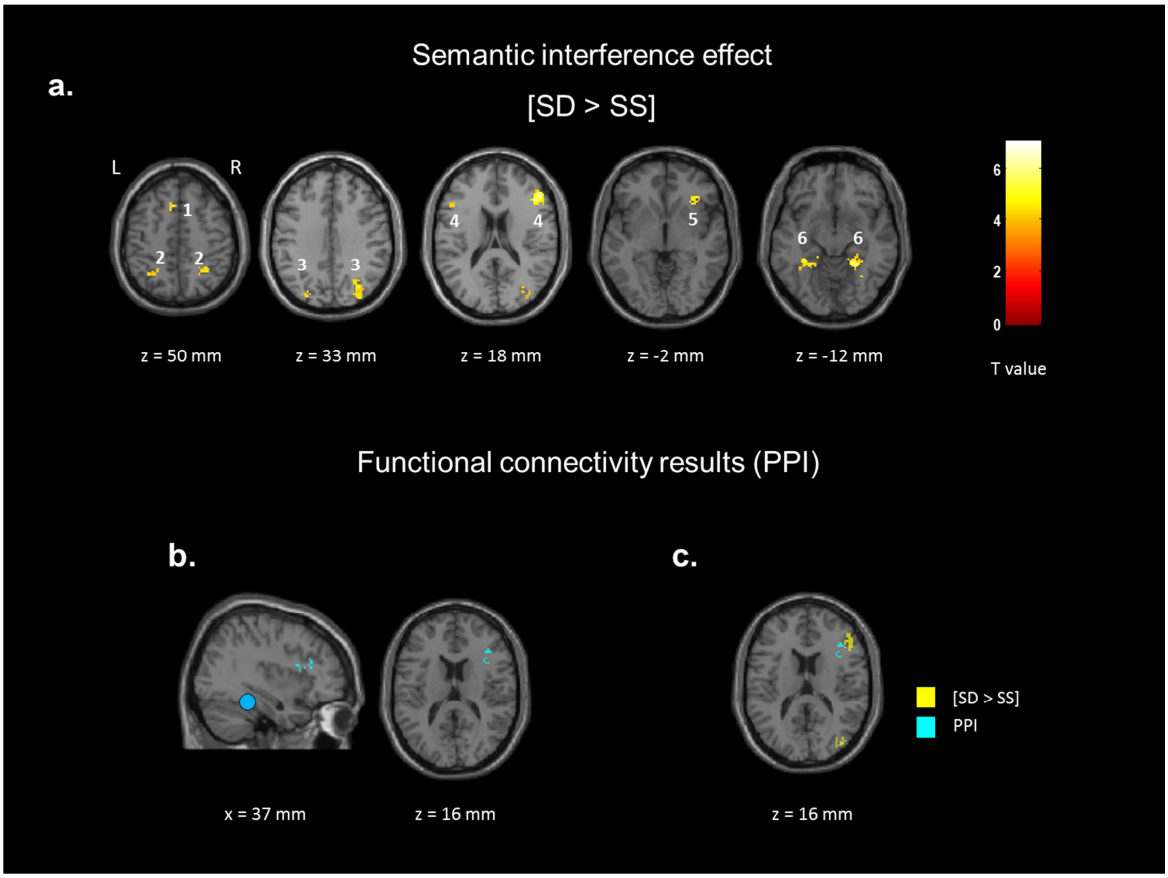


Figure 3: (a) Activations elicited by the semantic interference effect ([SD > SS] contrast): (1) the superior frontal gyrus, (2) the superior parietal lobules, (3) the superior/middle occipital gyri, (4) the inferior frontal gyri, (5) the orbitofrontal cortex, and (6) the parahippocampal and fusiform gyri. Activations are reported at a statistical threshold of $p < 0.001$ uncorrected for multiple comparisons at the cluster level, and a threshold of $p < 0.05$ FEW-corrected at the peak-level. (b) Region (right inferior frontal gyrus) showing increased connectivity with the right fusiform gyrus (blue circle) during the semantic interference effect, identified with the PPI analysis. (c) Superposition of clusters activated for the semantic interference effect ([SD > SS] contrast, in yellow) and clusters showing increase in connectivity with the parahippocampal gyrus for the semantic interference effect (PPI analysis, in blue).

Table 2: Cerebral regions specifically activated for the semantic interference effect ([SD > SS] and [SS > SD] contrasts). For each cluster, the region showing the maximum t value is listed first, followed by the other regions in the cluster [in brackets]. Talairach coordinates (x, y, z) are indicated. Abbreviations: R = right hemisphere; L = left hemisphere; BA = Brodmann area; k = number of voxels in the cluster. Activations are reported at a statistical threshold of $p < 0.001$, uncorrected for multiple comparisons, and a minimum cluster size of k = 15 voxels.

	Side	BA	*k*	*x*	*y*	*z*	*t*
Semantic interference effect							
[SD > SS]							
Temporal areas							
Parahippocampal gyrus	R	19	136	24	-46	-4	6.63
[Fusiform gyrus]		37		29	-54	-12	6.10
Parahippocampal gyrus	L	37	91	-33	-39	-10	5.56
[Fusiform gyrus]		20		-32	-37	-15	5.45
[Fusiform gyrus]		37		-23	-44	-10	5.33
Occipito-Parietal areas							
Superior parietal lobule	R	7	58	37	-50	53	6.57
Superior parietal lobule	L	7	15	-23	-60	54	5.17
Superior occipital gyrus	R	19	207	29	-71	29	6.98
[Middle occipital gyrus]	R	18/19		42	-82	8	6.02
[Inferior parietal lobule]		40		32	-65	40	4.55
Superior occipital gyrus	L	19	29	-25	-78	35	5.77
Frontal areas							
Inferior frontal gyrus	R	45	116	49	33	16	7.01
Orbitofrontal cortex	R	47	47	29	25	-3	6.07
[Inferior frontal gyrus]	R	45/47		37	32	2	5.73
Inferior frontal gyrus	L	44	23	-40	4	26	5.36
			31	-50	21	25	4.83
Medial frontal gyrus	L	6	36	-8	17	47	5.75
[Superior frontal gyrus]		6		-10	10	53	4.92
[SS > SD]							
Superior frontal gyrus	L	10	111	-5	61	-2	6.48
[superior frontal gyrus]	R	10		2	64	0	5.26
Cingular gyrus	L	31	27	-8	-39	39	6.14
Middle temporal gyrus	L	39	21	-60	-9	4	5.23
Physical interference effect							
[PD > PS]							
Lingual gyrus	L	17/18	592	-15	-87	0	6.77
[lingual gyrus]	R	17/18		17	-82	-5	6.43
[inferior occipital gyrus]	R	18/19		37	-75	-5	5.73
Cuneus	R/L	19	24	0	-85	30	5.86
Cuneus	R	19	33	24	-86	27	5.71
Middle occipital gyrus	L	19/20	34	-43	-77	10	5.73
Superior temporal gyrus	R	42	50	44	-33	11	5.26
[PS > PD]							
No significant cluster							

4. Discussion

Influential models of visual perception (Bar, 2003; Bar et al., 2006; Hegdé, 2008; Kauffmann et al., 2014; Schyns & Oliva, 1994) postulate that during scene categorization, LSF are processed earlier than HSF (coarse-to-fine processing sequence). This initial low-pass analysis would be used to generate predictions about the nature of the visual input, and facilitate the subsequent processing of HSF information (Bar, 2003; 2007; Bar et al., 2006). The aim of the present fMRI study was to gain better understanding of how the processing of LSF may modulate and interfere with the processing of HSF, by investigating the neural bases of the semantic interference effect using fMRI, and

within the theoretical framework of the coarse-to-fine categorization of scenes.

On a behavioral level, we observed that the categorization of HSF scenes in hybrids was impaired when the LSF scene was semantically dissimilar. This result replicates the previous findings of Mu and Li (2013) and indicates that, although it was not relevant to the task, the non-attended categorical information from LSF was processed automatically, and interfered with the processing of the task-relevant HSF semantic information. This semantic interference effect could be interpreted in the context of the proactive model of visual perception proposed by Bar (2003; 2007; Bar et al., 2006). In this model, coarse information from LSF is used to generate predictions about the visual input, and guides subsequent processing of HSF. Therefore, in the semantically dissimilar condition, the rapid processing of the LSF scene in the hybrid could have led to erroneous semantic predictions for the categorization of the HSF scene, resulting in an increase in error rates and reaction time for the correct categorization of HSF information. However, unlike Mu and Li (2013), we did not find that the semantic interference effect was modulated by the physical similarity between HSF and LSF scenes in hybrids. It should be noted that in the present study, reactions times were longer overall (800-1000 ms) than those reported in the Mu and Li (2013) study (under 800 ms). This may have been due to greater physical constraints during the fMRI experiment (i.e. participants had to adopt a lying-down position in the magnet, and were instructed not to move). These methodological differences may have prevented us from detecting any modulation of the semantic interference effect by the physical similarity of the scenes in reaction times.

On a neurobiological level, fMRI results supported Mu and Li's (2013) EEG results, as they revealed that the semantic interference effect ([SD > SS] contrast) was associated with greater activation in specific areas of the frontal, parietal, and temporal cortices. However, our fMRI results did not reveal significant interaction between the semantic and physical similarity of scenes in hybrids.

As expected, in the frontal cortex, the semantic interference effect was associated with the bilateral activation of the inferior frontal gyrus, extending to the right orbitofrontal cortex. These regions of the prefrontal cortex are thought to be particularly involved in tasks requiring the retrieval of semantic information related to visual stimuli (Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005; Freedman et al., 2001; Wagner, Pare, Clark, & Poldrack, 2001). They also contribute to the maintenance of visual information in visual awareness or working memory (Fletcher & Henson, 2001). Activation of the inferior frontal gyrus therefore seems consistent with the categorization task demands in the semantically dissimilar (SD) condition, which involved the semantic processing of different visual information. This result can be linked to Mu and Li's (2013) ERP results showing a frontal N1 component associated with the semantic interference effect. However, these authors observed that this component was greater in case of physical similarity between the scenes in hybrids only. In their study, this N1 component was observed very early on, 120 ms after stimulus onset, suggesting that frontal areas were involved in the semantic processing of spatial frequencies in the early stages of visual processing. The involvement of the inferior frontal gyrus during rapid scene categorization was also evidenced in a combined fMRI and ERP study conducted by Peyrin et al. (2010). In that study, the authors showed that the successive processing of an LSF scene followed by an HSF scene resulted in both greater fMRI activation and greater ERP source activity (150-194 ms after the onset of the LSF scene) in the inferior frontal gyrus. These findings suggest that the greater activation of the inferior frontal gyrus in the semantically dissimilar condition may be linked to the

rapid processing of LSF information in scenes. Overall, the greater activation of the inferior frontal gyrus associated with the semantic interference effect in the present study, together with findings from previous studies (Mu & Li, 2013; Peyrin et al., 2010), are consistent with the proactive model of visual recognition proposed by Bar and colleagues (Bar, 2003; 2007; Bar et al., 2006). According to this model, semantic predictions performed on the basis of LSF information occur very early in the inferior frontal gyrus (in the orbitofrontal cortex), and initiate top-down facilitation to guide subsequent processing of HSF. As previously mentioned, in our semantically dissimilar condition, the rapid processing of the LSF scene could have led to erroneous semantic predictions (as suggested by the increased error rates in behavioral results), which may have resulted in increased activity in the inferior frontal gyrus. Our results therefore provide a new argument in favor of the involvement of the inferior frontal gyrus in processing the semantic properties of spatial frequency information. It can be noted that activation of the inferior frontal gyrus, and mainly the orbitofrontal cortex, appears to be slightly right-lateralized. Interestingly, a recent fMRI study (Fintzi & Mahon, 2013) demonstrated that the right orbitofrontal cortex responds more strongly to the LSF components in a coarse-to-fine sequence processing of objects, whereas the left orbitofrontal cortex responds more strongly to the medium and HSF components. As previously suggested, categorization of the HSF scenes in the semantically dissimilar hybrid condition was interfered by the processing of the incongruent semantic information of the LSF scenes. The right orbitofrontal cortex could thus have been more sensitive to the incongruent semantic information contained in LSF.

Additional frontal cortex activation was observed in the left superior frontal gyrus, including the supplementary motor area (BA 8/6), when the two scenes were semantically dissimilar (compared to the semantically similar condition). These regions have been shown to be involved when conflict between alternative motor responses occurs, as is the case in go/no-go tasks (e.g., Garavan, Ross, Kaufman, & Stein, 2003; Garavan, Ross, Murphy, Roche, & Stein, 2002), and activity in this region increases with increasing uncertainty in decision making (Volz, Schubotz, & von Cramon, 2005). Stronger activation of this region in the semantically dissimilar condition, in which the incongruent category of the LSF scene interfered with the response to the HSF scene, could therefore reflect increased uncertainty and greater conflict between the motor responses to the HSF scene category.

In the parietal cortex, the semantic interference effect was associated with activation of the right occipito-parietal junctions. Activation of the right occipito-parietal junction has been observed previously in tasks requiring category-matching between two scenes (Peyrin et al., 2005) which may be consistent with our task involving the simultaneous processing of two different scenes. We also observed greater activation of the superior parietal lobules associated with the semantic interference effect. Involvement of these regions has previously been shown in tasks requiring high levels of attention (Behrmann, Geng, & Shomstein, 2004; Peyrin et al., 2010; Posner & Petersen, 1990), and attentional selection of relevant visual information (Capotosto et al., 2013; Shomstein, 2012; Yantis et al., 2002) or spatial frequency content (Baumgartner et al., 2013; Peyrin et al., 2010, 2004) in order to perform a visual task. The superior parietal lobule in particular appears to be predominantly involved in the top-down, task-driven, allocation of attention (Corbetta & Shulman, 2002; Shomstein, 2012; Yantis et al., 2002). Greater activation of this region in the semantically dissimilar condition could therefore reflect the use of greater attentional resources (as suggested by the increased error rate and reaction times in that condition; see behavioral results),

leading to an intensive recruitment of this parietal structure in order to allocate attention to the relevant spatial frequency information needed to perform the task.

Finally, in the occipito-temporal cortex, the semantic interference effect was associated with activation of the fusiform and parahippocampal gyri. The fusiform gyrus is known to be involved in object recognition (Grill-Spector, 2003; Haxby et al., 2001; Ishai et al., 1999; Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001; Tanaka, 1996). Activation in the parahippocampal gyrus overlaps with activation in the parahippocampal place area, which has been shown to be predominantly and selectively involved in scene perception and recognition (Epstein et al., 1999; Epstein & Kanwisher, 1998). Importantly, the fusiform and parahippocampal gyri both play a central role in Bar's model of visual recognition (Bar, 2003, 2007; Bar et al., 2006). In this model, these regions of the inferotemporal cortex combine top-down LSF-based semantic predictions from frontal areas with bottom-up HSF information from the ventral visual stream, in order to enable recognition. In that theoretical context, greater activation of these regions in the SD condition (compared to in the SS condition) could therefore be explained by a conflict between (1) top-down predictions based on the LSF component in hybrids and (2) the bottom-up processing of HSF information.

Results of the present study may raise the question of the interactions between regions involved in the semantic interference effect. In their study, Mu and Li (2013) showed that the first ERP component associated with the semantic interference effect occurred in frontal areas, followed by ERP components in parietal and occipital areas. According to these authors, this first ERP component indicated that information from both LSF and HSF was rapidly processed on a semantic level in frontal areas. Although this appears to be in disagreement with Bar's (2003; 2007; Bar et al., 2006) assumption that only LSF information is rapidly processed in inferior frontal areas, the authors suggested that the demands of the task, which involved focusing the attention on HSF information, may have resulted in the enhancement of HSF perception and processing speed (Abrams, Barbot, & Carrasco, 2010; Carrasco & Mcelree, 2001; Oliva & Schyns, 1997), enabling rapid processing of that information in inferior frontal areas. The semantic processing of LSF and HSF scenes in inferior frontal areas would have then initiated top-down influences to parieto-temporal areas, in order to facilitate subsequent perceptual analyses and allow the categorization task to be accomplished. However, since our fMRI results do not provide any information about the time course of activation, one possible alternative interpretation is that the greater activation of the inferior frontal gyrus associated with the semantic interference effect seen in the present study results from changes in connectivity with other regions, such as the parahippocampal place area and fusiform gyrus. It is possible that in the semantically dissimilar condition, for example, inferior frontal and occipito-temporal regions involved in the processing of visual scene information interacted more strongly, in order to disambiguate and enhance/inhibit processing of task-relevant/-irrelevant spatial frequency information.

In order to address whether connectivity between the inferior frontal gyrus and inferotemporal gyrus was modulated by the semantic interference effect, we conducted a functional connectivity analysis using PPI. The left and right fusiform gyri were defined as seed regions. Our results revealed that the semantic interference effect resulted in increasing connectivity between the right fusiform and right inferior frontal gyri. These results seem thus consistent with Bar's model of visual recognition (Bar, 2003, 2007; Bar et al., 2006) suggesting functional connectivity between the prefrontal cortex and the fusiform gyrus, and support our proposal that these regions

interacted more strongly in case of semantic dissimilarity between LSF and HSF scenes in hybrids. It should however be noted that the PPI method do not allow to determine the direction of interactions between regions, and only inform about regions which activity correlates with the one of the seed region, according to the experimental context. Therefore, it did not allow us to clarify which region influences the other.

Increase in functional coupling between inferotemporal areas and the inferior frontal gyrus in the context of an interference effect is supported by the results of a recent fMRI study (Wais et al., 2010) investigating the interference effect of irrelevant visual information (a visual scene) on a concurrent process (the recollection of objects) involving mental imagery. Firstly, results showed that the recollection process was associated with activation of the inferior frontal gyrus, the hippocampus and visual association cortex, and that this connectivity network was diminished when a visual stimulation interfered with the recollection process. This interference effect by a visual scene was associated with greater activation of the parahippocampal gyrus, and when the interference led to recollection errors, connectivity between the inferor frontal gyrus and the parahippocampal gyrus increased. The authors hypothesized that their results could reflect an overwhelming influence of a bottom-up processing of the interfering visual information. This would result in the inferior frontal gyrus being excluded from involvement in the recollection task in order to control and minimize these interfering influences. This would mean that in the last case the inferior frontal gyrus was involved in the cognitive control of task-relevant information processing (see also Chong, Williams, Cunnington, & Mattingley, 2008; Nelson, Reuter-Lorenz, Persson, Sylvester, & Jonides, 2009). Importantly, the inferior frontal gyrus has also been reported in a large number of studies implicating response conflict tasks, as monitoring response conflict resolution (e.g., in Stroop tasks or Go/No-Go tasks; see Nee, Wager, & Jonides, 2007 for a meta-analysis). It is thus possible that in the present study, greater activation of the inferior frontal gyrus for the semantic interference effect also reflect such management of response conflict. Unfortunately, the design used here does not allow us to dissociate the effects of semantic similarity from the effets of response conflict resolution since these factors were confounded.

To conclude, results of the present study provide additional support for influential models of visual perception (Bar, 2003, 2007; Bar et al., 2006; Peyrin et al., 2010; Schyns & Oliva, 1994), which suggest that LSF information in scenes is processed automatically and rapidly on a semantic level – in keeping with the coarse-to-fine hypothesis – and modulates subsequent processing of HSF information. Based on previous studies (Bar et al., 2006; Mu & Li, 2013; Peyrin et al., 2010), we propose that during scene categorization, information from LSF may be rapidly processed in the inferior frontal gyrus in order to activate plausible interpretations of the scene category. These coarse predictions may then initiate top-down influences on recognition-related areas of the inferotemporal cortex, and thus could interfere with the categorization of HSF information in the case of semantic dissimilarity between LSF and HSF. This may result in greater interactions between frontal and inferotemporal areas, in order disambiguate between LSF and HSF inconsistent information. During the categorization process, the parietal cortex may also be recruited to constrain the allocation of attention to the relevant visual information.

Acknowledgment

This work was supported by the RECOR ANR Grant (ANR-12-JHS2-0002-01 RECOR) and a subvention from "Grenoble Pôle Cognition". Louise Kauffmann was supported by Région Rhône-Alpes (Cible Grants). Grenoble MRI facility IRMaGe was

partly funded by the French program "Investissement d'Avenir" run by the "Agence Nationale pour la Recherche": Grant "Infrastructure d'Avenir en Biologie Santé" (ANR-11-INBS-0006). We thank Catherine Dal Molin for the English revision of the manuscript.

References

Badre, D., Poldrack, R. a, Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*, 907–18.

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

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*, 280–9.

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*, 449–454.

Baumgartner, F., Hanke, M., Geringswald, F., Zinke, W., Speck, O., & Pollmann, S. (2013). Evidence for feature binding in the superior parietal lobule. *NeuroImage*, *68*, 173–80.

Behrmann, M., Geng, J. J., & Shomstein, S. (2004). Parietal cortex and attention. *Current Opinion in Neurobiology*, *14*, 212–7. Bullier, J. (2001). Integrated model of visual processing. *Brain Research*, *36*, 96–107.

Capotosto, P., Tosoni, A., Spadone, S., Sestieri, C., Perrucci, M. G., Romani, G. L., … Corbetta, M. (2013). Anatomical segregation of visual selection mechanisms in human parietal cortex. *Journal of Neuroscience*, *33*, 6225–9. Chong, T. T.-J., Williams, M. a, Cunnington, R., & Mattingley, J. B. (2008). Selective attention modulates inferior frontal gyrus activity during action observation. *NeuroImage*, *40*, 298–307.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*, 201–15.

De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559.

De Valois, R., Yund, W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 531–544.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The Parahippocampal Place Area : Recognition , Navigation , or Encoding ? *Neuron*, *23*, 115–125.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 6–9.

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.*, *4*, 2379–94.

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory Insights from functional neuroimaging. *Brain*, *124*, 849–881.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–6.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1995). Statistical Parametric Maps in Functional Imaging : A General Linear Approach. *Human Brain Mapping*, *2*, 189–210.

Friston, K., Zarahn, E., Josephs, O., Henson, R. N., & Dale, a M. (1999). Stochastic designs in event-related fMRI. *NeuroImage*, *10*, 607–19.

Garavan, H., Ross, T. J., Kaufman, J., & Stein, E. A. (2003). A midline dissociation between error-processing and response-conflict monitoring. *NeuroImage*, *20*, 1132–9.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable Executive Functions in the Dynamic Control of Behavior: Inhibition, Error Detection, and Correction. *NeuroImage*, *17*, 1820–1829.

Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance.* (Ginsburg, ., pp. 1–41). Oxford, England: John Wiley & Sons.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*, 159–166.

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–30.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*, 405–39.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, *8*, 197– 230.

Hupé, J., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., Bullier, J., … E, J. H. U. P. (2001). Feedback Connections Act on the Early Part of the Responses in Monkey Visual Cortex. *Journal of Physiology*, *85*, 134–145.

Ishai, A., Ungerleider, L., Martrin, A., Schouten, J., & Haxby, J. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*, 9379–9384.

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*, 1–14. Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of topdown facilitation in recognition. *The Journal of Neuroscience*, *27*, 13232–40.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, *11*, 287–97.

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PloS One*, *7*, e38493.

Nelson, J. K., Reuter-Lorenz, P. a, Persson, J., Sylvester, C.-Y. C., & Jonides, J. (2009). Mapping interference resolution across task domains: a shared control process in left inferior frontal gyrus. *Brain Research*, *1256*, 92–100.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, *42*, 145–175.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–60.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *NeuroImage*, *23*, 698–707.

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*, 2768–80.

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., & Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *NeuroImage*, *28*, 464–73.

Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative Ophthalmology*, *11*, 368–77.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Schyns, P. G., & Oliva, A. (1994). Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*, 195–201.

Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, *42*, 2547–54.

Shomstein, S. (2012). Cognitive functions of the posterior parietal cortex: top-down and bottom-up attentional control. *Frontiers in Integrative Neuroscience*, *6*, 38.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–39.

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

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.

Volz, K. G., Schubotz, R. I., & von Cramon, D. Y. (2005). Variants of uncertainty in decision-making and their neural correlates. *Brain Research Bulletin*, *67*, 403–12.

Wagner, A. D., Pare, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering Meaning : Left Prefrontal Cortex Guides Controlled Semantic Retrieval. *Neuron*, *31*, 329–338.

Wais, P. E., Rubens, M. T., Boccanfuso, J., & Gazzaley, A. (2010). Neural mechanisms underlying the impact of visual distraction on retrieval of long-term memory. *Journal of Neuroscience 30*, 8541–50.

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. a, Pekar, J. J., & Courtney, S. M. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature Neuroscience*, *5*, 995–1002.

2.4. Conclusions

Les résultats des **Expériences 3 et 4** ont permis de préciser, respectivement, les bases cérébrales de la catégorisation « coarse-to-fine » des scènes, ainsi que celles de l'interférence du traitement des BSF sur le traitement des HFS. Ces études fournissent des arguments supplémentaires en faveur d'un mécanisme de facilitation « top-down » lors de la catégorisation, par le biais duquel le traitement rapide (voire automatique) de l'information sémantique contenue dans les BFS viendrait moduler et s'intégrer au traitement plus tardif des HFS pour aboutir à la reconnaissance finale du stimulus visuel. Ces études ont également permis de mettre en avant le rôle du cortex frontal inférieur au cours de ce processus de catégorisation. Selon plusieurs travaux, cette région serait impliquée, de façon très précoce lors de l'analyse visuelle, dans la génération de prédictions concernant la nature du stimulus visuel (Bar, 2003; Bar et al., 2006; Peyrin et al., 2010). Cependant, nous avons vu dans l'**Expérience 4** que cette région pourrait également intervenir dans l'intégration sémantique de l'information en fréquences spatiales (Mu & Li, 2013) et lors de la gestion d'interférence ou la résolution de conflit quant au choix de la réponse à apporter pour effectuer la tâche (Chong et al., 2008; Nelson et al., 2009; Wais et al., 2010). Il serait donc intéressant d'approfondir ces travaux sur les fonctions d'intégration et de pilotage du cortex frontal inférieur afin de préciser dans quelle mesure les activations de cette région observées dans notre étude reflètent spécifiquement son implication dans le traitement de l'information sémantique contenue dans les scènes. En effet, l'une des principales limites de l'**Expérience 4** réside dans le fait que les effets d'interférence sémantique liés à l'intégration des fréquences spatiales et les effets liés à la résolution du conflit quant au choix de la réponse à apporter étaient complètement confondus. Ce paradigme ne nous permettait donc pas de dissocier l'influence relative de ces deux facteurs sur les activations cérébrales observées.

Ces travaux ont également permis de souligner le rôle du cortex visuel comme région intégrative, codant à la fois les afférences rétino-thalamiques et les influences « feedback » de régions de haut niveau (Bullier, 2001; Peyrin et al., 2010). En effet, bien que de nombreuses études aient auparavant mis en évidence une influence « top-down » sur le cortex visuel primaire, par exemple lors de l'orientation de l'attention à des points particuliers du champ visuel, ou dans le contexte de l'apprentissage perceptuel (i.e. effets facilitateurs lors de la réalisation d'une tâche perceptive suite à un apprentissage; voir pour une revue Rauss et al., 2011), nos travaux permettent d'élargir ces mécanismes à la catégorisation visuelle.

Enfin, nos travaux mettent en évidence l'implication de régions inférotemporales, et notamment du gyrus parahippocampique dans le traitement des fréquences spatiales lors de la catégorisation des scènes. En effet, nos résultats montrent une activation systématique du gyrus parahippocampique, et notamment la « parahippocampal place area » (PPA), dans nos tâches de perception des scènes, que ce soit lors de l'analyse « coarse-to-fine » de scènes ou lors de l'interférence de l'analyse des BFS sur les HFS. Il est important de souligner que la PPA a été initialement décrite par Epstein et Kanwisher (1998) comme étant sélectivement impliquée dans la perception de scènes, par rapport à la perception d'autres catégories de stimuli tels que les visages et les objets. Au sein du cortex occipito-temporal, il existe d'autres régions sélectives au traitement des scènes : « l'occipital place area » (OPA) et le « retrosplenial cortex » (RSC). Si les fonctions de ces régions dans la perception des scènes ont été largement étudiées, leur rôle et leur implication dans le traitement des fréquences spatiales lors de la catégorisation de scènes est encore peu connu. Nous verrons donc en détail, dans le Chapitre 4 qui suit, le rôle de ces régions lors de la perception de scènes, et présenterons deux études en IRMf visant à spécifier comment ces régions participent au traitement des fréquences spatiales dans les scènes (**Expérience 5**), et si un mode de traitement « coarse-to-fine » est également privilégié au sein de ces régions **(Expérience 6)**.

RESUME

Ce chapitre portait sur les bases cérébrales du traitement des fréquences spatiales et de la catégorisation « coarse-to-fine » des scènes. Nous avons tout d'abord rappelé les travaux qui montrent que le traitement des fréquences spatiales se ferait en lien avec l'organisation rétinotopique du cortex visuel et de façon latéralisée entre les deux cortex occipitaux. Nous avons par la suite examiné, à travers deux études IRMf, les bases cérébrales de la catégorisation « coarse-to-fine » des scènes et avons mis en évidence l'implication d'un large réseau cérébral incluant le cortex occipital, mais également le cortex frontal inférieur et le cortex inférotemporal. Dans l'**Expérience 3**, une analyse de connectivité nous a permis de montrer que lors de l'analyse « coarse-to-fine » des scènes, l'information visuelle serait transmise du cortex occipital vers le cortex inférotemporal et vers le gyrus frontal inférieur. Ce dernier exercerait des influences « top-down » sur le cortex inférotemporal, mais également sur le cortex occipital, afin de guider le traitement de l'information visuelle dans ces régions. Dans l'**Expérience 4**, nous avons montré que le gyrus frontal inférieur et le cortex inférotemporal participeraient activement à l'intégration des BFS et HFS d'une scène. Par ailleurs, nous avons à nouveau observé une connectivité fonctionnelle importante entre ces deux régions. Les résultats de ces travaux permettent de conforter et préciser les modèles actuels de la catégorisation visuelle de scènes en soulignant notamment le rôle intégratif du cortex visuel primaire et du cortex inférotemporal, ainsi que les fonctions intégratives et régulatrices du cortex frontal inférieur qui lui permettraient de piloter et guider le traitement de l'information visuelle.

CHAPITRE 4 - TRAITEMENT DES FREQUENCES SPATIALES AU SEIN DES REGIONS OCCIPITO-TEMPORALES SELECTIVES AUX SCENES

Comme nous l'avons vu précédemment, de nombreux travaux suggèrent que la reconnaissance s'achève au niveau des aires inférotemporales, situées au bout de la voie ventrale (Tanaka et al., 1991; Tanaka, 1993, 1996). De façon intéressante, il existe au sein du cortex inférotemporal une multitude de régions dites « catégories-spécifiques », qui répondent sélectivement à différentes catégories de stimuli. Par exemple, la « fusiform face area » (FFA) répond sélectivement aux visages (Kanwisher et al., 1997), tandis que le complexe occipital latéral (LOC) est sélectif aux objets (Grill-Spector, Kourtzi, & Kanwisher, 2001; Lerner et al., 2001; Malach et al., 1995). En particulier, trois régions ont été mises en évidence comme étant sélectivement impliquées dans la perception de scènes : la « parahippocampal place area » (PPA), le cortex retrosplénial (dont l'acronyme anglais est RSC pour « retrosplenial cortex »), et, plus récemment, « l'occipital place area » (OPA).

Après une présentation succincte des travaux qui montrent que ces régions assureraient des fonctions cognitives complexes telles que la navigation dans l'espace, le traitement de l'agencement spatial des objets dans une scène ou le traitement d'informations sémantiques liées au contexte entre les objets d'une scène, nous présenterons des travaux plus récents – dont les nôtres – qui montrent que ces régions seraient également sélectives aux caractéristiques visuelles élémentaires des scènes (fréquences spatiales, orientations, contraste de luminance, etc.).

1. Rôle des régions sélectives aux scènes

1.1. Parahippocampal place area ou PPA

La PPA est une région du cortex parahippocampique, situé entre la partie postérieure du gyrus parahippocampique et la partie antérieure du gyrus lingual (cf. Figure 25). Elle a été mise pour la première fois en évidence dans une étude en IRMf de Epstein et Kanwisher (1998). Dans cette étude les auteurs présentaient des images de scènes, de maisons, de visages

et d'objets aux participants qui avaient pour tâche de détecter l'apparition successive de deux images identiques. Le contraste entre les activations cérébrales induites par la perception visuelle de scènes et celle des autres stimuli a révélé une activation plus importante de la PPA. Ces auteurs ont également montré que la PPA répondait plus fortement à des images de scènes contenant des objets ou à des images de scènes vides (i.e. ne contenant pas d'objets), qu'à des images de scènes ne contenant que les objets, sans le contexte spatial de la scène, ou qu'à des images de scènes dont les éléments du contexte spatial (e.g., le sol, la ligne d'horizon) étaient fractionnés et réorganisés. Ces résultats ont amené les auteurs à formuler l'hypothèse que la PPA serait spécifiquement impliquée dans l'encodage de l'agencement spatial et en trois dimensions de l'information visuelle d'une scène. Les résultats de nombreuses études en IRMf sont ensuite venus étayer cette hypothèse (Epstein, Graham, & Downing, 2003; Epstein et al., 1999; Epstein & Higgins, 2007; Epstein, 2005; Henderson, Larson, & Zhu, 2007, 2008; Henderson, Zhu, & Larson, 2011; Park, Brady, Greene, & Oliva, 2011; Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011). Par exemple, Epstein et al. (2003) ont utilisé dans une étude en IRMf un paradigme d'adaptation en présentant successivement deux scènes (vue d'une table) qui différaient soit en termes d'agencement spatial (i.e. différents points de vue d'une même scène ou différentes scènes contenant les mêmes objets), soit en termes d'objets contenus dans la scène, sans que leur agencement spatial soit altéré. Ce paradigme d'adaptation repose sur le fait que lorsqu'une région cérébrale est adaptée à un stimulus, la réponse de cette région diminue lors de la répétition du stimulus. Les auteurs ont observé que la PPA répondait de façon équivalente aux paires variant en termes d'agencement spatial, suggérant que cette région traitait ces scènes comme des stimuli distincts. En revanche, un effet d'adaptation était observé en réponse aux paires variant seulement en termes d'objets contenus dans les scènes, suggérant que la PPA traitait ces deux images comme des stimuli similaires. Ces résultats indiquent donc que la PPA encoderait l'agencement spatial des scènes, et ce, de façon spécifique au point de vue. D'autres travaux rapportent par ailleurs que la PPA serait impliquée non seulement lors de la perception de scènes, mais également lors de l'imagerie mentale de scènes (O'Craven & Kanwisher, 2000) ou la navigation mentale dans les scènes (Mellet et al., 2000).

Cependant, pour Bar et collaborateurs, le cortex parahippocampique ne serait pas spécifiquement dédié au traitement des scènes, mais répondrait aux éléments entretenant de forts liens contextuels entre eux en terme spatial (e.g., un mixeur associé au contexte d'une scène de cuisine), mais également au-delà du contexte spatial (e.g., un cœur associé à

Cupidon; Aminoff, Gronau, & Bar, 2007; Aminoff, Kveraga, & Bar, 2013; Bar, Aminoff, & Ishai, 2008; Bar, Aminoff, & Schacter, 2008; Bar & Aminoff, 2003; Bar, 2004). Ils montrent notamment que la partie postérieure du cortex parahippocampique, correspondant à la PPA décrite par Epstein et Kanwisher (1998), serait spécifiquement impliquée dans le traitement d'associations contextuelles spatiales tandis que la partie antérieure serait plus impliquée dans le traitement d'associations contextuelles non spatiales (Aminoff et al., 2007). Ces travaux viennent donc modérer le point de vue selon lequel la PPA serait dédiée exclusivement à la perception des scènes, en suggérant plus largement son implication dans l'extraction d'informations sémantiques liées au contexte spatial dans l'environnement visuel.

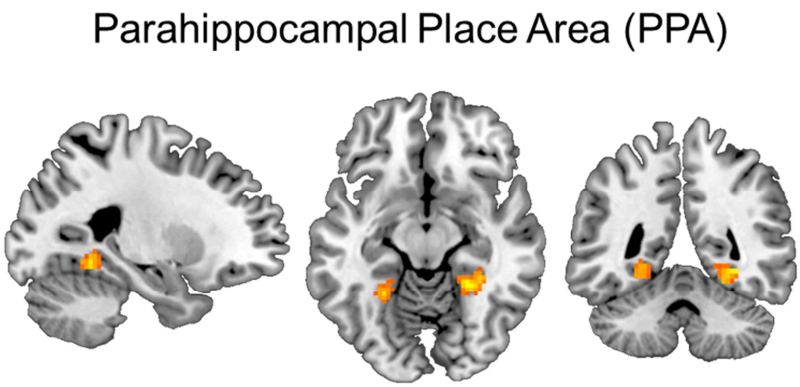


Figure 25 : Localisation anatomique de la PPA sur des coupes sagittale, axiale et coronale.

1.2. Cortex retrosplenial ou RSC

Le cortex retrosplénial ou « retrosplenial cortex » (RSC) est une région du gyrus cingulaire postérieur (cf. Figure 26) connue pour être impliquée dans la navigation et la mémoire spatiale (Aggleton & Vann, 2004; Epstein, 2008; Harker & Whishaw, 2004; Vann, Aggleton, & Maguire, 2009). Les études contrastant l'activité cérébrale induite par la perception de scènes à celle de visages ou d'objets ont fréquemment rapporté une activation plus importante du RSC pour les scènes, conjointe à celle de la PPA (Bar & Aminoff, 2003; Bar, 2004; Epstein et al., 2003; Epstein & Higgins, 2007; Nasr et al., 2011; O'Craven & Kanwisher, 2000; Park & Chun, 2009; Park, Intraub, Yi, Widders, & Chun, 2007; Sung, Kamba, & Ogawa, 2008; Walther, Caddigan, Fei-Fei, & Beck, 2009). Cependant, alors que la PPA serait impliquée dans l'encodage de l'agencement spatial de scènes individuelles, plusieurs travaux suggèrent que le RSC supporterait des représentations plus globales, permettant d'envisager les scènes dans un environnement spatial plus large (Epstein &

Higgins, 2007; Epstein, 2008; Park, Chun, & Johnson, 2010; Park et al., 2007; Park & Chun, 2009). Par exemple, Park et Chun (2009) ont utilisé un paradigme d'adaptation dans une étude en IRMf en présentant plusieurs versions panoramiques d'une même scène qui variaient selon leur point de vue. Ces auteurs ont observé que la PPA répondait de façon équivalente aux différents points de vue d'une même scène, tandis qu'un effet d'adaptation était observé au sein du RSC en réponse aux différents points de vue. Ces résultats ont donc conforté l'idée que la PPA traiterait les scènes de façon spécifique à leur point de vue (Epstein et al. 2003), tandis que le RSC intègrerait les différents points de vue comme faisant partie d'une même scène et envisagerait les scènes de façon plus globale. Le RSC aurait donc un rôle complémentaire à celui de la PPA pour la perception de scènes et la navigation spatiale.

Selon Bar et collaborateurs (Aminoff et al., 2007, 2013; Bar, Aminoff, & Ishai, 2008; Bar, Aminoff, & Schacter, 2008; Bar & Aminoff, 2003; Bar, 2004), le RSC interviendrait également dans le traitement et l'extraction d'associations contextuelles. Alors que la PPA serait principalement impliquée dans le traitement d'associations contextuelles spatiales, le RSC répondrait aussi bien à des associations contextuelles spatiales que non spatiales.

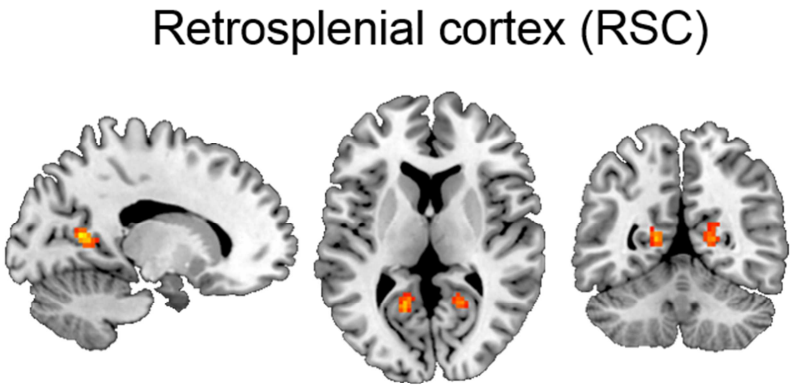


Figure 26 : Localisation anatomique du RSC sur des coupes sagittale, axiale et coronale.

1.3. Occipital place area ou OPA

L' « occipital place area » (OPA) est située au niveau du sillon occipital transverse (cf. Figure 27). L'activation de cette région a plusieurs fois été observée conjointement à celle de le PPA et du RSC lors de la perception de scènes, mais son rôle a été largement moins étudié (Dilks, Julian, Paunov, & Kanwisher, 2013; Epstein, Higgins, Jablonski, & Feiler, 2007; Epstein & Higgins, 2007; Grill-Spector, 2003; Hasson, Harel, Levy, & Malach, 2003; Nasr et

al., 2011). Dans une étude récente, Dilks et al. (2013) ont mis en évidence la sélectivité de cette région pour la perception et la catégorisation de scènes. Dans une première expérience, ces auteurs ont appliqué une stimulation magnétique transcrânienne (TMS) sur l'OPA afin d'inhiber l'activité de cette région pendant que les participants effectuaient une tâche impliquant la détection de différences fines dans l'agencement spatial de deux scènes ou deux visages issus d'un morphing. Les résultats ont montré que l'inhibition de l'OPA engendrait une baisse de performance pour la discrimination des scènes, mais pas pour la discrimination des visages, suggérant une sélectivité au sein de l'OPA pour le traitement de détails dans l'agencement spatial des scènes. Dans une deuxième expérience, la TMS était appliquée sur l'OPA lors de la présentation de scènes (plage, forêt, ville ou cuisine) ou d'objets (appareil photo, chaise, voiture ou chaussure) que les participants devaient catégoriser. Les auteurs ont observé que l'inhibition de l'OPA entraînait plus d'erreurs de catégorisation pour les scènes que les objets, suggérant également une sélectivité de l'OPA pour le processus même de catégorisation de scènes.

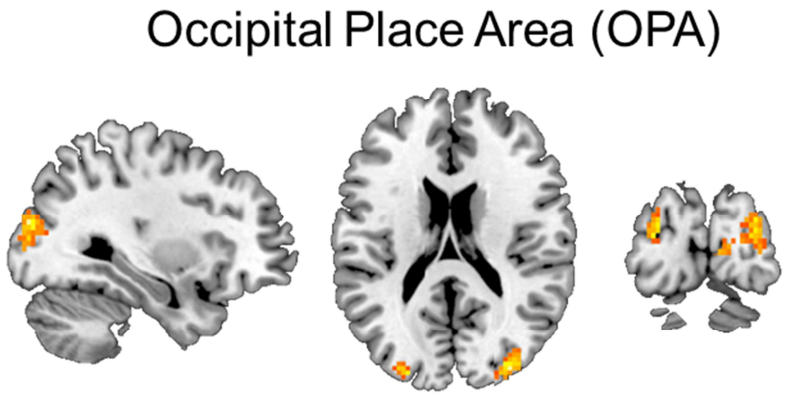


Figure 27 : Localisation anatomique de l'OPA sur des coupes sagittale, axiale et coronale.

L'ensemble des travaux présentés ci-dessus indique donc que la PPA, le RSC et l'OPA assurent des fonctions cognitives complexes et complémentaires lors de la perception de scènes. Alors que la PPA serait principalement impliquée dans l'encodage de la structure spatiale de scènes individuelles, de façon spécifique à leur point de vue, le RSC serait plus largement impliqué dans la navigation dans les scènes, les intégrant ainsi de façon plus globale et indépendante de leur point de vue. Ces deux régions joueraient également un rôle dans la mise en place d'associations contextuelles en terme spatial (PPA et RSC) et non spatial (RSC). Enfin, l'OPA serait impliquée dans le traitement de détails dans l'agencement

spatial des scènes. Dans leur étude, Dilks et al. (2013) ont soulevé la question de savoir quel type d'information est représenté au sein de l'OPA pour permettre de telles fonctions, et si celle-ci encode des informations complexes, telles que la catégorie sémantique des scènes, ou des caractéristiques plus élémentaires telles que le spectre d'amplitude ou les fréquences spatiales dans les scènes. De manière générale, cette question du type d'information représenté dans les régions sélectives aux scènes a jusqu'à présent été peu abordée.

2. Nature de l'information représentée au sein des régions sélectives aux scènes

Ce n'est que très récemment que plusieurs auteurs se sont intéressés à la nature de l'information (e.g., catégorie sémantique, caractéristiques du spectre d'amplitude…) extraite dans les scènes et représentée au sein des régions sélectives aux scènes. La plupart de ces travaux ont utilisé pour cela une méthode d'analyse multivariée des patterns de réponse (dont l'acronyme anglais est « MVPA » pour « multivariate pattern analysis »). Cette méthode permet d'examiner la réponse des aires cérébrales en fonction des conditions expérimentales en termes de configuration spatiale de l'activité cérébrale, et non en termes d'intensité des réponses d'un voxel individuel ou d'un groupe de voxel comme le permettent les contrastes d'activation cérébrale entre les conditions expérimentales (analyses univariées classiques ; voir pour une revue Norman, Polyn, Detre, & Haxby, 2006), tels que ceux utilisés dans les **Expériences 3 et 4**. Cette méthode repose sur un système de classification des patterns de réponses sur la base de classes prédéfinies (e.g., conditions expérimentales), afin de déterminer si les patterns de réponses peuvent être décodés sur la base de ces classes. Prenons l'exemple d'une étude IRMf au cours de laquelle on présente des scènes de ville et des scènes de plage. Le système est entraîné à classifier des patterns de réponses issus d'une partie des acquisitions fonctionnelles, par exemple, sur la base de la catégorie de la scène (e.g., ville ou plage) ayant généré chaque pattern. Une fois le système entraîné, celui-ci est ensuite testé avec un nouveau pattern de réponses (i.e. issu d'autres acquisitions fonctionnelles) n'ayant pas servi à l'entraînement, et doit déterminer quelle catégorie de scène a généré le pattern. Si les performances du classifieur sont au-dessus du hasard, cela suggère que les deux catégories de scènes ont généré des patterns de réponses assez distincts pour que le classifieur puisse les décoder.

En utilisant la méthode de MVPA, Walther et al., (2009; voir aussi Walther et al., 2011) entraînent un système à classifier les patterns de réponse de la PPA et du RSC induits par la vision passive de scènes. Ils utilisent comme classes les catégories des scènes (plages, villes, forêts, autoroutes, usines et montagnes). Les résultats ont montré que le système obtenait des performances au-dessus du hasard pour décoder et distinguer les différentes catégories dans des patterns de réponses non utilisés pour l'entraînement. Cette étude suggère que la catégorie des scènes permet de bien discriminer les patterns de réponses au sein de ces régions et, par extension, que la catégorie de la scène est une information potentiellement représentée au sein de la PPA et du RSC. D'autre études suggèrent néanmoins que ces différents patterns de réponses peuvent être décodés sur la base d'indice spatiaux, structuraux et texturaux des scènes, plutôt que sur la base d'informations de nature sémantique liées aux catégories des scènes (Cant & Goodale, 2011; Cant & Xu, 2012, 2014; Henderson et al., 2007, 2008; Konkle & Oliva, 2012; Kravitz, Peng, & Baker, 2011; Park et al., 2011; Park, Konkle, & Oliva, 2014; Walther et al., 2011). Par exemple, Park et al. (2011) ont spécifiquement étudié si les patterns de réponses au sein de la PPA et de la RSC lors de la catégorisation de scènes pouvaient être décodés sur la base du degré d'ouverture (i.e. degré de visibilité de l'horizon ; scènes ouvertes ou fermées) ou de la catégorie (naturelle ou artificielle) des scènes. En utilisant une méthode d'analyse corrélationnelle basée sur les résultats d'une MVPA, les auteurs ont observé qu'au sein de la PPA et du RSC, les patterns de réponses étaient mieux décodés sur la base du degré d'ouverture que sur la base de leur catégorie. Par exemple, les scènes naturelles ouvertes généraient la même réponse que les scènes artificielles ouvertes, tandis que les scènes naturelles fermées généraient la même réponse que les scènes artificielles fermées. Dans une étude plus récente, Park et al. (2014) ont montré qu'au sein de la PPA, les patterns de réponses pouvaient être aussi bien décodés sur la base de la taille de l'espace (e.g., petits espaces tels qu'une douche, ou grands espaces tels qu'un centre commercial) représenté dans les scènes que sur la base de leur niveau d'encombrement (i.e. quantité d'éléments dans la scène). Au sein du RSC, les patterns de réponses étaient mieux décodés sur la base de la taille de l'espace que du niveau d'encombrement. De façon intéressante, des travaux en simulation computationnelle ont montré que des caractéristiques des scènes telles que leur degré d'ouverture, leur surface, voire leur catégorie, pouvaient être estimées sur la base de leur spectre d'amplitude (Torralba & Oliva, 2003). Les résultats des travaux de Park (e.g., Park et al., 2011, 2014) soulèvent donc la question de savoir si les informations relatives par exemple au degré d'ouverture ou à

la surface des scènes représentées au sein de ces régions pouvaient en fait refléter le traitement de caractéristiques de plus bas niveau telles que le spectre d'amplitude des scènes.

C'est en effet ce que suggèrent plusieurs études qui indiquent que ces régions codent une information de bas niveau dans les scènes, telle que les régularités statistiques dans le spectre d'amplitude (i.e. en termes de distribution du contraste de luminance en fonction des fréquences spatiales et des orientations) spécifiques à différentes catégories de scènes (Andrews, Clarke, Pell, & Hartley, 2010; Nasr & Tootell, 2012; Rice, Watson, Hartley, & Andrews, 2014; Watson, Hartley, & Andrews, 2014). Andrews et al. (2010) ont spécifiquement examiné cette hypothèse dans une étude en IRMf en utilisant la méthode plus classique des contrastes d'activation cérébrale entre deux conditions. Ces auteurs présentaient des images de scènes et de visages, ainsi que des versions de ces stimuli pour lesquelles le spectre de phase (i.e. structure spatiale) était rendu aléatoire, mais dont le spectre d'amplitude était préservé (stimuli « scrambles », cf. Figure 28). Ces auteurs ont montré que la PPA répondait plus fortement aux « scrambles » de scènes qu'aux « scrambles » de visages, suggérant une sélectivité de la PPA aux propriétés de bas niveau caractéristiques des scènes, indépendamment de leur structuration spatiale.

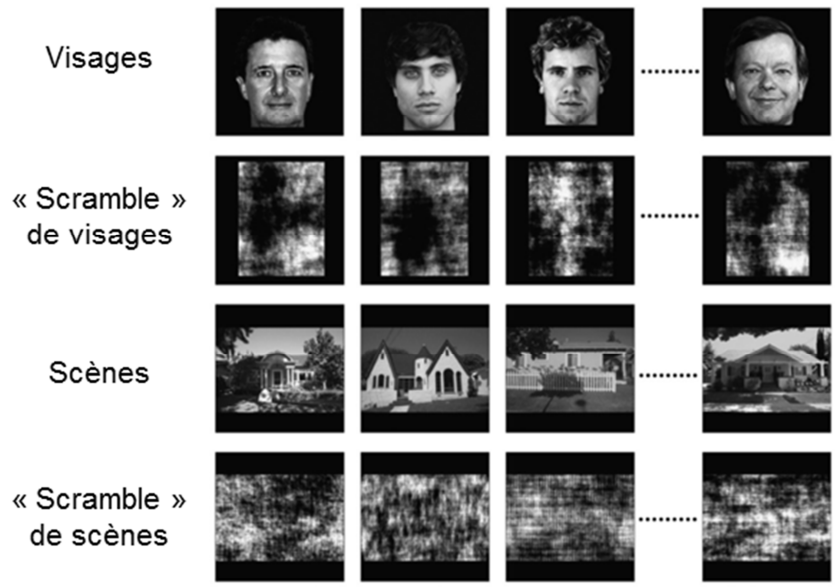


Figure 28 : Exemples de stimuli utilisés dans l'étude d'Andrews et al. (2010).

Dans le contexte des modèles de la perception visuelle qui reposent sur un traitement fréquentiel de l'information visuelle (Bar, 2003; Bar et al., 2006; Bullier, 2001; Peyrin et al., 2010; Schyns & Oliva, 1994), et au vu des travaux mentionnés ci-dessus suggérant que les

régions sélectives aux scènes sont impliquées dans le traitement de caractéristiques de très bas niveau des scènes visuelles (Andrews et al., 2010; Nasr & Tootell, 2012; Rice et al., 2014; Watson et al., 2014), en particulier relatives au spectre d'amplitude, il semble pertinent de se poser la question de l'implication de ces régions pour le traitement des fréquences spatiales lors de la catégorisation de scènes. De façon surprenante, cette question a peu été étudiée et les quelques travaux sur le sujet rapportent des résultats divergents en ce qui concerne la PPA. Par exemple, Peyrin et al. (2004) ont observé en IRMf une activité plus importante du gyrus parahippocampique droit lors de la catégorisation de scènes en BFS qu'en HFS. Cependant, Rajimehr et al. (2011) ont plus récemment montré chez le macaque et l'humain que la PPA répondait plus fortement à des objets et des scènes en HFS qu'en BFS. Par ailleurs, à notre connaissance, le traitement des fréquences spatiales au sein du RSC et de l'OPA n'a jamais été directement étudié. Nous avons donc mené une étude en IRMf dans laquelle nous avons spécifiquement étudié la réponse de la PPA, du RSC et de l'OPA aux fréquences spatiales dans les scènes (**Expérience 5**). A noter que nous nous intéressons tout particulièrement à la bande de fréquences spatiales privilégiée par chaque région. Nous avons donc analysé nos données en utilisant la méthode classique des contrastes entre les activations cérébrales induites par les différentes fréquences spatiales, plutôt que la méthode MVPA.

Dans ce contexte de recherche, nous avons vu précédemment que le traitement des fréquences spatiales se ferait selon une séquence, par défaut, « coarse-to-fine » (Musel et al., 2012; Schyns & Oliva, 1994). On peut alors se demander si une logique de traitement « coarse-to-fine » des fréquences spatiales est également privilégiée au sein des régions sélectives aux scènes. Une sensibilité aux fréquences spatiales, ainsi que la prédominance d'une stratégie d'analyse « coarse-to-fine » a en effet été démontrée au sein d'une région « catégorie-spécifique » de haut niveau, la FFA, connue pour répondre sélectivement aux visages (Goffaux et al., 2011). Dans cette étude en IRMf, les auteurs ont présenté au cours d'une première acquisition fonctionnelle des visages filtrés en BFS, HFS et moyennes fréquences spatiales (MFS), ainsi que des « scrambles » de phase de visages (i.e. aléatorisation du spectre de phase) également filtrés en BFS, HFS et MFS. Les stimuli « scrambles » avait pour but d'éliminer les activations en lien avec les traitements visuels de bas niveau (en soustrayant l'activité correspondant à la perception de visages « scrambles » à l'activité correspondant à la perception de visages intacts) et ainsi, d'examiner spécifiquement les activations en lien avec la construction de représentations de haut niveau. Les participants avaient pour tâche de déterminer si les stimuli étaient des visages ou non. Afin d'étudier le

traitement des fréquences spatiales au cours du temps au sein des régions sélectives aux visages, les auteurs ont manipulé le temps de présentation des stimuli. Les stimuli étaient donc présentés pendant 75 ms, 150 ms ou 300 ms et toujours suivis d'un masque. Les auteurs ont ensuite réalisé une deuxième acquisition fonctionnelle nommée « localizer ». Cette acquisition indépendante de la première avait pour but de localiser les régions sélectives aux visages. Pour cela, les auteurs présentaient des images de visages, de voitures, ainsi que des versions « scrambles » de ces deux types de stimuli. Les participants avaient pour tâche de détecter l'apparition successive de deux images identiques. Le contraste entre les activations induites par les visages intacts et les voitures intactes (contraste [visages > voitures]) permettait de localiser fonctionnellement, chez chaque participant, les régions répondant préférentiellement aux visages, et tout particulièrement la FFA. Les paramètres d'activation correspondant à la catégorisation des visages filtrés (moins celle des « scrambles » de phase de visage) de la première acquisition fonctionnelle étaient ensuite extraits au sein des régions sélectives aux visages identifiées à l'aide du « localizer » chez chaque participant. Les résultats ont montré que lorsque les stimuli étaient présentés 75 ms, la FFA de l'hémisphère droit répondait plus fortement aux visages filtrés en BFS et MFS qu'en HFS. Cependant, lorsque les stimuli étaient présentés plus longtemps (à partir de 150 ms), cette région répondait plus fortement aux visages filtrés en MFS et en HFS qu'en BFS (cf. Figure 29).

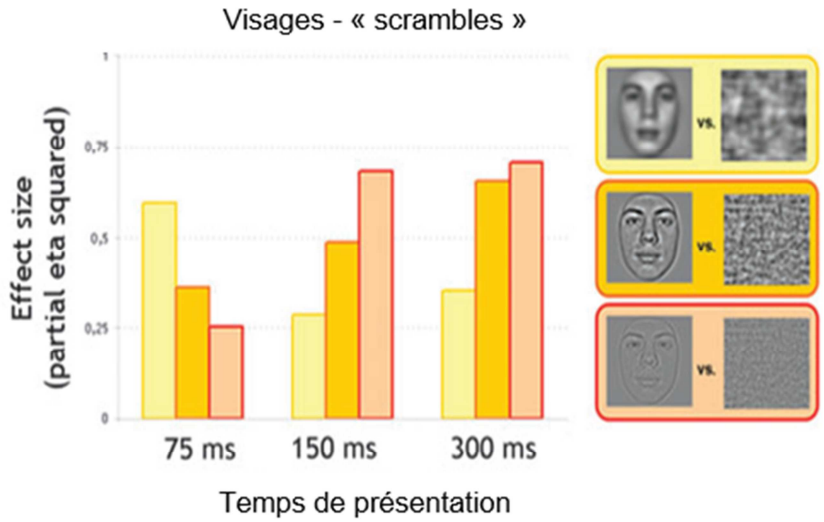


Figure 29 : Résumé des résultats obtenus par Goffaux et al. (2011). Différence de signal entre le traitement des visages intacts et « scrambles » en fonction du contenu en fréquences spatiales et du temps de présentation. On constate que les visages en BFS et en MFS induisent une activité plus importante que les visages en HFS pour des temps de présentation courts (75 ms) et ce pattern s'inverse en faveur des MFS et HFS avec l'augmentation du temps de présentation (dès 150 ms). Adapté de Goffaux et al. (2011).

Ces résultats indiquent donc que la FFA droite privilégierait l'information en BFS pour des temps de présentation courts et l'information en HFS pour des temps de présentation plus longs. Cette étude a donc été la première à mettre en évidence une intégration de type « coarse-to-fine » de l'information visuelle au sein d'une région du cortex inférotemporal sélective aux visages. Elle nous amène également à penser qu'une analyse « coarse-to-fine » pourrait être utilisée par les régions sélectives aux scènes.

Afin d'approfondir nos travaux sur les bases cérébrales du traitement des fréquences spatiales lors de la perception de scènes, nous avons donc spécifiquement étudié à travers deux expériences en IRMf, le traitement des fréquences spatiales (**Expérience 5**) et la catégorisation « coarse-to-fine » des scènes (**Expérience 6**) au sein de la PPA, du RSC et de l'OPA.

3. EXPERIENCE 5 : Traitement des fréquences spatiales et du contraste de luminance au sein des régions sélectives aux scènes

L'**Expérience 5** est détaillée dans l'article ci-après, qui a fait l'objet d'une publication dans la revue *NeuroImage* (Kauffmann et al., 2015). Dans cette étude, nous avons examiné comment la PPA, le RSC et l'OPA participent au traitement des fréquences spatiales lors de la catégorisation de scènes. Par ailleurs, rappelons que les résultats de l'**Expérience 2** suggéraient que l'égalisation du contraste RMS des scènes filtrées pouvait influencer le traitement des fréquences spatiales au niveau comportemental. Nous avons donc également examiné dans cette étude dans quelle mesure l'égalisation du contraste RMS pouvait influencer le traitement des fréquences spatiales au sein de ces régions.

Dans un premier temps, nous avons présenté aux participants des scènes naturelles filtrées en BFS, en HFS ou non filtrées (NF) dans un paradigme en IRMf de type bloc. Dans une première condition (LUM), le contraste de luminance des scènes filtrées et non filtrées n'était pas modifié. Dans une deuxième condition (RMS), le contraste RMS était égalisé entre les scènes filtrées et non filtrées. Les scènes étaient présentées 100 ms et les participant devaient les catégoriser comme étant soit des scènes d'extérieur, soit des scènes d'intérieur. Dans un second temps, nous avons localisé les régions sélectives aux scènes à l'aide d'un « localizer », c'est-à-dire une acquisition fonctionnelle au cours de laquelle les mêmes participants visualisaient des scènes, des visages et des objets et avaient pour tâche de détecter

l'apparition successive de deux images identiques. La comparaison entre l'activité cérébrale induite par la perception de scènes et celle induite par la perception de visages et d'objets (contraste [Scènes > Visages + Objets]) nous a permis d'identifier et de localiser fonctionnellement la PPA, le RSC et l'OPA chez chaque participant. Nous avons ensuite extrait dans ces régions, pour chaque participant, les paramètres d'activation correspondant à la catégorisation des scènes BFS, HFS et NF des conditions LUM et RMS.

Nos résultats (cf. Figure 30) ont montré que la PPA était plus activée par les scènes en BFS et NF que par les scènes en HFS lorsque le contraste RMS n'était pas égalisé, mais répondait plus fortement aux HFS lorsque le contraste était égalisé. La PPA semble donc sensible à l'interaction entre les fréquences spatiales et le contraste de luminance. Concernant le RSC, nous avons observé une activité plus importante en réponse aux BFS qu'aux HFS lorsque le contraste n'était pas égalisé. En revanche, aucune différence significative n'était observée entre les réponses aux différentes fréquences spatiales lorsque le contraste était égalisé. Cela suggère que le RSC est principalement sensible à de grandes différences de contraste de luminance dans les scènes. Enfin, l'OPA répondait plus fortement aux HFS quelle que soit la condition de contraste, suggérant une sélectivité pour les HFS dans cette région.

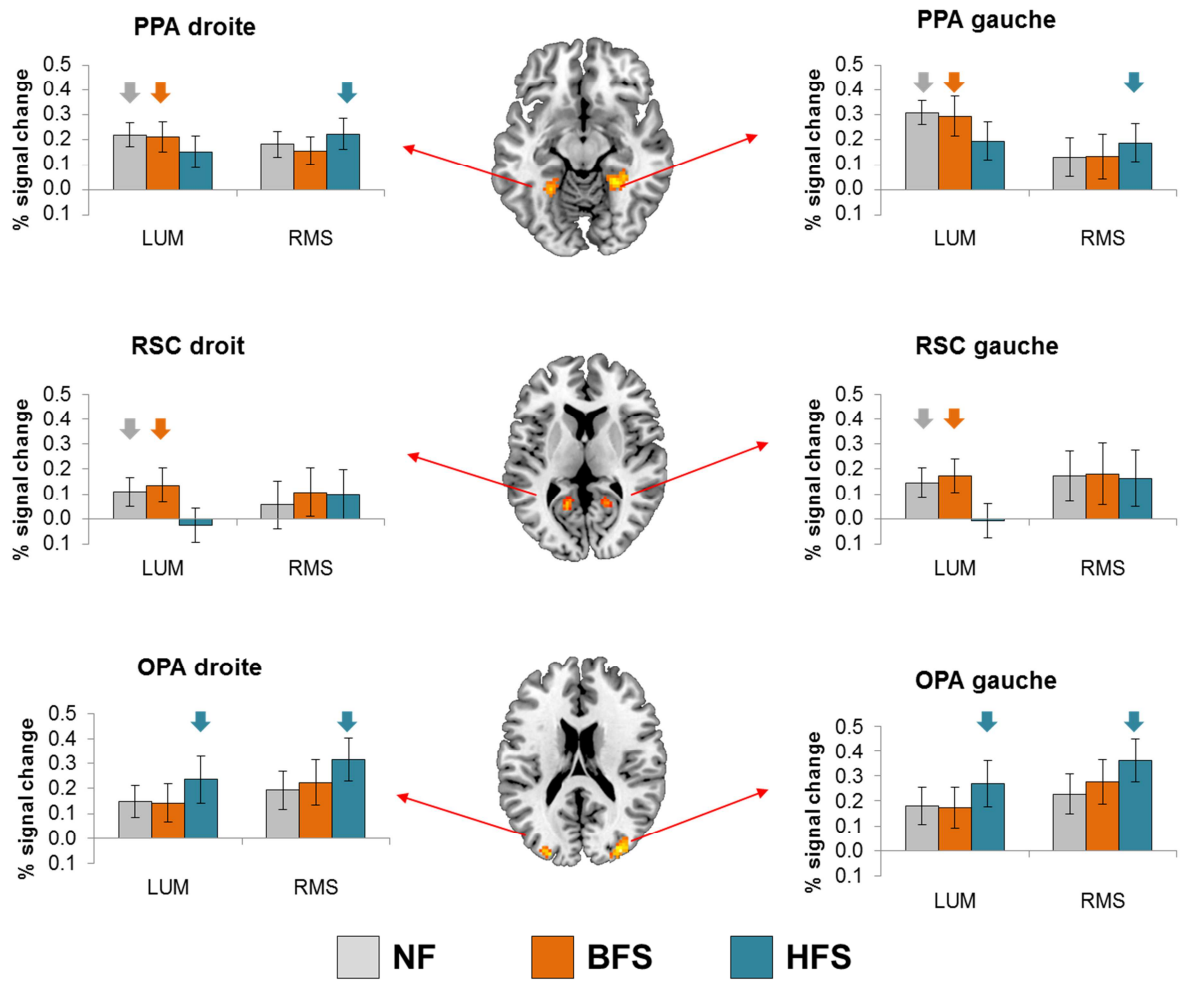


Figure 30 : Résumé des résultats obtenus dans l'Expérience 5. Un « localizer » indépendant nous a permis de localiser fonctionnellement la PPA, le RSC et l'OPA chez chaque participant, et d'extraire au sein de ces régions les paramètres d'activation correspondant à la catégorisation des scènes non filtrées (NF) et filtrées en BFS et HFS des conditions LUM et RMS.

Ces résultats permettent donc de préciser les bases cérébrales du traitement des fréquences spatiales lors de la catégorisation de scènes en démontrant une sensibilité au contenu fréquentiel des stimuli au sein des régions de haut niveau sélectives aux scènes. De plus, nos résultats révèlent des patterns de réponses distincts, suggérant des propriétés fonctionnelles différentes, au sein de chaque région. Par ailleurs, cette étude souligne encore une fois l'importance de considérer le rôle du contraste de luminance pour l'étude des fréquences spatiales, puisque des patterns d'activations très différents sont observés en fonction de l'égalisation ou non du contraste entre les fréquences spatiales au sein de la PPA et du RSC.

Article publié dans la revue *NeuroImage (2015)*

Spatial Frequency Processing in Scene-Selective Cortical Regions

Louise Kauffmann[1,2*], Stephen Ramanoël[1,2,3], Nathalie Guyader[4], Alan Chauvin[1,2], and Carole Peyrin[1,2]

[1] Univ. Grenoble Alpes, LPNC, F-38040 Grenoble
[2] CNRS, LPNC UMR 5105, F-38040 Grenoble
[3] INSERM U836, GIN, F-38706 Grenoble
[4] Univ. Grenoble Alpes, GIPSA-lab, F-38402 Grenoble

Abstract

Visual analysis begins with the parallel extraction of different attributes at different spatial frequencies. Low spatial frequencies (LSF) convey coarse information and are characterized by high luminance contrast, while high spatial frequencies (HSF) convey fine details and are characterized by low luminance contrast. In the present fMRI study, we examined how scene-selective regions - the parahippocampal place area (PPA), the retrosplenial cortex (RSC) and the occipital place area (OPA) - responded to spatial frequencies when contrast was either equalized or not equalized across spatial frequencies. Participants performed a categorization task on LSF, HSF and non-filtered scenes belonging to two different categories (indoors and outdoors). We either left contrast across scenes untouched, or equalized it using a root-mean-square contrast normalization. We found that when contrast remained unmodified, LSF and NF scenes elicited greater activation than HSF scenes in the PPA. However, when contrast was equalized across spatial frequencies, the PPA was selective to HFS. This suggests that PPA activity relies on an interaction between spatial frequency and contrast in scenes. In the RSC, LSF and NF elicited greater response than HSF scenes when contrast was not modified, while no effect of spatial frequencies appeared when contrast was equalized across filtered scenes, suggesting that the RSC is sensitive to high-contrast information. Finally, we observed selective activation of the OPA in response to HSF, irrespective of contrast manipulation. These results provide new insights into how scene-selective areas operate during scene processing.

Keywords: Spatial Frequencies; Root Mean Square contrast; Parahippocampal place area; Retrosplenial cortex; Occipital place area.

1. Introduction

Over the past 20 years, there has been mounting scientific excitement about the perception of scenes containing more realistic and complex stimuli than simple objects or drawings. It is now widely agreed that visual recognition of scenes is a fast, automatic and reliable process. In the case of signal representation, a scene can be expressed in the Fourier domain, in terms of both amplitude and phase spectra (Field, 1987; Ginsburg, 1986; Hughes et al., 1996; Tolhurst et al., 1992). The amplitude spectrum refers to the luminance contrast distribution across spatial frequencies and orientations. Contrast refers here to the magnitude of luminance variation in a stimulus relative to the mean luminance (Robert Shapley & Enroth-Cugell, 1984). On a neurobiological level, it has been established that the cells of the primary visual cortex respond to contrast, spatial frequency and orientation (Boynton, 2005; De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Poggio, 1972; Shams & von der Malsburg, 2002; Shapley & Lam, 1993). According to influential models of visual perception (Bar & Aminoff, 2003; Bullier, 2001; Hegdé, 2008; Kauffmann et al., 2014; Schyns & Oliva, 1994), visual analysis begins with the parallel extraction of different attributes at different spatial frequencies and follows a predominantly "coarse-to-fine" default processing sequence. Low spatial frequencies (LSF) in a scene, conveyed by fast magnocellular visual channels, are thought to activate visual pathways and subsequently access the occipital cortex and high-order areas in the dorsal cortical stream (parietal and frontal) more rapidly than high spatial frequencies (HSF). This permits an initial coarse parsing of visual inputs, prior to their complete propagation along the ventral (inferotemporal) cortical stream, which ultimately mediates object recognition. This initial low-pass visual analysis is believed to guide the subsequent finer analysis of HSF, conveyed more slowly by parvocellular visual channels to the ventral cortical stream.

Interestingly, the ventral cortical stream contains a mosaic of different areas that respond selectively to different categories of visual stimuli (Haxby et al., 2001; Lerner et al., 2001). Three occipito-temporal regions have been shown to be scene-selective (Dilks et al., 2013; Epstein et al., 2007; Epstein & Kanwisher, 1998): the parahippocampal place area (PPA), the retrosplenial cortex (RSC), and the occipital place area (OPA). These regions have been linked to high-order functions during scene perception, such as navigation (Epstein et al., 2007; Vass & Epstein, 2013), spatial layout processing and scene recognition (Dilks et al., 2013; Epstein, 2005, 2008; Epstein, Graham, & Downing, 2003; Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Higgins, 2007; Epstein & Kanwisher, 1998; Epstein & Ward, 2010), and contextual association (Aminoff, Gronau, & Bar, 2007; Bar, 2004, 2007; Bar & Aminoff, 2003; Bar, Aminoff, & Ishai, 2008; Bar, Aminoff, & Schacter, 2009). Interestingly, some studies have shown that scene-selective regions are also sensitive to low-level visual features in scenes, such as statistical regularities in the amplitude spectrum of visual stimuli (Andrews et al., 2010; Watson et al., 2014), dominant cardinal orientations (Nasr & Tootell, 2012), and spatial frequencies (Rajimehr et al., 2011).

The aim of the present study was, therefore, to further investigate the neural bases of spatial frequency processing during scene categorization, by examining how scene-selective regions of the ventral cortical stream analyze spatial frequencies. The involvement of scene-selective regions has not, as yet, been systematically investigated in this theoretical context. It is however a key issue for the development of current neurally-grounded models of visual perception and for obtaining a better understanding of the role played by spatial frequency at high-level stages of visual processing. In a recent study, Rajimehr et al.

(2011) examined the PPA's response to spatial frequencies in scenes and showed that this region was more strongly activated by HSF than by LSF. These authors suggested that during spatial perception and navigation, the PPA relies predominantly on HSF information in order to detect details and the borders of objects. However, other studies have come up with divergent results. In a functional magnetic resonance imaging (fMRI) study, Peyrin, Baciu, Segebarth, & Marendaz (2004) showed greater activation in the right PPA and a number of other regions for the categorization of LSF than of HSF scenes. Similarly, in an event-related brain potential (ERPs) study, Schettino, Loeys, Delplanque, & Pourtois (2011) observed strong activation in the parahippocampal cortex (PHC), including the PPA, during the categorization of LSF scenes. This activation decreased when HSF information was progressively added to LSF scenes, suggesting that the PHC is more sensitive to LSF information. Overall, these results indicated that the PPA is involved in the spatial frequency processing of scenes. Whether or not it responds selectively to a particular spatial frequency content remains unclear. Furthermore, very little evidence is available on spatial frequency processing in the RSC and the OPA.

A number of methodological aspects also need to be taken into consideration when investigating spatial frequency processing. LSF and HSF stimuli are created by using low and high pass filters which attenuate signals for frequencies which are higher and lower, respectively, than the cut-off frequencies applied when filtering. However luminance contrast in scenes decreases as spatial frequency increases, following a $1/f\alpha$ function (Field, 1987). This means that luminance contrast is higher for LSF than for HSF. Importantly, differences in contrast were found to influence cerebral activation. Several fMRI studies have shown, for example, that as the contrast level of stimuli increased, so did activation in the primary visual cortex (Boynton et al., 1996; Goodyear & Menon, 1998; Olman et al., 2004; Rieger et al., 2013). In the studies described previously, differences in contrast may, therefore, partially account for variations in brain activation related to the different spatial frequencies. In order to avoid any confusion between spatial frequency content and luminance contrast in scenes, recent studies have equalized the contrast of filtered stimuli used (see for example, Goffaux et al., 2011; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Mu & Li, 2013; Vlamings, Goffaux, & Kemner, 2009). RMS contrast (root mean square) is the most frequently-used normalization. RMS contrast corresponds to the standard deviation of luminance values and has been shown to be the most reliable indicator of the visibility of broadband filtered images (Bex & Makous, 2002). In their ERP study, Vlamings et al. (2009) used, LSF and HSF filtered faces either with or without contrast equalization between spatial frequencies. On a behavioral level, they found that LSF faces were categorized more rapidly than HSF faces when contrast was not modified. This difference decreased significantly when contrast was equalized across LSF and HSF faces. These results suggest that contrast normalization plays a crucial role in highlighting differences in LSF and HSF processing. However, the effects of equalizing contrast between LSF and HSF have never been directly investigated at neurobiological level in the context of natural scene perception.

The present fMRI study aimed to investigate spatial frequency processing during scene categorization in scene-selective regions, by examining how the PPA, and for the first time the RSC and OPA, process spatial frequency information. Our study also aimed to address methodological issues, by examining the effects of contrast normalization on spatial frequency processing within these regions. In order to do so, we used a categorization task of scenes from two categories (indoors and outdoors) filtered in LSF and HSF, and non-filtered (NF) scenes. In one experimental condition, mean luminance in

each scene was equalized to a fixed value and the contrast level was not modified (LUM experimental condition). In another experimental condition, mean luminance in each scene was again equalized, but this time contrast was also equalized using the RMS contrast normalization (RMS experimental condition). We also examined the possible existence of a relationship between spatial frequency processing in scene-selective regions and the spatial frequency cut-off used to filter scenes. We used a block-design fMRI paradigm in which scenes from each spatial frequency content (3 LSF cut-offs, 3 HSF cut-offs, and NF) and contrast condition (LUM and RMS) were displayed in separate experimental blocks.

We began by identifying scene-selective regions in both hemispheres in each participant using a localizer adapted from previous studies (Bar et al., 2008; Epstein et al., 2003; Epstein & Kanwisher, 1998; Musel et al., in press; Walther, Caddigan, Fei-Fei, & Beck, 2009). Participants were shown grey-scale pictures of scenes, faces, and common objects. The contrast between scenes and other categories was used to localize the regions specifically involved in scene processing. Once localized, we compared the activation elicited by the different spatial frequency and contrast conditions in the areas defined as the PPA, RSC, and OPA. Based on influential models of visual perception which show that scenes are processed in terms of spatial frequencies (Bar, 2003; Kauffmann et al., 2014; Peyrin et al., 2010; Schyns & Oliva, 1994), and on previous studies showing a sensitivity of scene-selective areas to low-level visual features (Andrews et al., 2010; Nasr & Tootell, 2012; Rajimehr et al., 2011; Watson et al., 2014), we hypothesized that these regions respond to spatial frequencies, and we expected to observe different responses to LSF, HSF, and NF scenes in each of these regions. Furthermore, given that these three scene-selective areas have been seen to support different functions during scene perception (e.g., Bar, Aminoff, & Schacter, 2008; Dilks et al., 2013; Epstein & Higgins, 2007), we also hypothesized that response to spatial frequencies would vary in the three regions. We therefore expected to observe different patterns of response to spatial frequency information in the regions concerned. Finally, we hypothesized that spatial frequency processing in these regions would be influenced by luminance contrast equalization. We therefore expected to observe different responses to spatial frequencies in each region depending on the contrast condition.

2. Method

2.1. Participants

Sixteen right-handed participants (9 males; 23 ± 2 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, which was approved by the local ethics committee. All participants were included in two experiments: the Spatial Frequency experiment and the Localizer experiment.

2.2. Stimuli and Procedure in the Spatial Frequency experiment

Stimuli consisted of 20 black and white photographs (256-level grey-scales, 1042x768 pixels) of scenes classified into two distinct categories (10 indoor scenes and 10 outdoor scenes) with a visual angle of 24x18 degrees, ensuring to stimulate both the fovea and the peripheral visual field (Musel et al., 2013). Scenes were displayed in their original version and in their mirrored version (left and right were reversed) in order to avoid any effect of the visual asymmetry of these large scene images. Exemplar from both categories were chosen to ensure similarity of amplitude spectrum and to prevent categorization from being based on this type of visual cue (Guyader et al., 2004), and to avoid contrast energy differences between categories that could interfere with spatial frequency processing. In both categories, images had the same

distribution of energy in spatial frequencies and dominant orientations (as shown by the mean amplitude spectrum of non-filtered natural scenes in each category; Figure 1a). To ensure that the chosen scenes have similar amplitude spectra, we first calculated the mean amplitude spectrum for the 10 indoor scenes (mean AS indoor) and the 10 outdoor scenes (mean AS outdoor). Then, for each scene, we calculated two 2D correlation coefficients, one between the scene's amplitude spectrum and the mean AS indoor, and the other one between the scene's amplitude spectrum and the mean AS outdoor. The mean AS of the category corresponding to the scene of interest was calculated by excluding the scene's amplitude spectrum (i.e. for an indoor scene, the mean AS indoor was calculated based on the 9 remaining indoor scenes, while the mean AS outdoor was calculated based on the 10 outdoor scenes). The 2D correlation coefficient was calculated using the Matlab function "corr2d". A 2 × 2 variance analyse (ANOVA) with the Category of the scene (indoor and outdoor) and the Category of the mean AS (indoor and outdoor) as within-subject factors were conducted on the 2D correlation coefficients. Results show that the 2D correlation coefficients calculated between indoor scenes and the mean AS indoor did not significantly differ from those calculated between indoor scenes and the mean AS outdoor (0.76±0.05 and 0.76±0.05, respectively; $F_{1,18} < 1$). Similarly, the 2D correlation coefficients calculated between outdoor scenes and the mean AS outdoor did not significantly differ from those calculated between outdoor scenes and the mean AS indoor (0.78±0.05 and 0.78±0.04, respectively; $F_{1,18} < 1$).

Outdoor and indoor categories were equivalent in terms of visual cluttering (Subband Entropy measures; see Rosenholtz, Li, & Nakano, 2007). The mean subband entropy was equivalent for outdoors and indoors (2.91 ± 0.15 and 2.91 ± 0.15, respectively; $F_{1,38} < 1$). Stimuli were elaborated using the MATLAB image processing toolbox (Mathworks Inc., Sherborn, MA, USA). The spatial frequency content of scenes was filtered by multiplying the Fourier transformation of original images by Gaussian filters. The standard deviation of the Gaussian filter is a function of the spatial frequency cut-off, for a standard attenuation of 3 dB. For LSF stimuli, we removed spatial frequency content above 0.5, 1, and 2 cycles per degree (cpd; i.e. low-pass cut-offs of 12, 24, 49 cycles per image). For HSF stimuli, we removed spatial frequency information below 3, 6, and 12 cpd (i.e. high-pass cut-offs of 71, 144, and 293 cycles per image). Cut-off frequencies followed a logarithmic scale in order to obtain a better sampling of the amplitude spectrum of natural scenes (see Willenbockel et al., 2010 for a similar procedure). Furthermore, these values were chosen in order to include cut-off frequencies of 2 cpd for LSF and 6 cpd for HSF, as used in Schyns & Oliva (1994)'s pioneer study on spatial frequency processing during scene perception. The spatial frequency content of scenes was not modified for NF stimuli. The resulting images were then normalized to obtain a mean luminance of 0.5 for luminance values of between 0 and 1 (corresponding to a mean luminance of 128 on a grey-level scale). The contrast of filtered scenes was not modified in the LUM condition (see Table 1). In the RMS condition, scenes were normalized to obtain an RMS contrast of 0.1 (i.e. 25.6 on a grey-level scale). We chose a value situated between LSF and HSF contrast values in natural conditions (i.e. 0.21 and 0.04, respectively in the LUM condition), in order to avoid affecting one spatial frequency condition more than another. This resulted in 14 versions of each scene (1 NF, 3 LSF and 3 HSF for the LUM condition and 1 NF, 3 LSF and 3 HSF for the RMS condition, see Figure 1). Stimuli were displayed using E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, USA) and back-projected onto a translucent screen positioned at the rear of the magnet. Participants viewed this screen at a distance of about 222 cm via a mirror fixed on the

head coil. We used a backward mask, to prevent retinal persistence of scenes. In the LUM condition, the mask was built using a 1/f amplitude spectrum and a random phase spectrum, to match the amplitude spectrum of NF scenes in which luminance decreases as spatial frequency increases following a 1/f function (Field, 1987). In the RMS condition, the mask was also built using a 1/f amplitude spectrum and a random phase spectrum. It was then normalized to match the mean luminance and RMS contrast of filtered scene stimuli in the RMS condition (i.e. mean luminance of 0.5 and RMS contrast of 0.1 for luminance values between 0 and 1).

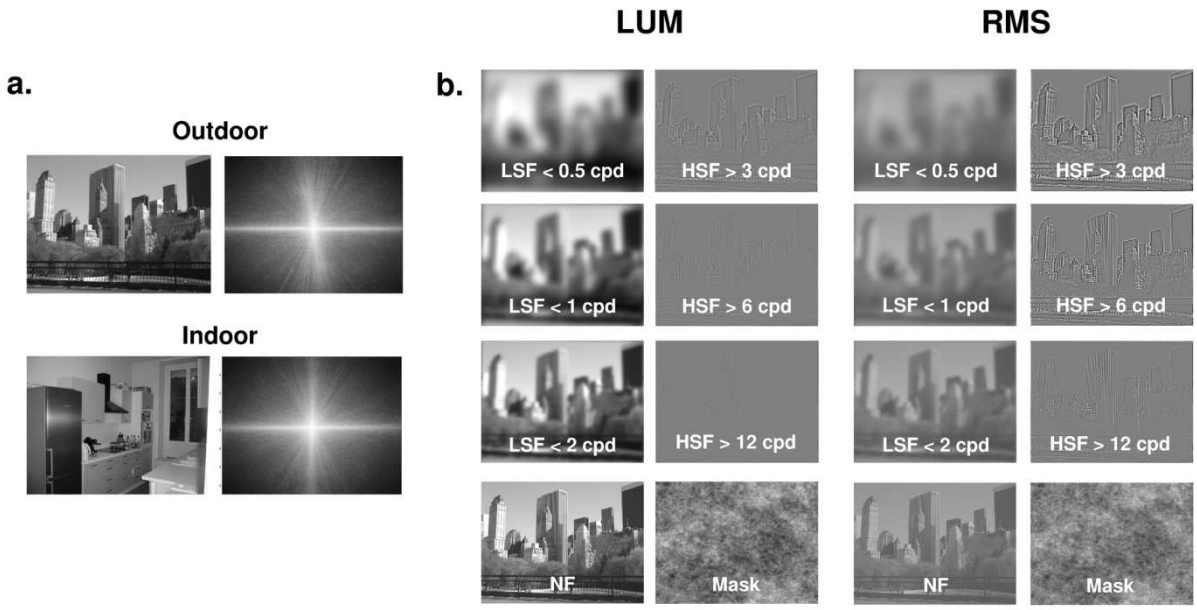


Figure 1: (a) Example of scenes belonging to two different categories (outdoors and indoors) with the mean amplitude spectrum of each category. On the amplitude spectrum images, low spatial frequencies are close to the center and high spatial frequencies are on the periphery. Vertical orientations are represented on the x-axis and horizontal orientations on the y-axis. (b) Example of non-filtered scenes (NF), low-spatial frequency scenes (LSF) below 0.5, 1, and 2 cycles per degree (cpd), and high-spatial frequency scenes (HSF) above 3, 6, and 12 cpd in the LUM (right) and RMS (left) contrast conditions. The masks used and equalized according to the LUM and RMS conditions are also presented (bottom). It should be noted that the perception of spatial frequencies could be affected by the picture reduction of scenes for the illustrative purpose. The picture size is about 10 times smaller than the actual picture size on the screen used for the experiment.

Table 1: Luminance contrast (Mean ± *standard deviation*) for each contrast condition (LUM and RMS), each spatial frequency condition (non-filtered scenes (NF), low-spatial frequency scenes (LSF) below 0.5, 1, and 2 cycles per degree (cpd), and high-spatial frequency scenes (HSF) above 3, 6, and 12 cpd) and each category (indoor and outdoor). Luminance values were normalized in the range [0 1]. The mean luminance value was equal to 0.5 for all stimuli.

	LUM				**RMS**			
Spatial Frequency	**Indoor**		**Outdoor**		**Indoor**		**Outdoor**	
NF	0.21	±*0.03*	0.25	±*0.03*	0.10	±*0.00*	0.10	±*0.00*
LSF < 0.5 cpd	0.21	±*0.04*	0.26	±*0.03*	0.10	±*0.00*	0.10	±*0.00*
LSF < 1 cpd	0.21	±*0.04*	0.26	±*0.03*	0.10	±*0.00*	0.10	±*0.00*
LSF < 2 cpd	0.21	±*0.03*	0.26	±*0.03*	0.10	±*0.00*	0.10	±*0.00*
HSF > 3 cpd	0.04	±*0.01*	0.04	±*0.01*	0.10	±*0.00*	0.10	±*0.00*
HSF > 6 cpd	0.03	±*0.01*	0.04	±*0.01*	0.10	±*0.00*	0.10	±*0.00*
HSF > 12 cpd	0.03	±*0.01*	0.03	±*0.01*	0.10	±*0.00*	0.10	±*0.00*

A block-designed paradigm with the NF, LSF, and HSF scenes was used. The Spatial Frequency experiment consisted of eight functional runs. There were four runs for the LUM condition and four runs for the RMS condition. Each functional run lasted five minutes and was composed of nine 25-second task blocks (one block per spatial frequency cut-off and three blocks of NF scenes), including 10 scenes (5 indoors and 5 outdoors), interspersed with three 25-second blocks with a fixation dot in the center of the screen (Fixation condition) displayed against a grey background. Each scene was presented in all of the spatial frequency conditions within the experiment. The order of scene stimuli was randomized between spatial frequency blocks and runs across participants. It should be noted that a block-design paradigm did not allow us to analyze individual response to trials as an event-related paradigm would have allowed, and it did not allow us to investigate the neural correlates of priming effect in our study. Each stimulus was displayed for 100 ms followed by a mask for 30 ms and a fixation dot in the center of the screen. Stimulus exposure lasted between 75 and 150 ms, in order to avoid favouring LSF or HSF processing, respectively (see Goffaux et al., 2011; Schyns & Oliva, 1994) and was consistent with the exposure duration of filtered scenes used in previous studies (e.g., Peyrin et al., 2010, 2004) which reported no bias towards LSF or HSF processing. The interval between the onsets of two successive stimuli was 2.5 seconds. The participants had to give a categorical answer on the scenes ("indoors" or "outdoors") by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. They were instructed to fixate on the center of the screen (fixation dot) during the entire run and to respond as quickly and as accurately as possible by pressing one of two response buttons. Half of the participants had to answer "indoors" with their forefinger and "outdoors" with the middle finger, while the second half of the participants had to answer 'indoors" with their middle finger and "outdoors" with the forefinger. Response accuracy and reaction times (RT, in milliseconds) were recorded.

2.3. Stimuli and Procedure in the Localizer experiment

Following the main Spatial Frequency experiment, we performed a separate functional Localizer experiment to localize the functional regions of interest specifically involved in the processing of natural scenes. The localizer experiment was adapted from previous studies (Bar et al., 2008; Epstein & Kanwisher, 1998; Walther

et al., 2009). Participants viewed greyscale photographs of scenes, faces and common objects in a block design paradigm. Scene pictures used in the localizer experiment were not shown during the Spatial Frequency experiment. Stimuli were black and white photographs (256 grey-scales), all sized to 700x700 pixels (or 16.4x16.4 degrees of visual angle). The Localizer experiment consisted of one functional run. The functional run lasted three minutes and was composed of eight 15-second task blocks (two blocks of faces, two blocks of objects and four blocks of scenes), including 15 different photographs of the same type, interspersed with four 15-second blocks with a fixation dot in the center of the screen displayed against a grey background. Participants performed a "one-back" repetition detection task. They were instructed to press a button whenever they saw two identical stimuli repeated. Only two repetitions per block were presented. Each stimulus was presented for 300 ms, with a 700 ms interstimulus interval with a fixation dot in the center of the screen. For each participant, the PPA, the RSC and the OPA were identified in both hemispheres by a [scenes > objects and faces] contrast.

2.4. fMRI Acquisition

Experiments were performed using a whole-body 3T Philips scanner (Philips Medical Systems, Eugene, OR) with a 32-channel head coil at the University Hospital Center of Grenoble in France. For all functional Spatial Frequency and Localizer scans, the manufacturer-provided gradient-echo/T2* weighted EPI method was used. Thirty-nine adjacent axial slices parallel to the bi-commissural plane were acquired in interleaved mode. Slice thickness was 3 mm. The in-plane voxel size was 2.5×2.5 mm (220×220 mm field of view acquired with a 88×85 pixel data matrix; reconstructed with zero filling to 96×96 pixels). The main sequence parameters were: TR = 2.5 s, TE = 30 ms, flip angle = 80°. Finally, a T1-weighted high-resolution three-dimensional anatomical volume was acquired, by using a 3D T1 TFE sequence a 3D Modified Driven Equilibrium Fourier Transform (MDEFT) sequence (field of view = 256×224×175 mm; resolution: 1.333×1.750×1.375 mm; acquisition matrix: 192×115×128 pixels; reconstruction matrix: 288×288×128 pixels).

2.5. Data analysis

Data analysis was performed using the general linear model (Friston et al., 1995) for block designs in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK, www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB 7 (Mathworks Inc., Sherborn, MA, USA). Individual scans were realigned, time-corrected, normalized to the MNI space and spatially smoothed by an 8-mm FWHM (Full Width at Half Maximum) Gaussian kernel. Times-series for each voxel were high-pass filtered (1/128 Hz cutoff) to remove low-frequency noise and signal drift.

For the localizer experiment the fMRI signal in the localizer run was analyzed using single-participant general linear model. For each participant, four conditions of interest (scenes, faces, objects, and fixation) were modeled as four regressors, constructed as box-car functions convolved with a canonical hemodynamic response function. Movement parameters derived from realignment corrections (three translations and three rotations) were also entered into the design matrix as additional factors of no interest. The areas responding to scenes were defined independently for each participant using the [Scenes > Faces + Objects] contrast. Significant voxel clusters on individual t maps were identified using a false-discovery correction at qFDR < .05 to control for the overall false-positive rate (Benjamini & Hochberg, 1995). Scene-selective voxel clusters were located within the PPA, RSC and OPA (see Figure 3a for an illustration on a representative participant). To facilitate comparisons with other studies, a transformation of MNI into Talairach and Tournoux (1988) coordinates was performed using the MNI2TAL function (created by Matthew Brett,

available at www.fil.ion.ucl.ac.uk/spm). As the clusters were generally large and involved several regions of interest, small sphere ROIs (3 mm radius) were created at individual peaks of activation in each scene-selective region and in each hemisphere (cf. Figure 3b). These clusters were selected as ROIs for the data analysis in the Spatial Frequency experiment.

For the Spatial Frequency experiment, 16 conditions of interest (NF, LSF-0.5, LSF-1, LSF-2, HSF-3, HSF-6, HFS-12, and Fixation in the LUM and RMS conditions) were modeled as 16 regressors constructed as box-car functions convolved with a canonical hemodynamic response function. Movement parameters derived from realignment corrections (three translations and three rotations) were also entered into the design matrix as additional factors of no interest, to account for head motion-related variance. Parameter estimates (% signal change relative to fixation periods) of block responses were then extracted from these six sphere ROIs for each participant). The average parameter of activity was calculated for the LSF, HSF, and NF conditions, in both Contrast conditions. These values were submitted to a repeated-measure ANOVA with Region (PPA, OPA and RSC), Hemisphere (Left and right), Spatial Frequency condition (NF, LSF, HSF, and Fixation), and Contrast (LUM and RMS) as within-subject factors. In order to examine whether activity in scene-selective regions was influenced by the spatial frequency cut-offs, we additionally performed two ANOVA, one with Region (PPA, RSC, and OPA), Hemisphere (Left and Right), Spatial Frequency cut-off for LSF (0.5, 1, and 2 cpd) and Contrasts (LUM and RMS) as within-subject factors, and the other one with Region (PPA, RSC, and OPA), Hemisphere (Left and Right), Spatial Frequency cut-off for HSF (3, 6 and, 12 cpd) and Contrasts (LUM and RMS).

3. Results

3.1. Behavioral results

Two 2 × 3 × 2 variance analyses (ANOVA) with Spatial Frequencies (LSF, HSF, NF), Contrasts (LUM and RMS), and Category (indoor and outdoor) as within-subjects factors were conducted on the mean error rates (mER) and mean correct reaction times (mRT). The ANOVA conducted on mER revealed a main effect of Spatial Frequencies ($F_{2,30} = 27.78$, $p < 0.05$) and a main effect of Contrast ($F_{1,15} = 34.4$, $p < 0.05$). Notably, there was significant interaction between Spatial frequency and Contrast ($F_{2,30} = 45.03$, $p < 0.05$). Post-hoc comparisons revealed that that for the LUM condition, participants made more errors when categorizing HSF scenes (16.41 ± 8.96%) than when categorizing NF scenes (2.71 ± 5.11%, $p < 0.05$) and LSF scenes (4.74 ± 6.61, $p < 0.05$), and there was no difference between LSF and NF scenes ($p = 0.51$). For the RMS condition, there was no effect of spatial frequencies (NF: 3.54 ± 4.60%; LSF: 4.48 ± 5.04%; HSF: 3.28 ± 3.87; NF vs. HSF: $p = 0.99$; LSF vs. HSF: $p = 0.90$; NF vs. LSF: $p = 0.96$). Furthermore, mER were significantly higher for LUM scenes than RMS scenes only for HSF scenes (16.41 ± 8.96% and 3.28 ± 3.87%, respectively, $p < .05$). There was a significant interaction between Spatial Frequencies, Contrast, and Category ($F_{2,30} = 5.42$, $p < 0.05$). Post-hoc comparisons revealed that participants made more errors for categorizing indoors (21.67 ± 14.17 %) than outdoors (11.15 ± 7.81 %) only for HSF scenes of the LUM condition ($p < 0.05$; all other comparisons: $p > 0.05$).

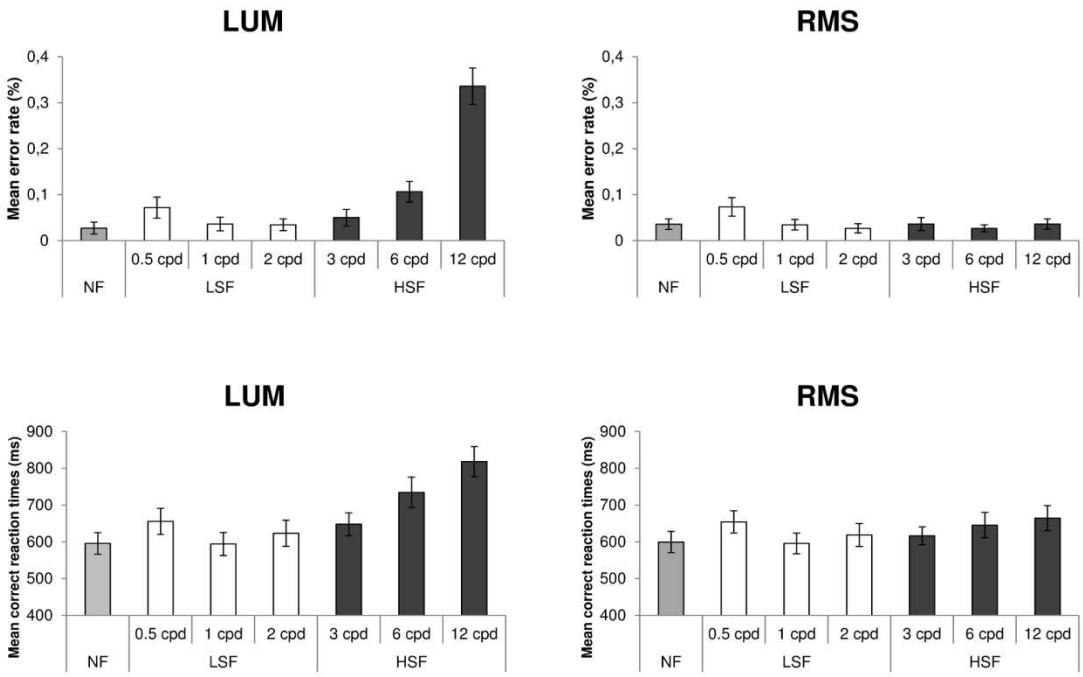


Figure 2: Mean error rates (%) and mean correct reaction times (in milliseconds) to categorize the non-filtered scenes (NF), low-spatial frequency scenes (LSF) below 0.5, 1, and 2 cycles per degree (cpd), and high-spatial frequency scenes (HSF) above 3, 6, and 12 cpd in the LUM (right) and RMS (left) contrast conditions. Error bars correspond to the standard error.

The ANOVA conducted on mRTs again revealed a main effect of Spatial frequency ($F_{2,30}$ = 47.55, $p < 0.05$) and a main effect of Contrast ($F_{1,15}$ = 18.25, $p < 0.05$). Again, there was a significant interaction between Spatial frequency and Contrast ($F_{2,30}$ = 28.15, $p < 0.05$). Post-hoc comparisons showed that for the LUM condition, participants categorized NF scenes (596 ± 118 ms) more quickly than they did for LSF scenes (624 ± 134 ms, $p < 0.05$) and HSF scenes (733 ± 141 ms, $p < 0.05$), and they categorized LSF scenes more quickly than they did for HSF scenes ($p < 0.05$). Similarly, for the RMS condition, participants categorized NF scenes (599 ± 116 ms) more quickly than they did for LSF scenes (623 ± 116 ms, $p < 0.05$) and HSF scenes (642 ± 121 ms, $p < 0.05$), and they categorized LSF scenes more quickly than they did for HSF scenes ($p < 0.05$). Furthermore, mRT were significantly longer for LUM scenes than for RMS scenes, only for HSF scenes (733 ± 141 ms and 642 ± 121 ms, respectively, $p < .05$). There was no main effect of Category, or interaction with the other factors (all ps > 0.20).

We also tested whether the spatial frequency cut-off for the LSF and HSF conditions could influence mER and mRT. Separate ANOVAs were conducted for the LSF and HSF stimuli with Spatial frequency cut-off (either 0.5, 1, and 2 cpd for LSF scenes or 3, 6, and 12 cpd for HSF scenes) and Contrasts (LUM and RMS). For the LSF scenes, there was a main effect of the cut-off on both mER ($F_{2,30}$ = 11.21, $p < 0.05$) and mRT ($F_{2,30}$ = 17.52, $p < 0.05$), but this factor did not interact with Contrast (all $F_{2,30} < 1$). Post-hoc comparisons performed irrespective of contrast showed that participants made more errors as the cut-off for the LSF information decreased (0.5 cpd:7.27 ± 8.15%; 1 cpd:3.52 ± 5.21%; 2 cpd:3.05 ± 4.28%; 0.5 vs. 1 cpd: $p < 0.05$; 0.5 vs. 2 cpd: $p < 0.05$; 1 vs. 2 cpd: $p < 0.05$), and that they were faster for 1 and 2 cpd than for 0.5 cpd (0.5 cpd: 655 ± 130 ms; 1 cpd: 595 ± 118 ms; 2 cpd: 621 ±132 ms; 0.5 vs. 1 cpd: $p < 0.05$; 0.5 vs. 2 cpd: $p <$

0.05; 1 vs. 2 cpd: $p < 0.05$). For the HSF scenes, there was a main effect of the cut-off on both mER ($F_{2,30} = 40.63$, $p < 0.05$) and mRT ($F_{2,30} = 25.63$, $p < 0.05$), and this factor interacted with Contrast on both mER ($F_{2,30} = 55.45$, $p < 0.05$) and mRT ($F_{2,30} = 14.12$, $p < 0.05$). Post-hoc comparisons showed that participants made more errors as the cut-off for the HSF information increased for the LUM condition (3 cpd: 5.00 ± 7.19%; 6 cpd: 10.63 ± 9.11%; 12 cpd: 33.59 ± 15.76%; 3 vs. 6 cpd: $p = 0.08$; 3 vs. 12 cpd: $p < 0.05$; 6 vs. 12 cpd: $p < 0.05$), but no effect was found for the cut-off of the RMS condition (3 cpd: 3.59 ± 5.55%; 6 cpd: 2.66 ± 3.09%; 12 cpd: 3.59 ± 4.47%; all $p > 0.90$). Similarly, participants were slower as the cut-off for the HSF information increased for the LUM condition (3 cpd: 648 ± 124 ms; 6 cpd: 731 ± 156 ms; 12 cpd: 818 ± 166 ms; 3 vs. 6 cpd: $p < 0.05$; 3 vs. 12 cpd: $p < 0.05$; 6 vs. 12 cpd: $p < 0.05$) and tendentially for the RMS condition (3 cpd: 616 ± 98 ms; 6 cpd: 645 ± 137 ms; 12 cpd: 664 ± 135 ms; 3 vs. 6 cpd: $p = 0.38$; 3 vs. 12 cpd: $p < 0.05$; 6 vs. 12 cpd: $p = 0.80$).

3.2. fMRI results

The PPA, RSC and OPA ROIs were defined in each individual, based on the independent Localizer experiment. This served as the structural constraint for the analysis of the data in the Spatial Frequency experiment, in which the categorization of NF, LSF, and HSF scenes was examined. Using the contrast in which scenes were compared with faces and common objects, greater activity elicited by the scenes was observed in different scene-selective regions. Scenes elicited stronger activation than did faces and objects ([Scenes > Faces + Objects] contrast; see Figure 3a for an illustration on a representative participants), within the bilateral parahippocampal gyrus (including the PPA), cingulate gyrus (including the RSC) and the occipital gyrus (in the OPA). These three regions were activated for all participants in the two hemispheres. Peak coordinates of the ROIs were consistent with the previous studies (Talairach coordinates are reported in Table 2). Parameter estimates (% signal change relative to fixation periods) of block responses were then extracted from these six sphere ROIs for each participant. The average parameter of activity was calculated for each experimental condition. These values were submitted to a repeated-measure ANOVA with Region (PPA, RSC and OPA), Hemisphere (Left and Right), Spatial frequency (NF, LSF and HSF) and Contrast (LUM and RMS) as within-subject factors. It should be noted that the three spatial frequency cut-offs used for LSF and HSF scenes were grouped together for this ANOVA.

We first examined whether the three ROIs were differentially sensitive to spatial frequencies in scenes and to luminance contrast equalization by testing the interaction between the ROIs, Spatial frequencies and Contrast. The ANOVA revealed a significant interaction between the ROIs and Spatial Frequency ($F_{4,60} = 24.49$, $p < 0.05$). No interaction was observed between ROIs and Contrast ($F_{2,30} = 1.82$, $p = 0.18$). However, there was a significant Region × Spatial Frequency × Contrast interaction ($F_{4,60} = 2.85$, $p < 0.05$). We further explored the triple interaction by testing the effects of Spatial frequencies, Contrast, and their interaction for each ROI separately. Post-hoc comparisons (Tukey's HSD) were used to perform pairwise comparisons.

Concerning first the PPA, there was no effect of the spatial frequency content ($F_{2,30} = 1.75$, $p = 0.19$) but there was a significant Spatial Frequency × Contrast interaction ($F_{2,30} = 15.78$, $p < 0.05$), suggesting an influence of contrast equalization on spatial frequency processing within the PPA. Post-hoc comparisons (Tukey's HSD) showed that the NF and LSF scenes elicited greater activation than did the HSF scenes for the LUM condition, (NF vs. HSF: $p < 0.05$; LSF vs. HSF: $p < 0.05$; NF vs. LSF: $p = 0.96$). This suggests that the PPA is selective to LSF information in scenes when contrast differs between LSF

and HSF. However, HSF scenes elicited significantly greater activation than did the LSF scenes in the RMS condition (NF vs. HSF: $p = 0.61$; LSF vs. HSF: $p < 0.05$; NF vs. LSF: $p = 0.69$). This may suggest that the PPA is sensitive to contrast in scenes and is selective to HSF information in absence of contrast differences between spatial frequencies. Hemispheres did not interact with either the Spatial frequency ($F_{2,30} = 2.61$, $p = 0.09$), the Contrast ($F_{1,15} < 1$) or their interaction ($F_{2,30} < 1$).

Concerning the RSC, there was a main effect of Spatial Frequencies ($F_{2,30} = 5.12$, $p < 0.05$, but this factor interacted with Contrast ($F_{2,30} = 10.41$, $p < 0.05$) suggesting an influence of contrast equalization on spatial frequency processing within the RSC. Post-hoc comparisons (Tukey's HSD) showed that the NF and LSF scenes elicited greater activation than did the HSF scenes for the LUM condition, (NF vs. HSF: $p < 0.05$; LSF vs. HSF: $p < 0.05$; NF vs. LSF: $p = 0.50$). This may suggest that the RSC is selective to LSF information in scenes. However, for the RMS condition, there was no effect of spatial frequencies (all ps > 0.05). These results thus suggest that in the LUM condition, the different responses to the different spatial frequencies were mainly driven by differences in contrast between spatial frequencies. Hemispheres did not interact with either the Spatial Frequency ($F_{2,30} < 1$), the Contrast ($F_{1,15} < 1$) or their interaction ($F_{2,30} = 2.00$, $p = 0.15$).

Finally, concerning the OPA, HSF scenes elicited greater activation than did NF and LSF scenes (NF vs. HSF: $p < 0.05$; LSF vs. HSF: $p < 0.05$; NF vs. LSF: $p = 0.24$). The Spatial Frequency × Contrast interaction was not significant ($F_{2,30} < 1$). Again, the hemispheres did not interact with either the Spatial frequency ($F_{2,30} < 1$), the Contrast ($F_{1,15} < 1$) or their interaction ($F_{2,30} < 1$). These results suggest that the OPA is selective to HSF information in scenes.

As with the behavioral data, we tested whether the spatial frequency cut-off for the LSF and HSF conditions could influence the activation observed within the ROIs. Two ANOVA were conducted, one with Region (PPA, RSC, and OPA), Hemisphere (Left and Right), Spatial Frequency cut-off for LSF (0.5, 1, and 2 cpd) and Contrasts (LUM and RMS) as within-subject factors, and the other one with Region (PPA, RSC, and OPA), Hemisphere (Left and Right), Spatial Frequency cut-off for HSF (3, 6 and, 12 cpd) and Contrasts (LUM and RMS). There was no effect of the spatial frequency cut-off for either region. However, the number of blocks for each spatial frequency cut-off was maybe too small (only 4 per spatial frequency cut-off in the whole experiment) for producing robust statistics.

To sum up, these results indicate that the PPA is sensitive to spatial frequencies in scenes, but also to the contrast associated with the different spatial frequencies. This region responds more strongly to LSF associated with a higher contrast than HSF, but is selective to HSF in absence of contrast difference between spatial frequencies. This therefore suggests that the contrast normalization could change the sensitivity of the PPA to spatial frequency processing, and that PPA activity relies on the interaction between spatial frequencies and contrast processing. Furthermore, results in the RSC suggested that this region is mainly sensitive to contrast differences in scenes, rather than to spatial frequencies. Finally, the OPA was found to be selective to HSF information in scenes.

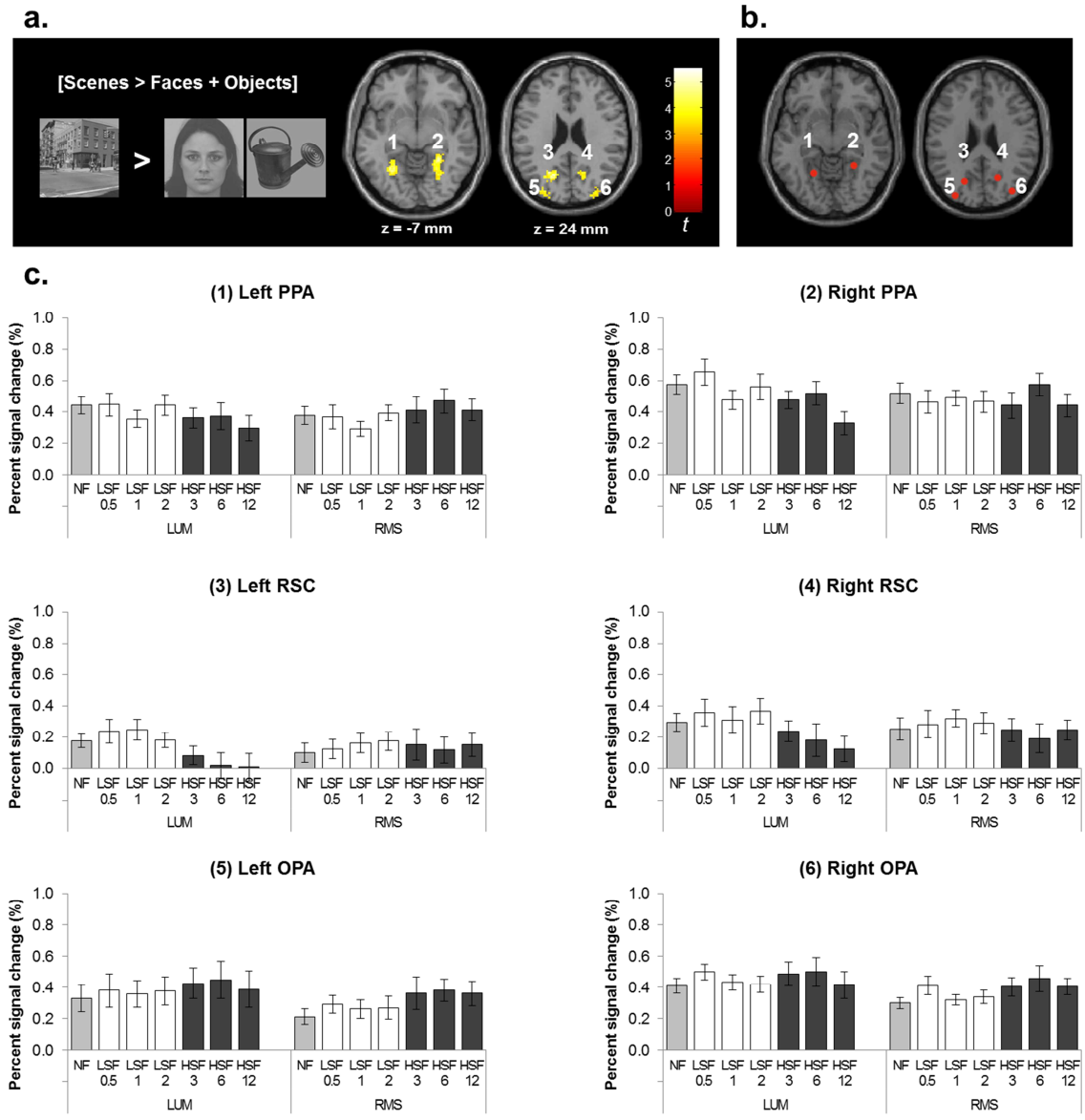


Figure 3: Cerebral regions activated during the perception of scenes compared to faces and objects ([Scenes > Faces + Objects] contrast): (1) left parahippocampal place area (PPA), (2) right PPA, (3) lefts retrosplenial cortex (RSC), (4) right RSC, (5) left occipital place area (OPA), (6) right OPA are illustrated on a representative participant. (b) Example of 3-mm spheres centered on the individual activation peaks of each ROI from which parameter estimates are extracted. (c) The ROIs were defined independently for each participant by contrasting scenes to other stimuli: [Scenes > Faces + Objects]. Signal changes relative to fixation periods were then extracted from the scene-selective ROIs for each participant and each Spatial Frequency condition: non-filtered (NF), low-spatial frequency below 0.5, 1, and 2 cpd (LSF 0.5, LSF 1, and LSF, 2, respectively), and high-spatial frequency above 3, 6, and 12 cpd (HSF 3, HSF 6, and HSF 12, respectively) and according to the two Contrast conditions (LUM and RMS). Graphics represent the mean percentage of signal change of the 16 participants for each ROI. Error bars correspond to the standard errors.

Table 2: Mean coordinates (standard deviations are shown in italics) of scene-selective regions as identified by the [Scenes > Faces + Objects] contrast in the Localizer experiment. Coordinates (x, y, z) are indicated in the Talairach space.

	Mean x		**Mean y**		**Mean z**	
Right PPA	26	± *4*	-43	± *6*	-7	± *3*
Left PPA	-24	± *3*	-48	± *5*	-5	± *3*
Right RSC	17	± *3*	-52	± *3*	12	± *4*
Left RSC	-16	± *3*	-55	± *3*	14	± *5*
Right OPA	34	± *3*	-82	± *4*	22	± *3*
Left OPA	-30	± *5*	-84	± *5*	23	± *6*

4. Discussion

Influential models of visual perception (Bar, 2003; Kauffmann et al., 2014; Peyrin et al., 2010; Schyns & Oliva, 1994) assume that scene perception and categorization is based mainly on the analysis of spatial frequency information. In the present fMRI study, we directly investigated the processing of spatial frequencies in high-level regions of the occipito-temporal cortex dedicated to scene perception. In order to do so, we presented scenes filtered in LSF and HSF, as well as non-filtered scenes, to participants who performed a categorization task on those stimuli. We also addressed methodological issues by examining how contrast equalization across spatial frequencies influences spatial frequency processing in scene-selective regions. We examined the effect of contrast normalization across spatial frequencies using two contrast conditions. Contrast was not modified in any of the scenes (NF, LSF, and HSF) in the LUM condition, while in the RMS condition, contrast was equalized between all scenes. Our results revealed, firstly, that when contrast is not modified, all scene-selective regions respond preferentially to a particular type of spatial frequency content. Furthermore, we observed that this selectivity was different in each of the three scene-selective regions. Finally, we showed that some of these regions were sensitive to luminance contrast equalization across spatial frequencies. These results indicate that these high-level regions are differentially involved in the spatial frequency processing of scenes, and suggest the existence of a potential link between low-visual features in complex images and responses in scene-selective regions. Our study also has methodological implications, as it demonstrates the effect of contrast equalization between spatial frequencies on the responses of these regions.

On a behavioral level, results revealed faster categorization of LSF scenes than of HSF scenes in both contrast conditions. Although the temporal precedence of LSF over HSF processing was not the focus of the present study, these behavioral results support the hypothesis of a coarse-to-fine time course of spatial frequency processing in the visual system, suggested in a considerable number of other studies. Psychophysical and computational studies have, for example, demonstrated temporal precedence of LSF over HSF processing using sine wave gratings and spatial frequency filtered scenes (De Cesarei & Loftus, 2011; Hughes, Nozawa, & Kitterle, 1996; Loftus & Harley, 2005; Mermillod et al., 2005; Parker, Lishman, & Hughes, 1992; Schyns & Oliva, 1994). However, the difference between LSF and HSF categorizations was smaller in the RMS condition, than in the LUM condition. This may be because HSF scenes are substantially less visible than LSF scenes in the LUM condition. Differences in contrast might therefore partially account for behavioural differences in LSF and HSF processing. We went on to investigate spatial frequency

processing and the effects of contrast normalization in scene-selective regions.

Firstly, we observed that the PPA exhibited different response patterns to spatial frequencies depending on the contrast condition. This region responded more strongly to LSF and NF scenes, when associated with a higher contrast, than to HSF scenes (LUM condition). However, HSF elicited greater activation than LSF and NF scenes in the absence of any contrast difference across spatial frequencies (RMS condition). These results therefore suggest that contrast normalization could change the sensitivity of the PPA to spatial frequency processing, and that the PPA activity relies on the interaction between spatial frequency and contrast processing. Studies investigating the specific functions of the PPA during scene perception have shown, for example, that this region encodes the geometric structure (i.e. spatial layout) of scenes and responds preferentially to three-dimensional spatial contexts (Epstein & Kanwisher, 1998; Epstein & Ward, 2010; Epstein, 2005, 2008). In their recent study, Rajimehr et al. (2011) showed that the PPA responded selectively to HSF in scenes, and hypothesized that this information is mainly used to detect scene borders and geometrical shapes during spatial perception. It could, however, be argued that the perception of 3D spatial information also requires global perception of the scene, which may be based mainly on LSF processing (Farell, Li, & Mckee, 2004). As our results in the LUM condition suggest, the use of high-contrast LSF information, may indeed be relevant in the context of spatial perception and navigation, in order to build a structured representation of a scene (e.g., by extracting coarse blobs that depict spatial invariants). This is also consistent with data from previous fMRI and ERP studies (Peyrin et al., 2004; Schettino et al., 2011) showing higher levels of response to LSF in the PPA The discrepancy between our results in the LUM condition and those of Rajimehr et al. (2011) could be due to the fact that their study differed from ours in several aspects such as task demands (i.e. detection of the appearance of a small dot in the center of the screen), and exposure duration of stimuli (i.e. 500 ms). Experimental factors of this type are known to influence spatial frequency processing (Goffaux et al., 2011; Oliva & Schyns, 1997; Schyns & Oliva, 1994). Short exposure duration, for example, favors LSF over HSF processing, whereas long exposure duration favors HSF over LSF processing. The fine-grained task and long exposure duration used in their study may therefore have favored HSF information, whereas the categorization task and the short exposure duration (i.e. 100 ms) used in the present study may have favored LSF processing. Importantly, we assume that the PPA analyses spatial frequency information in a flexible way based on the demands of a particular visual process, and that it uses the most relevant spatial frequency content to perform the task efficiently. Therefore, the PPA may use HSF information preferentially when the task requires the processing of fine information such as spatial layout, and scene borders. For tasks requiring rapid scene categorization or the extraction of spatial invariants in the environment, high-contrasted LSF information would be preferentially used.

Results revealed higher levels of response in the RSC region for LSF than for HSF when contrast was not equalized across spatial frequencies. However, this region did not exhibit any differential response to spatial frequencies in the absence of contrast differences across LSF and HSF. This indicates that the RSC is sensitive mainly to high contrast differences. The RSC has been shown to be involved during the perception of scenes (Aminoff et al., 2007; Bar et al., 2008; Epstein & Higgins, 2007; Park, Brady, Greene, & Oliva, 2011; Park & Chun, 2009; Sung, Kamba, & Ogawa, 2008; Walther et al., 2009), in the coding of "spatial layout" and location (Epstein & Higgins, 2007; Vass & Epstein, 2013) and in contextual associations (Aminoff et al., 2007; Bar, Aminoff, & Ishai, 2008). Furthermore, it was recently shown to be particularly

sensitive to permanent landmarks (e.g. buildings, traffic lights) which represent the most stable features in the environment, compared to non-permanent landmarks (e.g., a bus or a bicycle; Auger, Mullally, & Maguire, 2012). Again, the extraction of high-contrast information in scenes appears to be relevant in the detection of this kind of scene invariant.

Finally, we investigated spatial frequency processing in the OPA. Results for this region revealed a greater level of response to HSF than to LSF scenes, irrespective of the contrast condition. Dilks et al. (2013) recently demonstrated the selective and causal involvement of the OPA in scene categorization and spatial layout processing. In their study, the authors administered transcranial magnetic stimulation (TMS) to the OPA while participants performed either a categorization task with scenes and objects (e.g., participants had to decide if a scene depicted a city or a beach, or if the object which appeared was a car or a shoe), or a discrimination task involving the detection of fine-grained differences in faces and scenes. The authors demonstrated that TMS to the OPA impaired participants' performance only for scene stimuli in both tasks. The authors raised questions about the specific role of the OPA during scene perception, and about whether the OPA represents high-level scene information or extracts more basic perceptual information, such as spatial frequencies or spatial envelope properties (Torralba & Oliva, 2003). They did not, however, manipulate the spatial frequency content of scenes in their study. Our results indicate clearly that the OPA extracts spatial frequency information, and is selectively tuned to HSF and to the details in scenes. This is consistent with the results of Dilks et al. (2013), who also suggested that the OPA was involved in the detection of fine details in scenes. It should be noted that the OPA is situated in the posterior part of the occipital cortex (in the transverse occipital sulcus; Nasr et al., 2011) and corresponds to the projection of the central visual field, according to the retinotopic organization of the visual cortex. Interestingly, Musel et al. (2013) recently showed greater involvement of the posterior part of the occipital cortex related to the fovea in the categorization of HSF than of LSF scenes, suggesting that this area specializes in HSF processing. The right OPA activation (x,y,z: 34, -82, 22) in our study is close to the lateral occipital activation observed by Musel et al. (2013) for HSF scene categorization (x,y,z: 27, -88, 16). OPA selectivity for HSF during scene perception could therefore also be linked to the fact that this region is located in the retinotopic visual areas dedicated to HSF.

Results of the present study may raise the question of the connections between, and relative roles of, scene-selective regions in the processing of low-level visual features in the construction of scene representations. In a recent study, Baldassano, Beck, and Fei-Fei (2013) demonstrated that the PPA exhibits a gradient in connectivity with the other scene-specific regions along the anterior-posterior axis. It does so in a way which suggests that the posterior part of the PPA is more closely connected to the occipital areas (including the OPA) and contributes more to the processing of low level visual features (possibly spatial frequencies and amplitude spectrum properties), while the anterior part of the PPA is more closely connected to the RSC and contributes to the construction of a global scene representation. These results suggest a certain degree of hierarchy in the functions supported by the different scene-selective regions, and that the anterior and posterior PPA have different functional properties. Although the present study was not designed to investigate a potential hierarchical organization of scene-selective regions in the construction of scene representation, our results support the proposal of Baldassano et al. (2013) of specific OPA and PPA involvement in the processing low-level features, such as the spatial frequency, in scenes. In addition, our results indicate that the RSC may also be

involved in the processing of low-level properties, such as contrast, in scenes. Future studies may wish to examine the possible differential involvement of the anterior and posterior parts of the PPA in spatial frequency processing.

Our findings on the methodological issues addressed in the present study also demonstrate that contrast equalization across spatial frequencies drastically influences results in the PPA and the RSC, since opposing patterns of cerebral activation were observed in the different contrast conditions. This specific control therefore has to be taken into account when discussing or comparing studies using spatial frequency filtered stimuli. In particular, although the equalization of RMS contrast to 0.1 in the present study allowed us to avoid affecting one type of spatial frequency content more than another, equalization of the RMS contrast to such a low value should be used with care as it reduces contrast in LSF while enhancing HSF contrast. In doing so, it induces severe modifications in the natural properties of scenes, in which luminance contrast decreases as spatial frequency increases (Field, 1987). Numerous psychophysical and computational studies have observed that the distribution of contrast value across spatial frequencies and orientations in scenes follows statistical regularities which are exploited to categorize scenes (Guyader et al., 2004; Oliva & Torralba, 2001; Torralba & Oliva, 2003). Modification of these regularities via contrast equalization, may, therefore, result in irrelevant analysis of scenes by the visual system and bias behavioral and neurobiological responses. Although we did not observe significant differences between the LUM and RMS conditions for each type of spatial frequency content, it is indeed possible that RMS contrast normalization differentially affected LSF and HSF processing. As can be seen in Figure 1 (see also Table 1 for luminance contrast values), contrast normalization resulted mainly in the reduction of specific high contrast information in LSF, while HSF were much more visible and salient. Further research is needed to clarify the extent to which the association of a particular contrast value to a particular spatial frequency band affects spatial frequency processing during scene perception. Closer examination of how spatial frequency processing is influenced by contrast could be particularly interesting. This could be done by systematically manipulating different levels of luminance contrast as a function of different spatial frequencies Although a large number of psychophysical studies have already investigated contrast detection as a function of spatial frequency using sine-wave gratings as stimuli, there has been no in-depth investigation in the context of scene perception and categorization (e.g., Bex & Makous, 2002; Bex, Solomon, & Dakin, 2009). Future studies on this topic could be of interest in order to gain better understanding of the relative role of spatial frequency and contrast in scene-selective regions.

To conclude, the present fMRI experiment looked at how scene-selective cortical areas, known to be involved in high-level processes during scene perception, process the different spatial frequencies in scenes, and how they are influenced by the equalization of luminance contrast across spatial frequencies. More specifically we investigated how the PPA, the RSC, and the OPA responded to LSF, HSF and non-filtered scenes, in which contrast had either been equalized across scenes or left unmodified We demonstrated greater activation in the PPA in response to LSF and NF scenes than to HSF scenes when contrast was not equalized across spatial frequencies. Activation was, however, selective to HSF in the absence of contrast differences across spatial frequencies. This suggests that PPA activity relies on an interaction between spatial frequency and contrast in scenes. We observed that LSF and NF scenes elicited a greater response in the RSC than HSF scenes when contrast was not equalized across spatial frequencies. No effect of spatial frequency appeared when contrast

was equalized across the filtered scenes. These results suggest that the RSC is sensitive only to high contrast differences. Finally, we found that the OPA responded selectively to HSF, irrespective of the contrast condition.

Overall, the present results allow us to refine the neural bases of spatial frequency processing during scene categorization, by demonstrating a sensitivity to the spatial frequency content of scenes at high-level stages of visual processing, in scene-selective cortical regions. We suggest that this information may contribute to the complex functions supported by these regions during scene perception. The present results therefore provide new insights into how scene-selective areas operate during scene perception and add critical support for influential models of visual perception based mainly on a spatial frequency analysis (Bar, 2003; Kauffmann et al., 2014; Peyrin et al., 2010; Schyns and Oliva, 1994). They also provide a full picture of the use of spatial frequency information in scene-selective cortex and additional data for the development of a neutrally-grounded model of scene perception. Results of the present study may serve as a starting point from which further questions can be asked about the influence of these low-level features on the high-level functions supported by scene-selective regions.

Acknowledgments

This work was supported by the RECOR ANR Grant (ANR-12-JHS2-0002-01 RECOR). Louise Kauffmann was supported by Région Rhône-Alpes (Cible Grants). The authors warmly thank the "Délégation à la Recherche Clinique et à l'Innovation" of the University Hospital of Grenoble and the "Cellule de Neuroimagerie Fonctionnelle" of the Laboratory of Psychology and NeuroCognition for sponsoring. We thank Catherine Dal Molin for the English revision of the manuscript.

References

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, *17*, 1493–503.

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *NeuroImage*, *49*, 703–11.

Auger, S. D., Mullally, S. L., & Maguire, E. a. (2012). Retrosplenial cortex codes for permanent landmarks. *PloS One*, *7*, e43620.

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. *NeuroImage*, *75*, 228–37.

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

Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience*, *5*, 617–29.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*, 280–9.

Bar, M., & Aminoff, E. (2003). Cortical Analysis of Visual Context. *Neuron*, *38*, 347–358. Bar, M., Aminoff, E., & Ishai, A. (2008). Famous faces activate contextual associations in the parahippocampal cortex. *Cerebral Cortex*, *18*, 1233–8.

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience*, *28*, 8539–8544.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289– 300.

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of Optical Society of America*, *19*, 1096–1106.

Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision*, *9*, 1.1–19.

Boynton, G. M. (2005). Contras gain in the brain. *Neuron*, *47*, 476–477.

Boynton, G. M., Engel, S. a, Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, *16*, 4207–21.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research*, *36*, 96–107.

De Cesarei, A., & Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychonomic Bulletin & Review*, *18*, 840–7.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559.

De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 531–544.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of Neuroscience*, *33*, 1331–6a.

Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*, 954– 978.

Epstein, R. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*, 388–396.

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-Specific Scene Representations in Human Parahippocampal Cortex. *Neuron*, *37*, 865–876.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The Parahippocampal Place Area : Recognition , Navigation , or Encoding ? *Neuron*, *23*, 115–125.

Epstein, R., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cerebral Cortex*, *17*, 1680–93.

Epstein, R., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology*, *97*, 3670–83.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 6–9.

Epstein, R., & Ward, E. J. (2010). How reliable are visual context effects in the parahippocampal place area? *Cerebral Cortex*, *20*, 294–303.

Farell, B., Li, S., & Mckee, S. P. (2004). Coarse scales , fine scales , and their interactions in stereo vision. *Journal of Vision*, *4*, 488–499.

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.*, *4*, 2379–94.

Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In *Handbook of Perception and Human Performance* (Ginsburg, ., pp. 1–41). Wiley, New York,.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*, 467–76.

Goodyear, B. G., & Menon, R. S. (1998). Effect of luminance contrast on BOLD fMRI response in human primary visual areas. *Journal of Neurophysiology*, *79*, 2204–7.

Guyader, N., Chauvin, A., Peyrin, C., Hérault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, *327*, 313–318.

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–30.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*, 405–39.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global Precedence, Spatial Frequency Channels, and the Statistics of Natural Images. *Journal of Cognitive Neuroscience*, *8*, 197–230.

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*, 1–14.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, *11*, 287–97.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review*, *12*, 43–65.

Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling. *Brain and Cognition*, *57*, 151–7.

Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental Brain Research.*, *227*, 367–77.

Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Bas, J. Le, & Peyrin, C. (2013). Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. *Journal of Cognitive Neuroscience*, *25*, 1315–1331.

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of Cognitive Neuroscience*, 1–11.

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-selective cortical regions in human and nonhuman primates. *The Journal of Neuroscience*, *31*, 13771–85.

Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual cortex. *The Journal of Neuroscience*, *32*, 14921–6.

Navon, D. (1977). Forest Before Trees: The Precedence of Global Features in Visual Perception. *Cognitive Psychology*, 353–383.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges ? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *107*, 72–107.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope ∗. *International Journal of Computer Vision*, *42*, 145–175.

Olman, C. A., Ugurbil, K., Schrater, P., & Kersten, D. (2004). BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Research*, *44*, 669–683.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *The Journal of Neuroscience*, *31*, 1333–40.

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage*, *47*, 1747–1756.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–60.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *NeuroImage*, *23*, 698–707.

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*, 2768–80.

Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative Ophthalmology*, *11*, 368–77.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology*, *9*, e1000608.

Rieger, J. W., Gegenfurtner, K. R., Koechy, N., Heinze, H., & Grueschow, M. (2013). BOLD responses in human V1 to local

structure in natural scenes : Implications for theories of visual coding, *13*, 1–15.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*, 1–22.

Schettino, A., Loeys, T., Delplanque, S., & Pourtois, G. (2011). Brain dynamics of upstream perceptual processes leading to visual object recognition: a high density ERP topographic mapping study. *NeuroImage*, *55*, 1227–41.

Schyns, P. G., & Oliva, A. (1994). Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*, 195–201.

Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, *42*, 2547–54.

Shapley, R., & Enroth-Cugell, C. (1984). Visual Adaptation and Retinal Gain Controls. *Progress in Retinal Research*, *3*, 263–346.

Shapley, R., & Lam, D. (1993). *Contrast sensitivity*. Retrieved from http://scholar.google.fr/scholar?q=Shapley+RM%2C+Lam+Contrast+sensitivity&btnG= &hl=fr&as_sdt=0%2C5#1

Sung, Y.-W., Kamba, M., & Ogawa, S. (2008). Building-specific categorical processing in the retrosplenial cortex. *Brain Research*, *1234*, 87–93.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic & Physiological Optics_: The Journal of the British College of Ophthalmic Opticians (Optometrists)*, *12*, 229–32.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network (Bristol, England)*, *14*, 391–412.

Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *The Journal of Neuroscience*, *33*, 6133–42.

Vlamings, P. H. J. M., Goffaux, V., & Kemner, C. (2009). Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information ? *Journal of Vision*, *9(5)*, 1–13.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *The Journal of Neuroscience*, *29*, 10573–81.

Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., … Gosselin, F. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology. Human Perception and Performance*, *36*, 122–35.

4. EXPERIENCE 6 : Prédominance d'une stratégie d'analyse « coarse-to-fine » au sein des régions sélectives aux scènes

Les résultats de l'**Expérience 5** indiquent que les régions de haut niveau sélectives aux scènes participent au traitement des fréquences spatiales lors de la catégorisation de scènes. Dans l'**Expérience 6**, nous avons testé en IRMf si ces régions privilégient également un mode de traitement « coarse-to-fine » de l'information en fréquences spatiales. L'**Expérience 6** est détaillée dans l'article ci-après qui a fait l'objet d'une publication dans la revue *Journal of Cognitive Neuorscience* (Musel, Kauffmann et al., 2014).

Pour cela, nous avons dans un premier temps présenté aux participants des séquences dynamiques imitant une analyse « coarse-to-fine » ou « fine-to-coarse » des scènes, adaptées de **l'Expérience 2** et utilisées dans l'**Expérience 3**. Les séquences étaient présentées 150 ms et les participants devaient catégoriser les scènes comme étant soit des scènes d'intérieur, soit des scènes d'extérieur. Dans un second temps, nous avons localisé les aires sélectives aux scènes à l'aide du même « localizer » que dans l'**Expérience 5**. Nous avons ensuite extrait dans ces régions, pour chaque participant, les paramètres d'activation correspondant à la catégorisation des scènes selon une séquence CtF ou FtC.

Les résultats ont révélé que la PPA répondait plus fortement aux séquences CtF que FtC. Cependant, nous n'avons pas observé de différence entre les deux types de séquences au niveau du RSC et de l'OPA (cf. Figure 31). Cela suggère qu'une stratégie d'analyse « coarse-to-fine » semble privilégiée au sein de la PPA. Les résultats de cette étude nous permettent donc de préciser les bases cérébrales de la catégorisation « coarse-to-fine » des scènes.

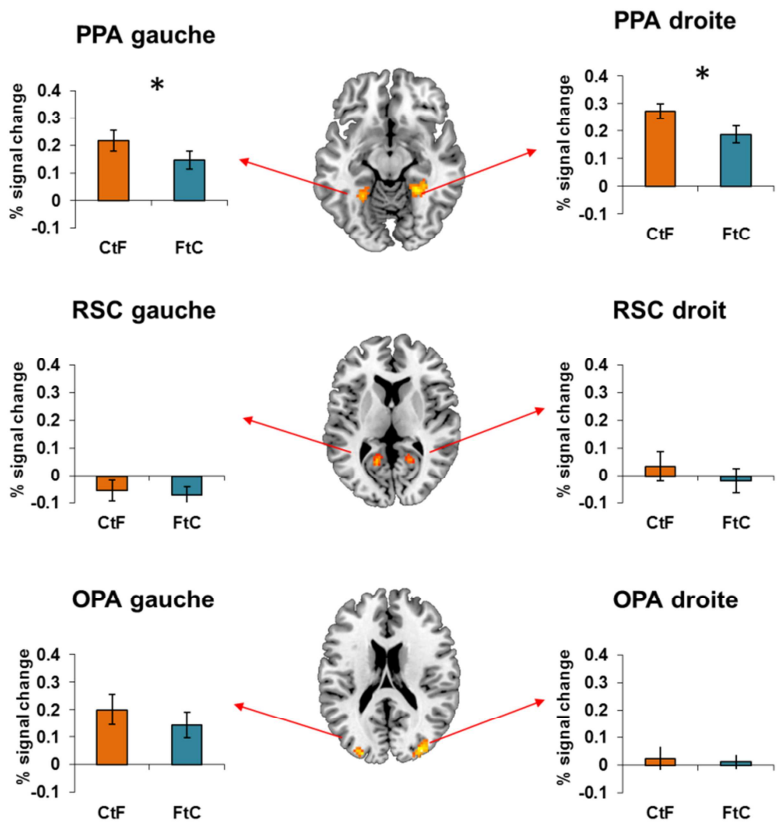


Figure 31 : Résumé des résultats obtenus dans l'Expérience 6. Le « localizer » nous a permis de localiser fonctionnellement la PPA, le RSC et l'OPA chez chaque participant et d'extraire au sein de ces régions les paramètres d'activation correspondant à la catégorisation des scènes selon une séquence CtF ou FtC.

Article publié dans la revue *Journal of Cognitive Neuroscience (2014)*

Coarse-to-fine Categorization of Visual Scenes in Scene-selective Cortex

Benoit Musel[1,2], Louise Kauffmann[1], Stephen Ramanoël[1], Coralie Giavarini[1], Nathalie Guyader[3], Alan Chauvin[1], and Carole Peyrin[1]

[1] Univ. Grenoble Alpes, LPNC, F-38040 Grenoble
[2] University of Geneva
[3] Univ. Grenoble Alpes, GIPSA-lab, F-38402 Grenoble

Abstract

Neurophysiological, behavioral, and computational data indicate that visual analysis may start with the parallel extraction of different elementary attributes at different spatial frequencies and follows a predominantly coarse-to-fine (CtF) processing sequence (low spatial frequencies [LSF] are extracted first, followed by high spatial frequencies [HSF]). Evidence for CtF processing within scene-selective cortical regions is however still lacking. In the present fMRI study, we tested whether such processing occurs in three scene-selective cortical regions: the parahippocampal place area (PPA), the retrosplenial cortex, and the occipital place area. Fourteen participants were subjected to functional scans during which they performed a categorization task of indoor versus outdoor scenes using dynamic scene stimuli. Dynamic scenes were composed of six filtered images of the same scene, from LSF to HSF or from HSF to LSF, allowing us to mimic a CtF or the reverse fine-to-coarse (FtC) sequence. Results showed that only the PPA was more activated for CtF than FtC sequences. Equivalent activations were observed for both sequences in the retrosplenial cortex and occipital place area. This study suggests for the first time that CtF sequence processing constitutes the predominant strategy for scene categorization in the PPA.

Keywords: Natural scene; Spatial frequency; Parahippocampal place area; Retrosplenial cortex; Occipital place area.

Introduction

One of the main functions of the visual system is to categorize the environment. A considerable number of studies on the visual system in humans and animals suggest that spatial frequencies are crucial in the visual categorization of scenes. In terms of signal representation, the image of a scene can be expressed in the Fourier domain as both amplitude and phase spectra (Hughes, Nozawa, & Kitterle, 1996; Tolhurst, Tadmor, & Chao, 1992; Field, 1987, Ginsburg, 1986). The amplitude spectrum decomposes the scene in terms of spatial frequencies and orientations, and the phase spectrum describes the relationship between spatial frequencies. In primates, the primary visual cortex is mainly dominated by simple and complex cells that respond preferentially to orientations and spatial frequencies (Shams & von der Malsburg, 2002; De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Poggio, 1972). In humans, simulations and psychophysical experiments have shown that information from low/medium spatial frequencies is sufficient to allow rapid scene categorization (Guyader, Chauvin, Peyrin, Herault, & Marendaz, 2004; Torralba & Oliva, 2003; Schyns & Oliva, 1994). For example, Schyns and Oliva (1994) used hybrid stimuli made of two superimposed images of natural scenes, taken from different semantic categories and containing different spatial frequencies (e.g., a highway scene in low spatial frequencies superimposed on a city scene in high spatial frequencies). The perception of these hybrid scenes was dominated by low spatial frequency information when presentation time was very brief (30 msec), but by high spatial frequency information when presentation time was longer (150 msec), suggesting precedence of low spatial frequencies (LSF) over high spatial frequencies (HSF) in the visual processing time course. These data support the influential neurobiological models of visual recognition (Peyrin et al., 2010; Bar, 2003; Bullier, 2001). According to these models, visual analysis starts with the parallel extraction of different elementary visual attributes at different spatial frequencies in a predominantly coarse-to-fine (CtF) processing sequence. The LSF in a scene, conveyed by fast magnocellular visual channels, might therefore activate visual pathways and subsequently access the occipital cortex and high-order areas in the dorsal visual cortex (extending to the parietal and frontal cortices) and ventral visual cortex (to the inferotemporal cortex) more rapidly than HSF. The rapid analysis of LSF allows an initial perceptual parsing of visual inputs before their complete propagation along the ventral visual stream, which ultimately mediates object recognition. This initial low-pass visual analysis might serve to refine the subsequent processing of HSF, conveyed more slowly by parvocellular visual channels to the ventral visual cortex.

The ventral visual cortex contains a mosaic of different areas that respond selectively to different categories of visual stimuli (Spiridon & Kanwisher, 2002; Haxby et al., 2001; Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001). For example, faces selectively activate a lateral region of the fusiform gyrus, called the fusiform face area (FFA; Kanwisher, McDermott, & Chun, 1997), whereas man-made objects primarily activate the lateral occipital complex (LOC; Grill-Spector, Kourtzi, & Kanwisher, 2001; Malach et al., 1995). A number of fMRI studies have specifically investigated the cerebral structures involved in complex scene processing compared with other visual stimuli (e.g., faces and objects). Most studies agree that a prominent region in the inferotemporal cortex, known as the parahippocampal place area (PPA), and the retrosplenial cortex (RSC) are regions of the human cortex primarily involved in processing "spatial layout" during the perception of scenes (Epstein & Ward, 2010; Epstein, 2005, 2008; Epstein & Higgins, 2007; Epstein, Graham, & Downing, 2003; Epstein, Harris, Stanley, & Kanwisher,

1999; Epstein & Kanwisher, 1998), navigationally relevant spatial information within a familiar real world environment (Vass & Epstein, 2013; Epstein, Higgins, Jablonski, & Feiler, 2007) or contextual associations (Bar, Aminoff, & Ishai, 2008; Bar, Aminoff, & Schacter, 2008; Aminoff, Gronau, & Bar, 2007; Bar, 2004, 2007; Bar & Aminoff, 2003). A region around the transverse occipital sulcus was also more recently found to be involved in both the "spatial layout" processing and the semantic categorization process during the perception of scenes (Dilks, Julian, Paunov, & Kanwisher, 2013). As a result, this region is now called the occipital place area (OPA).

However, the specific functions supported by scene-selective regions during the categorization of scenes remain unclear. We do not as yet know if scene-selective regions use CtF categorization because, to our knowledge, no imaging studies have explored the effects of different spatial frequency orders during the explicit categorization of complex visual scenes within these regions. In the present fMRI study, we measured the activation of scene-preferring regions during the categorization of dynamic natural scene stimuli (see Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012), in which they resorted to either a CtF sequence or a reverse fine-to-coarse (FtC) sequence. We first identified scene-selective regions in each individual using a localizer adapted from previous studies (Walther, Caddigan, Fei-Fei, & Beck, 2009; Bar, Aminoff, & Ishai, 2008; Epstein et al., 2003; Epstein & Kanwisher, 1998) in which participants viewed grayscale photographs of scenes, faces, and common objects. The contrast between scenes and other categories was intended to enable the localization of the regions involved in the perception of scenes. Once localized, we compared activation elicited by CtF and FtC dynamic scenes within the areas defined as the PPA, RSC, and OPA.

Methods

Participants

Sixteen right-handed participants (eight men; 23±2 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, which was approved by the local ethics committee. All participants were submitted to two experiments: the dynamic scene experiment and the localizer experiment.

Stimuli and Procedure in the Dynamic Scene Experiment

Stimuli consisted of 40 black and white photographs (256-level grayscales, 1024 × 768 pixels) of man-made scenes classified into two distinct categories (20 indoor scenes and 20 outdoor scenes) with a visual angle of 24 × 18 degrees. Exemplars from both categories were chosen to ensure similarity of amplitude spectrum and to prevent categorization from being based on this type of visual cue (Guyader et al., 2004) and to avoid contrast energy differences between categories that could interfere with the sequence of spatial frequency processing. In both categories, images had the same distribution of energy in spatial frequencies and dominant orientations (as shown by the mean amplitude spectrum of non-filtered natural scenes in each category; Figure 1A). To ensure that the chosen scenes have similar amplitude spectra, we first calculated the mean amplitude spectrum for the 20 indoor scenes (mean AS indoor) and the 20 outdoor scenes (mean AS outdoor). Then, for each scene, we calculated two 2-D correlation coefficients, one between the scene's amplitude spectrum and the mean AS indoor and the other one between the scene's amplitude spectrum and the mean AS outdoor. The mean AS of the category corresponding to the scene of interest was calculated by excluding the scene's amplitude spectrum (i.e., for an indoor scene, the mean AS indoor was calculated based on the 19 remaining indoor scenes, whereas the mean AS outdoor was calculated based on the 20 outdoor scenes).

The 2-D correlation coefficient was calculated using the Matlab function "corr2d." A 2 × 2 variance analyses (ANOVA) with the category of the scene (indoor and outdoor) and the category of the mean AS (indoor and outdoor) as within-subject factors were conducted on the 2-D correlation coefficients. Results show that the 2-D correlation coefficients calculated between indoor scenes and the mean AS indoor did not significantly differ from those calculated between indoor scenes and the mean AS outdoor (0.76 ± 0.05 and 0.76 ± 0.06, respectively; $F(1, 38) < 1$). Similarly, the 2-D correlation coefficients calculated between outdoor scenes and the mean AS outdoor did not significantly differ from those calculated between outdoor scenes and the mean AS indoor (0.78 ± 0.05 and 0.78 ± 0.04, respectively; $F(1, 38) < 1$).

Furthermore, outdoor and indoor categories were equivalent in terms of visual cluttering (subband entropy measures; see Rosenholtz, Li, & Nakano, 2007). The mean subband entropy was equivalent for outdoors and indoors (2.95 ± 0.16 and 2.95 ± 0.14, respectively; $F(1, 38) < 1$). Stimuli were created using the image processing toolbox on MATLAB (Mathworks, Inc., Sherborn, MA). On the basis of our previous studies (Musel et al., 2012), each scene was filtered by six band-pass filters with different central spatial frequencies: 24, 34, 49, 71, 101, 144 cycles/image and a standard deviation of 25.6 cycles/image (or 1, 1.42, 2.04, 2.96, 4.21, 6 cycles/ degree and a standard deviation of 1.07 cycles/degree). The central frequencies of filters followed a logarithmic scale. This enabled us to obtain a better sample of the amplitude spectrum of natural scene, in which the energy decreases as frequency increases ("1/f" shape; Field, 1987) and more filters centered on low spatial frequencies (for a similar approach, see Willenbockel et al., 2010). For the first filter, we removed information contained in the frequencies below 7.2 cycles/image (or 0.3 cycles/degree) and retained higher LSF to apply a band-pass filter even for the lower spatial frequency image (see Figure 1). The cut off frequencies at 67% of the height of each Gaussian were, therefore, [7 47] [11 58] [26 72] [47 93] [78 124] [121 167] cycles/image (i.e., [0.29 1.96] [0.46 2.42] [1.08 3] [1.96 3.88] [3.25 5.17] [5.04 6.97] cycles/degree). Images were then normalized to obtain a mean luminance of 128 with a standard deviation of 25.5 (i.e., root mean square [RMS]) on a 256 gray-level scale. This resulted in six versions of each scene, all accurately categorized,1 and from these we created a movie. For each scene, we created two movies: one following a CtF sequence and one following a FtC sequence. Each movie lasted 150 msec and was composed of the same scene filtered in the six different frequency bands (presented for 25 msec). Stimuli were displayed using E-prime software (E-prime Psychology Software Tools, Inc., Pittsburgh, PA) and back-projected on a translucent screen positioned at the rear of the magnet. Participants viewed this screen at a distance of about 222 cm via a mirror fixed on the head coil. We used a backward mask, built with 1/f white noise, to prevent retinal persistence of the scene.

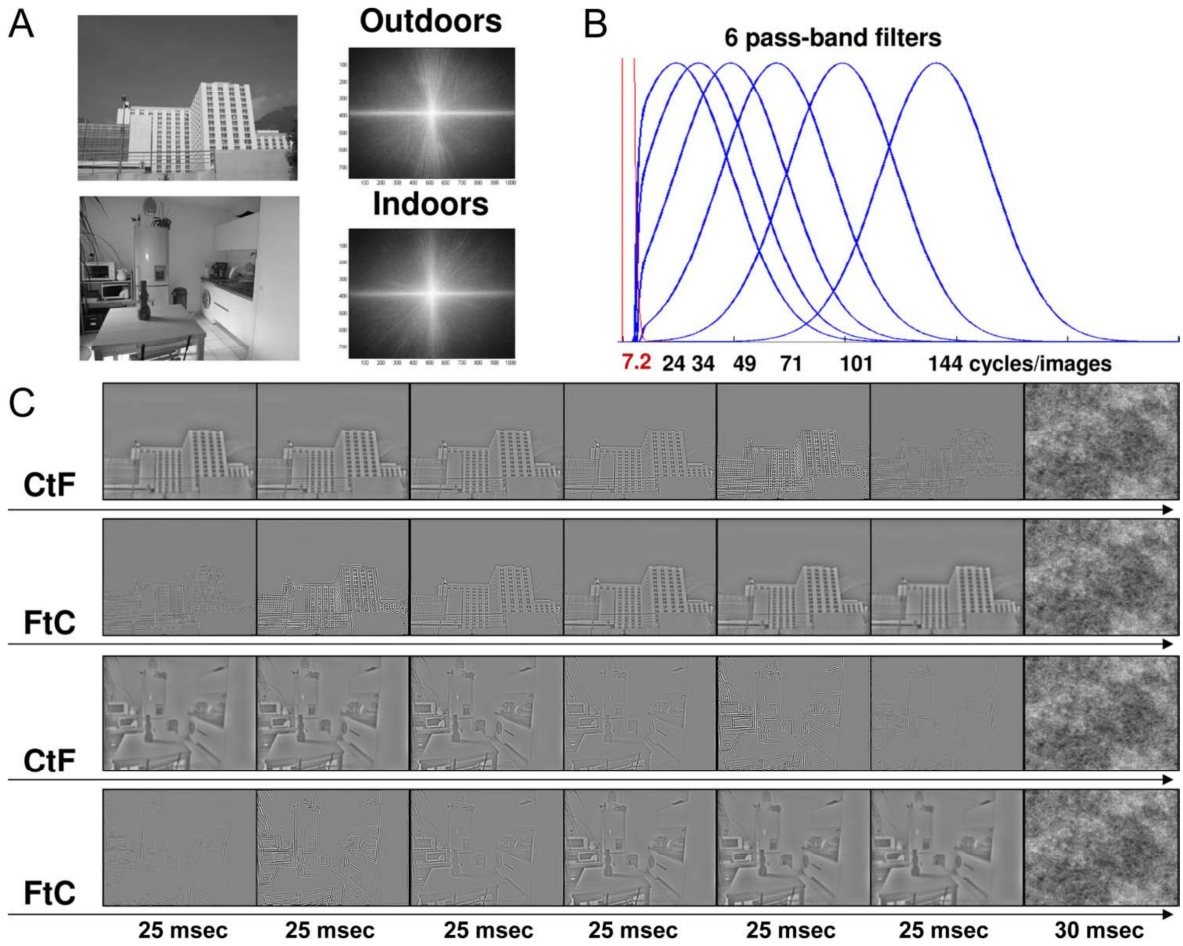


Figure 1. (A) Example of scenes belonging to different categories (outdoors and indoors). Mean amplitude spectra of each category. On each amplitude spectrum, low spatial frequencies are close to the center, and high spatial frequencies are on the periphery. Vertical orientations are represented on the x axis, and horizontal orientations are on the y axis. (B) Example of six spatial frequency filtered images of scenes depicting (C) the CtF and FtC sequences.

A block design paradigm was used with CtF and FtC movies. The dynamic scene experiment consisted of four functional runs. Each functional scan lasted 5 min and was composed of eight 25-sec task blocks (four CtF blocks and four FtC blocks) including 10 dynamic scenes (five indoors and five outdoors), interspersed with 4 25-sec blocks with a fixation dot in the center of the screen (Fixation condition) displayed against a gray background. Each scene was presented in the CtF and FtC movie conditions within a run, but a given scene appear only once in each sequence condition within a run. It should be noted that a block design paradigm did not allow to analyze individual response to trials as an event-related paradigm allowed, and it did not allow to investigate the neural correlates of priming effect in our study. However, a block design paradigm increases the statistical power (Friston, Zarahn, Josephs, Henson, & Dale, 1999) and the relatively large BOLD signal change relative to baseline (Glover, 1999). The robustness of activations is particularly important for this study in which we used two experimental conditions involving the exact same images for the same total exposure duration but differing only by their relative temporal order. The choice of a block design paradigm was guided by an unpublished previous work in which the use of an event-related paradigm with CtF and FtC movies, as well as null events, did not reveal significant activation.

Each functional run consisted of 80 experimental trials. Each stimulus was displayed for 150 msec, followed by a mask for 30 msec and a fixation dot in the center of the screen. The average interval between the onsets of two successive stimuli was 2.5 sec. Participants had to give a categorical answer on movies ("indoors" or "outdoors") by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. They were instructed to fixate the center of the screen (fixation dot) during the whole run and to respond as quickly and as accurately as possible by pressing one of two response buttons. Half the participants had to answer "indoor" with the forefinger and "outdoor" with the middle finger, whereas the other half had to answer "indoor" with the middle finger and "outdoor" with the forefinger. Response accuracy (ACC) and RTs (in milliseconds) were recorded.

Stimuli and Procedure in the Localizer Experiment

Following the main dynamic scene experiment, we performed a separate functional localizer experiment to localize the functional ROIs specifically involved in the processing of natural scenes. The localizer experiment was adapted from previous studies (Walther et al., 2009; Bar, Aminoff, & Ishai, 2008; Epstein & Kanwisher, 1998). Participants viewed scrambled and intact versions of grayscale photographs of scenes, faces, and common objects in a block design paradigm. Scene pictures used in the localizer experiment were not shown during the dynamic scene experiment. Stimuli were black and white photographs (256 grayscales), all sized 700 × 700 pixels (with a visual angle of 16.4 × 16.4 degrees), and RMS contrast of images was normalized. Scrambled pictures of these stimuli were created by dividing and randomizing intact picture scenes into 100 × 100 pixel squares. Participants viewed intact and scrambled versions of photographs of scenes, faces, and common objects in separate blocks of a block design paradigm. The localizer experiment consisted of two functional runs. Each functional scan lasted 3 min 30 sec and was composed of ten 15-sec task blocks (one block for each stimulus type) including 15 different photographs of the same type, interspersed with four 15-sec blocks with a fixation dot in the center of the screen displayed against a gray background. Participants performed a "1-back" repetition detection task. They were instructed to press a button whenever they saw two identical stimuli repeated. This task ensured that participants paid at least as much attention to unmotivating stimuli (e.g., scrambled images) as to more interesting stimuli (scenes, faces, and objects). Only two repetitions were presented per block. Each stimulus was presented for 300 msec, with a 700-msec ISI with a fixation dot in the center of the screen. The PPA, RSC, and OPA were identified in both hemispheres mainly by a [Scenes > Faces + Objects] contrast.

fMRI Acquisition

Experiments were performed on a whole-body 3T Philips scanner (Philips Medical Systems, Eugene, OR) at the University Hospital Center of Grenoble (France). For all functional dynamic scene and localizer scans, the manufacturer-provided gradient-echo/T2* weighted EPI method was used. Thirty-nine adjacent axial slices parallel to the bicommissural plane were acquired in interleaved mode. Slice thickness was 3.5 mm. The in-plane voxel size was 3 × 3 mm (216 × 216 mm field of view acquired with a 72 × 72 pixel data matrix; reconstructed with zero filling to 128 × 128 pixels). The main sequence parameters were repetition time= 2.5 sec, echo time = 30 msec, flip angle=77°. To correct images for geometric distortions induced by local B0 inhomogeneity, a B0 field map was obtained from two gradient-echo data sets acquired with a standard 3-D FLASH sequence (Δecho time = 9.1 msec). The field map was subsequently used during data processing. Finally, a T1-weighted high-resolution 3-D anatomical volume was

acquired by using a 3-D Modified Driven Equilibrium Fourier Transform sequence (field of view = 256 × 224 × 176 mm; resolution = 1.333 × 1.750 × 1.375 mm; acquisition matrix = 192 × 128 × 128 pixels; reconstruction matrix = 256 × 128 × 128 pixels).

Data Analysis

Data analysis was performed using the general linear model (Friston et al., 1995) for block designs in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK, www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB 7 (Mathworks, Inc., Sherborn, MA). Individual scans were realigned, time-corrected, normalized to the MNI space and spatially smoothed by an 8-mmFWHMGaussian kernel. Time series for each voxel were high-pass filtered (1/128 Hz cutoff) to remove low-frequency noise and signal drift.

Individual ROIs were isolated based on the two localizer scans. The fMRI signal in the localizer runs was analyzed using single-participant general linear model. For each participant, five conditions of interest (scenes, faces, objects, scrambled scenes, and fixation) were modeled as five regressors convolved with a canonical hemodynamic response function. Movement parameters derived from realignment corrections (three translations and three rotations) were also entered in the design matrix as additional factors of no interest. The areas responding to scenes were defined independently for each participant using two contrasts: [Scenes > Faces + Objects] and [Scenes > Scrambled scenes]. The contrast eliciting greater activity in ROIs for all participants was the one used to continue the analysis of data. Significant voxel clusters on individual t maps were identified using a false discovery correction at qFDR < 0.05 to control the overall false positive rate (Benjamini & Hochberg, 1995). Scene-selective voxel clusters were located in the PPA, RSC, and OPA. To facilitate comparisons with other studies, a transformation of MNI into Talairach and Tournoux (1988) coordinates was performed using the MNI2TAL function (created by Matthew Brett, available at www.mrc-cbu.cam.ac.uk/Imaging). These clusters were selected as ROIs for the analysis of data in the dynamic scene experiment in which categorization of CtF and FtC sequences was examined.

For the dynamic scene experiment, three conditions of interest (CtF, FtC, and Fixation) were modeled as three regressors convolved with a canonical hemodynamic response function. Note that the CtF and FtC experimental conditions involved the exact same images for the same total exposure duration (150 msec) but differed by their relative temporal order, which was too close to produce distinct hemodynamic responses to each of the six images during fMRI. RTs for each trial and movement parameters derived from realignment corrections (three translations and three rotations) were also entered in the design matrix as additional factors of no interest to account for RT-related variance and head motion, respectively. Parameter estimates (percent signal change relative to the global mean intensity of signal) of block responses were then extracted from the scene-selective ROIs for each participant. The average parameter of activity was calculated for CtF, FtC, and Fixation conditions. These values were submitted to a repeated measure ANOVA with Conditions (CtF, FtC, and Fixation), Regions (PPA, RSC, and OPA), and Hemispheres (Left and Right) as within-subject factors.

Results

Behavioral Results

Two 2 × 2 variance analyses (ANOVA) with sequences (CtF and FtC) and categories (outdoor and indoor) as within-subject factors were conducted on mean error rates (mER) and mean correct RTs (mRT). The ANOVA conducted on mER showed no effect of Sequences (CtF: 3.98 ± 4.68%; FtC: 5.43 ± 8.41%; $F(1, 16) = 1.48$, $p = .24$), but a main effect of Categories ($F(1, 15) = 9.91$, $p < .05$).

Participants made more errors when categorizing indoor (5.47 ± 7.25%) than outdoor scenes (3.95 ± 6.32%). No interaction was observed between Sequences and Categories ($F(1, 15) < 1$). The ANOVA conducted on mRT showed that participants categorized CtF sequences more quickly than FtC sequences (CtF: 640 ± 128 msec; FtC: 657 ± 130 msec; $F(1, 15) = 7.50$, $p < .05$). There was no effect of Categories (outdoor: 650 ± 125 msec; indoor: 647 ± 132 msec; $F(1, 15) < 1$), and no interaction was observed between Sequences and Categories ($F(1, 15) < 1$; Figure 2).

fMRI Results

PPA, RSC, and OPA ROIs were defined in each individual based on the independent localizer experiment. This served as the structural constraint for the analysis of data in the dynamic scene experiment in which categorization of CtF and FtC sequence was examined. Using the contrast in which scenes were compared with faces and common objects, greater activity elicited by scenes was observed in different scene-selective regions. Scenes elicited stronger activation than faces and objects ([Scenes > Faces + Objects] contrast; Figure 3A) within the bilateral parahippocampal gyrus (including the PPA), cingulate gyrus (including the RSC), and the occipital gyrus (in the OPA) for all participants. Peak coordinates of ROIs were consistent with previous studies (Talairach coordinates are reported in Table 1). As the clusters were generally large involving several ROIs, small sphere ROIs (3-mm radius) were created at the peak of activation clusters. To ensure that the sphere only contains voxels that were truly activated, these spheres were masked with the thresholded activation map (Poldrack, 2007). It should be noted that the contrast in which scenes were compared with scrambled images did not allow to identify all the ROIs for all participants. Parameter estimates (percent signal change relative to the global mean intensity of signal) of block responses were then extracted from these six sphere ROIs for each participant. The average parameter of activity was calculated for CtF, FtC, and Fixation condition. These values were submitted to a repeated-measure ANOVA with Conditions (CtF, FtC, and Fixation), Regions (PPA, RSC, and OPA), and Hemispheres (Left and Right) as within-subject factors.

The ANOVA revealed a significant interaction between the ROIs and the Sequences ($F(2, 30) = 3.35$, $p < .05$). Planned comparisons showed that the CtF periods elicited greater activation than FtC periods only within the PPA ($F(1, 15) = 12.52$, $p < .05$; RSC: $F(1, 15) < 1$; OPA: $F(1, 15) < 1$; Figure 3B). Hemispheres did not interact with Sequences in either region (PPA: $F(1, 15) < 1$; RSC: $F(1, 15) = 2.29$, $p = .15$; OPA: $F(1, 15) < 1$). The right and left PPA were more activated for the CtF than FtC (right PPA: $F(1, 15)=13.27$, $p<.05$; left PPA: $F(1, 15)=9.93$, $p< .05$). This was not the case for the right and left RSC (right RSC: $F(1, 15) = 1.50$, $p = .24$; left RSC: $F(1, 15) < 1$) or the right and left OPA (right OPA: $F(1, 15) = 1.48$, $p = .24$; left OPA: $F(1, 15) < 1$). Furthermore, CtF and FtC sequences showed greater activation than Fixation periods within the right PPA (CtF vs. Fixation: $F(1, 15) = 42.50$, $p < .05$; FtC vs. Fixation: $F(1, 15) = 23.40$, $p < .05$) and left PPA (CtF vs. Fixation: $F(1, 15)= 32.60$, $p < .05$; FtC vs. Fixation: $F(1, 15) = 31.63$, $p < .05$). This was not the case for the other scene-selective regions (all $F(1, 15) < 1$).

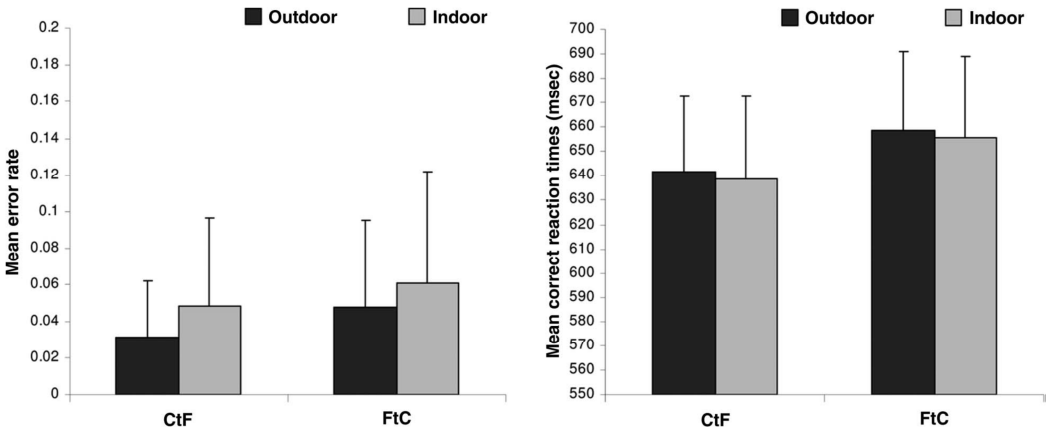


Figure 2. Mean error rates and mean correct RTs in milliseconds according to CtF and FtC sequences and scene categories (Outdoor and Indoor). Error bars correspond to standard errors.

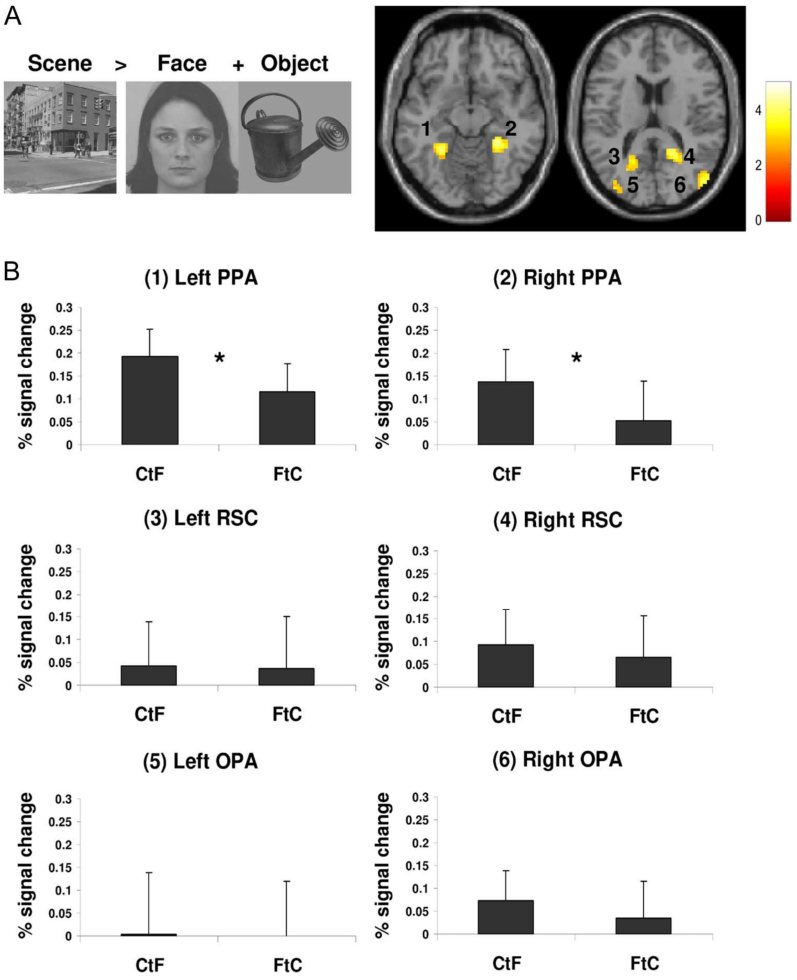


Figure 3. (A) Cerebral regions activated during the perception of scenes compared with faces and objects ([Scenes > Faces + Objects] contrast): (1) left PPA, (2) right PPA, (3) left restrosplenial cortex/ RSC, (4) right RSC, (5) left OPA, (6) right OPA. ROIs are illustrated on a representative participant. (B) The ROIs were defined independently for each participant by contrasting intact scenes to meaningful intact other stimuli: [Scenes > Faces + Objects]. Signal changes relative to the global mean intensity of signal were then extracted from the scene-selective ROIs for each participant and each sequence (CtF and FtC). Graphics represent for each ROI the mean percentage of signal change for 16 participants. Error bars indicate 95% confidence intervals. * indicate significant differences.

Table 1: Mean coordinates (standard errors are shown in italics) of scene-selective regions identified by the [Scenes > Faces + Objects] contrast in the Localizer experiment. Coordinates (x, y, z) are indicated in the Talairach space.

	Mean x		**Mean y**		**Mean z**	
Right PPA	26	*±2.9*	-38	±3	-10	*±5.5*
Left PPA	-26	*±4.4*	-43	±4.7	-8.2	*±2.9*
Right RSC	21	*±4.2*	-53	±7.7	15	*±11*
Left RSC	-19	*±4.2*	-57	±7.4	14	*±11*
Right OPA	-29	*±9*	-78	±6.5	28	*±18*
Left OPA	-22	*±8*	-70	±4	27	*±15*

Discussion

The present fMRI study investigated for the first time the temporal order of spatial frequency processing within scene-selective regions. For this purpose, we created two types of movies composed of the same scene filtered in six different spatial frequency bands but displayed in different orders, one following a CtF sequence and one following a FtC sequence. These sequences therefore imposed the order of spatial frequency processing. Although this procedure was obviously not physiological, it allowed us to experimentally mimic the sequential processing of spatial frequencies postulated by influential model of visual categorization (Peyrin et al., 2010; Hegde, 2008; Bar, 2003; Schyns & Oliva, 1994). Crucially, all visual information content was the same in both sequences, and only the order of spatial frequency images changed. None of the participants realized that we manipulated the spatial frequency order. However, when we used these stimuli in a previous behavioral study (Musel et al., 2012), we showed that CtF sequences were categorized more rapidly than FtC sequences in young adults. This provided new arguments in favor of a predominantly CtF categorization of natural scenes and a new experimental tool, which imposes a CtF processing and allows investigations of the neural substrates of CtF processing. In this study, we replicated the behavioral results of Musel et al. (2012). Participants categorized CtF more rapidly than FtC sequences. Furthermore, in terms of scene-selective regions, only the PPA showed stronger activation during the CtF than FtC categorization of filtered scenes. Sequences of filtered scenes (with blank screens occurring between scenes) were also used by Schettino, Loeys, Delplanque, and Pourtois (2011) in an evoked potential study to investigate the neural correlates of the accumulation of visual information during object recognition and their time course. For this purpose, the authors used sequences in which the first scene was always in LSF and the scene was gradually revealed in six successive images by progressively adding HSF information. In line with our results, these authors observed that activation in the parahippocampal cortex decreases when the spatial frequency content of scenes increases, suggesting that this region is sensitive to the primary processing of LSF information, even if this study did not investigate explicit CtF processing. CtF processing of faces in high-level visual cortex was recently the central focus of Goffaux et al. (2011), who showed an intriguing effect of spatial frequencies in a face-selective region, the FFA. By manipulating exposure duration and the spatial frequency content of faces, the authors showed that the FFA responded more strongly to LSF for short exposure durations of faces, whereas it responded more strongly to HSF for longer exposure durations. These results suggest that CtF processing is the predominant strategy in the

most prominent regions of the ventral visual stream (inferotemporal cortex). Interestingly, the authors used scrambled faces (phase of the face images was scrambled in the Fourier domain via random permutation) as control stimuli, from which no face representation can be extracted. Therefore, the activation elicited by contrasting intact and scrambled faces are related to high-level representations of faces and not to low-level aspects of spatial frequency processing. Note, however, that participants had to perform an intact versus scrambled face categorization task. This task possibly engaged other cognitive demands and neural processes that the ones usually involved in face recognition. In our present dynamic scene experiment, we did not use scrambled stimuli because we were rather concerned by the effect of temporal order of different spatial frequency bands during an explicit categorization of dynamic scene and also because the CtF and FtC movies involved the exact same images and thus the same low-level visual information.

This study does not directly address the role played by the PPA in scene perception, but it suggests that the PPA would be selectively tuned to the processing of LSF before HSF information. Early studies (Epstein, 2008; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Epstein et al., 1999; Epstein & Kanwisher, 1998) assumed that the PPA responds more strongly to images of real-world scenes (such as cityscapes and landscapes) than to scrambled images and other meaningful visual stimuli (such as faces and objects). We obtained similar results in our localizer experiment. According to Epstein and colleagues (Epstein & Ward, 2010; Epstein, 2005, 2008; Epstein & Kanwisher, 1998), the PPA encodes the geometric structure (i.e., spatial layout) of scenes. The PPA is sensitive to real scenes without discrete objects (an empty room) and to complex meaningful scenes containing multiple objects (the same room furnished). However, it responds weakly to objects lacking a 3-D spatial context (objects in this room on a blank background). The perception of 3-D spatial information usually requires global perception of the scene and might therefore be preferentially performed on a first LSF-based analysis (Farell, Li, & McKee, 2004). Bar and colleagues (Bar, Aminoff, & Ishai, 2008; Bar, Aminoff, & Schacter, 2008; Aminoff et al., 2007; Bar, 2004; Bar & Aminoff, 2003) suggest that the PPA would be more particularly sensitive to visual contextual associations in scenes. These authors have shown in several studies that the PPA responds significantly more strongly to images with highly associative contextual objects both in spatial (e.g., a traffic light is strongly associated with a street context) and in nonspatial domains (e.g., a crown is strongly associated with royalty, but not with a specific place) compared with images of equal visual qualities but containing objects with weak contextual associations (e.g., a cell phone, which is not strongly associated with a single context). The PPA should therefore be viewed not as being exclusively dedicated to the analysis of place, but rather as more generally mediating spatial contextual associations (Aminoff et al., 2007). Computational data again suggest that the early processing of LSF information is sufficient to extract the context of a scene (Torralba & Oliva, 2003; Oliva & Torralba, 2001). Therefore, this study would provide support for both the spatial layout and contextual hypothesis of PPA function.

PPA sensitivity to LSF (when displayed before HSF) could be linked to previous fMRI results showing greater activation of the parahippocampal gyrus for LSF than HSF scene categorization (Peyrin, Baciu, Segebarth, & Marendaz, 2004). This result differed from recent fMRI studies conducted by Rajimehr, Devaney, Bilenko, Young, and Tootell (2011), who showed that the PPA responds preferentially to spatial discontinuities in geometrical shapes (such as boundary edges and corners) and HSF in

natural images. The discrepancy between our results and those of Rajimehr et al. (2011) could be because of several experimental parameters. First, Goffaux et al. (2011) clearly demonstrated that the presentation time of stimuli could drastically modify the level of activity of category-specific regions. Face-specific regions were more strongly activated by LSF faces in the early stages of visual processing (up to 75 msec of face exposure), and this activation decreased as a function of exposure duration (mostly to 150 msec). In contrast, activation in response to HSF faces increased over time within these regions. In Rajimehr et al. (2011), the presentation time of stimuli (checkerboard, face and scene) was always beyond 500 msec. Such a long presentation time may favor the analysis of HSF information and the stronger activity of the PPA for HSF stimuli. Second, like psychophysical studies, which revealed a certain degree of flexibility in the extraction and analysis of spatial frequencies depending on task demands (Rotshtein, Schofield, Funes, & Humphreys, 2010; Oliva & Schyns, 1997; Schyns & Oliva, 1997), we believe that the sensitivity of the PPA to a particular bandwidth may also depend on the demands of the visual task. Thus, the PPA would be more activated for LSF when visual tasks require a global visual processing and for HSF when visual tasks require finer and more detailed visual processing. This assumption could explain why Zeidman, Mullally, Schwarzkopf, and Maguire (2012) observed stronger activation of the PPA in response to HSF than LSF 3-D spaces. In this study, the spaces were depicted by positioning small white dots on a black background following an exponential distribution and then either displayed in HSF (intact dot pattern) or filtered in LSF. Participants had to detect whether a small proportion of dots disappeared irrespective of spatial frequency content. This type of detection task involves a fine analysis of visual information, which may favor the analysis of HSF spaces by the PPA. In the same way, LSF information could be sufficient to allow rapid discrimination and categorization of indoor and outdoor scenes by the PPA. It should however be noted that, in this study, we used dynamic scenes containing both LSF and HSF information displayed in different relative orders. It is therefore possible that each image that made up the movies may have induced differential activation within the PPA. Future studies manipulating explicitly spatial frequencies in different stimuli are needed to fully investigate the spatial frequency selectivity of the PPA during scene categorization.

Concerning the other scene-selective regions, although both RSC and OPA were more activated by the passive viewing of scenes compared with faces and objects in the Localizer experiment, they were not more activated by the categorization of CtF and FtC sequences compared with fixation periods in the dynamic scene experiment. These results suggest that these two regions were not involved in the processing of spatial frequencies or the categorization process. However, according to evidence previously provided by Dilks et al. (2013), the OPA was shown to be explicitly involved in the categorization of scenes such as beaches, forests, cities, or kitchens (compared with the categorization of objects as a camera, chair, car, or shoes). The authors even suggest that the OPA was preferentially involved in the processing of low-level visual features in scenes such as spatial frequencies or spatial envelope properties (i.e., the representation of the shape of a scene based on its degree of naturalness, openness, roughness, expansion, and ruggedness of the scene; Oliva & Torralba, 2001). Our study did not reveal significant activation of the OPA for either CtF or FtC sequences relative to fixation periods, excluding the potential specialization of the OPA for spatial frequency processing during the categorization. It should be noted that Dilks et al. (2013) used a task involving the

perception of scenes and objects whose spatial envelope properties and spatial representations differed greatly. This suggests that the OPA might be more involved in the spatial discrimination between scenes and other categories through the analysis of the spatial envelope than in the analysis of spatial frequency content used in the categorization process. In keeping with this assumption, the two categories used in this study (outdoor and indoor categories) were composed exclusively of man-made environments. Outdoor scenes consisted mainly of front views of buildings or close-up views of cities. This drastically reduced the degree of expansion that usually characterizes outdoors and allowed them to be distinguished from indoor scenes. Furthermore, both categories had similar amplitude spectra and equivalent levels of visual cluttering to prevent task performance being based on these low-level visual features. The two categories therefore contained large objects with flat surfaces (either a kitchen cupboard door for the indoor category or shutters on a building for the outdoor category), although outdoor scenes are usually more highly textured and made up of small elements. Given the visual properties of our scenes and based on previous computational works (Oliva & Torralba, 2001), we assumed similarity of spatial envelope in both categories. Finally, mean luminance and contrast were normalized for the six images that composed the movie. As a result, the absence of activation in the OPA during our categorization task may have been because of an irrelevant analysis of the amplitude spectrum (i.e., the distribution of energy in spatial frequencies and dominant orientations), the spatial envelope properties or the contrast to perform the visual task. Alternatively, the OPA could be selective to a particular spatial frequency range within the CtF and FtC sequences, but the selective activation would be disturbed by the other spatial frequency ranges.

The RSC, region of the posterior cingulate cortex, is mainly involved in spatial memory and navigation (Kravitz, Saleem, Baker,& Mishkin, 2011; Vann, Aggleton,&Maguire, 2009; Aggleton & Vann, 2004; Harker & Whishaw, 2004). These interpretations are mainly based on the extensive connections linking the RSC and the visual and hippocampal regions and the presence of cells responding to specific directions of the head (Vann et al., 2009; Cho & Sharp, 2001). However, in a number of recent studies, involvement of this region was also evidenced during the perception of natural scenes (Park, Brady, Greene, & Oliva, 2011; Park & Chun, 2009; Bar, Aminoff, & Ishai, 2008; Sung, Kamba, & Ogawa, 2008; Walther et al., 2008; Aminoff et al., 2007; Epstein & Higgins, 2007) in the coding of “spatial layout” and location (Vass & Epstein, 2013; Epstein & Higgins, 2007) or contextual associations (Bar, Aminoff, & Ishai, 2008; Aminoff et al., 2007). More particularly, the RSC may support an expansive spatial representation of the scenes, which allows the current view of the scene to be replaced into a larger spatial environment extending beyond the borders of the image (Kravitz et al., 2011; Park et al., 2011; Epstein & Higgins, 2007). Using three pictures from panoramic scenes representing different views of the same scene in an fMRI paradigm of adaptation, Park et al. (2011) found a significant attenuation of activation in the RSC for repeated panoramic scenes suggesting that different viewpoints of the same scene are integrated into a single visual context. It has also been demonstrated that the RSC responds more strongly to the identification of specific locations around a familiar environment than to general judgments about the scene category (Epstein & Higgins, 2007). Consequently, it could explain why the RSC was not activated during our categorization task.

To sum up, we have demonstrated here that the CtF strategy is a plausible modus operandi in the PPA. This result

provides critical support for influential models of visual perception based mainly on a spatial frequency analysis, which follows a CtF strategy (Peyrin et al., 2010; Hegde, 2008; Bar, 2003; Schyns & Oliva, 1994). It also provides additional data for the development of a neurally grounded model of scene perception.

Acknowledgments

This work was supported by the RECOR ANR grant (ANR-12- JHS2-0002-01 RECOR). Benoit Musel and Louise Kauffmann were supported by Région Rhône-Alpes (ARC2 and Cible grants). The authors warmly thank the "Délégation à la Recherche Clinique et à l'Innovation" of the University Hospital of Grenoble for sponsoring. We thank Catherine Dal Molin for the English revision of the manuscript.

Footnotes

[1] A control behavioral study was conducted on 17 participants to verify that the six spatial frequency band images were correctly categorized. Whatever the spatial frequency band, the mean error rate was low (mER for the pass-band centered on 24 cycles/image [1 cycle/degree]: 6.03 ± 1.25%; on 34 cycles/ image [1.42 cycles/degree]: 4.59 ± 0.95%; on 49 cycles/image [2.04 cycles/degree]: 5.59 ± 1.02%; on 71 cycles/image [2.96 cycles/ degree]: 5.15 ± 1.36%; on 101 cycles/image [4.21 cycles/degree]: 4.71 ± 1.02%; on 144 cycles/image [6 cycles/degree]: 5.29 ± 1.21%) and there was no effect of the spatial frequency band, $F(5, 80) < 1$.

References

Aggleton, J. P., & Vann, S. D. (2004). Testing the importance of the retrosplenial navigation system: lesion size but not strain matters: a reply to Harker and Whishaw. *Neurosci Biobehav Rev*, *28*(5), 525-531.

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb Cortex*, *17*(7), 1493-1503.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J Cogn Neurosci*, *15*(4), 600-609.

Bar, M. (2004). Visual objects in context. *Nat Rev Neurosci*, *5*(8), 617-629.

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

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), 347-358.

Bar, M., Aminoff, E., & Ishai, A. (2008). Famous faces activate contextual associations in the parahippocampal cortex. *Cereb Cortex*, *18*(6), 1233-1238.

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *J Neurosci*, *28*(34), 8539-8544.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*, 96-107.

Cho, J., & Sharp, P. E. (2001). Head direction, place, and movement correlates for cells in the rat retrosplenial cortex. *Behav Neurosci*, *115*(1), 3-25.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res*, *22*(5), 545-559.

De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res*, *22*(5), 531-544.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *J Neurosci*, *33*(4), 1331-1336a.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cereb Cortex*, *16*(10), 1453-1461.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, *23*(1), 115-125.

Epstein, R. A. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*(6), 954-978.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci*, *12*(10), 388-396.

Epstein, R. A., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene

representations in human parahippocampal cortex. On *Neuron*.

Epstein, R. A., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex, 17*(7), 1680-1693.

Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *J Neurophysiol, 97*(5), 3670-3683.

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

Epstein, R. A., & Ward, E. J. (2010). How reliable are visual context effects in the parahippocampal place area? *Cereb Cortex, 20*(2), 294-303.

Farell, B., Li, S., & McKee, S. P. (2004). Coarse scales, fine scales, and their interactions in stereo vision. *J Vis, 4*(6), 488-499.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A, 4*(12), 2379-2394.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., et al. (1995). Analysis of fMRI time-series revisited. *Neuroimage, 2*(1), 45-53.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *NeuroImage, 10*(5), 607–619.

Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In K. Boff, L. Kauman & J. Thomas (Eds.), *Hanbook of perception and human performance* (Vol. II, chap.34, pp. 1-41). NY: Wiley.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cereb Cortex, 21*(2), 467-476.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage, 9*(4), 416–429.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res, 41*(10-11), 1409-1422.

Guyader, N., Chauvin, A., Peyrin, C., Herault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *C R Biol, 327*(4), 313-318.

Harker, K. T., & Whishaw, I. Q. (2004). A reaffirmation of the retrosplenial contribution to rodent navigation: reviewing the influences of lesion, strain, and task. *Neurosci Biobehav Rev, 28*(5), 485-496.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Hegde, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Prog Neurobiol, 84*(4), 405-439.

Hughes, H. C., Nozawa, G., & Kitterle, F. L. (1996). Global precedence, spatial frequency channels, and the statistic of the natural image. *Journal of Cognitive Neuroscience, 8*, 197-230.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nat Rev Neurosci, 12*(4), 217-230.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cereb Cortex, 11*(4), 287-297.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A, 92*(18), 8135-8139.

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.

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*(1), 72-107.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope. *International Journal in Computer Vision, 42*, 145-175.

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

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage, 47*(4), 1747-1756.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *Neuroimage, 23*(2), 698-707.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., et al. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *J Cogn Neurosci, 22*(12), 2768-2780.

Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Invest Ophthalmol, 11*(5), 368-377.

Poldrack, R. A. (2007).Region of interest analysis for fMRI. *Scan, 2*, 67-70.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol, 9*(4), e1000608.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *J Vis, 7*(2), 17 11-22.

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. *J Vis, 10*(10), 25.

Schettino, A., Loeys, T., Bossi, M., & Pourtois, G. (2012). Valence-specific modulation in the accumulation of perceptual evidence prior to visual scene recognition. *PLoS One, 7*(5), e38064.

Schettino, A., Loeys, T., Delplanque, S., & Pourtois, G. (2011). Brain dynamics of upstream perceptual processes leading to visual object recognition: a high density ERP topographic mapping study. *Neuroimage, 55*(3), 1227-1241.

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

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception, 26*(8), 1027-1038.

Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Res, 42*(22), 2547-2554.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron, 35*(6), 1157-1165.

Sung, Y. W., Kamba, M., & Ogawa, S. (2008). Building-specific categorical processing in the retrosplenial cortex. *Brain Res, 1234*, 87-93.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic Physiol Opt, 12*(2), 229-232.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14*(3), 391-412.

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nat Rev Neurosci, 10*(11), 792-802.

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nat Rev Neurosci, 10*(11), 792-802.

Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *J Neurosci, 33*(14), 6133-6142.

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

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J., et al. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology: Human Perception and Performance, 36*, 122-135.

Zeidman, P., Mullally, S. L., Schwarzkopf, D. S., & Maguire, E. A. (2012). Exploring the parahippocampal cortex response to high and low spatial frequency spaces. *Neuroreport, 23*(8), 503-507.

Zhang, X., Kedar, S., Lynn, M. J., Newman, N. J., & Biousse, V. (2006). Homonymous hemianopias: clinical-anatomic correlations in 904 cases. *Neurology, 66*(6), 906-910.

5. Conclusions

Les deux études présentées ci-dessus indiquent que la PPA, le RSC et l'OPA répondent de façon distincte aux différentes gammes de fréquences spatiales, et que seule la PPA est sensible à l'ordre de présentation des fréquences spatiales (i.e. CtF ou FtC). Nous proposons d'établir un lien entre ces réponses et les fonctions qui ont été jusqu'ici attribuées à ces régions de la façon suivante.

La PPA est sensible à l'association des fréquences spatiales à des valeurs de contraste particulières et répond plus fortement aux BFS lorsque celles-ci sont associées à un contraste plus élevé que les HFS, et aux HFS lorsque celles-ci sont associées à un contraste équivalent aux BFS (**Expérience 5**). De plus cette région est sensible à l'ordre de présentation des fréquences spatiales et semble privilégier un mode de traitement « coarse-to-fine », même lorsque le contraste est égalisé entre les fréquences spatiales (**Expérience 6**). Nous avons proposé que la PPA utilise une information grossière et très contrastée en BFS pour extraire rapidement la structure spatiale et les invariants présents dans les scènes (e.g., ségrégation entre le ciel et le sol) permettant d'accéder rapidement à la catégorie de la scène et/ou à certaines de leurs caractéristiques telles que leur degré d'ouverture ou le niveau d'encombrement des objets qui les composent (Park et al., 2011, 2014). L'information en HFS serait utilisée pour l'analyse des contours et de la géométrie de la scène, et éventuellement déterminer son identité (Epstein & Higgins, 2007; Rajimehr et al., 2011). On peut par ailleurs noter que des études récentes suggèrent une dissociation entre les parties antérieure et postérieure de la PPA (Aminoff & Tarr, 2014; Arcaro, McMains, Singer, & Kastner, 2009; Baldassano et al., 2013; Berman & Walther, 2014; Nasr, Devaney, & Tootell, 2013; Park et al., 2014) de telle sorte que la partie postérieure serait plutôt impliquée dans le traitement des propriétés de bas niveau des scènes tandis que la partie antérieure de la PPA serait principalement impliquée dans la construction de représentations de haut niveau du contenu des scènes (Aminoff & Tarr, 2014[2]; Baldassano et al., 2013). En particulier, une étude récente suggère que la partie postérieure de la PPA serait plus sensible aux BFS que la partie antérieure, qui serait sélective aux HFS (Berman & Walther, 2014[2]). Ces données suggèrent donc des propriétés fonctionnelles différentes au sein de la PPA, selon un axe antéro-postérieur, qu'il serait intéressant de prendre en compte et de préciser dans de futurs travaux.

[2] Il est à noter que ces citations font référence à des communications affichées lors d'un congrès international (Vision Science Society) et n'ont pas, à ce jour, fait l'objet de publication.

Nos résultats ont également permis de mettre en évidence, pour la première fois, une sensibilité aux fréquences spatiales au sein du RSC et de l'OPA. Concernant le RSC, nos résultats suggèrent que cette région répond principalement à une information très contrastée en BFS dans les scènes et ne semble pas sensible à l'ordre de présentation des fréquences spatiales. Nous avons vu que ce type d'information pourrait être utile pour l'extraction d'invariants et de repères dans l'environnement, permettant de le structurer et de naviguer au sein de celui-ci, ainsi que pour envisager les scènes dans une perspective globale (Auger et al., 2012; Epstein et al., 1999; Epstein & Higgins, 2007; Epstein, 2008; Park et al., 2010, 2007; Park & Chun, 2009; Vann et al., 2009). L'OPA semble au contraire sélective aux HFS, suggérant l'extraction des contours et de détails fins dans les scènes. Ces résultats sont compatibles avec ceux de Dilks et al. (2013) qui montraient un rôle crucial de cette région pour la détection de différences fines entre les scènes. Notons cependant que les propriétés de cette région et son rôle relatif au sein des régions occipito-temporales sélectives aux scènes sont encore peu connus. A noter également que l'OPA est située dans la partie postérieure du cortex occipital qui correspond à la représentation du champ visuel central, en accord avec les propriétés rétinotopiques du cortex visuel. Rappelons que l'étude de Musel et al. (2013), ainsi que celle menée en collaboration avec Stephen Ramanoël (Ramanoël, Kauffmann, Cousin, Dojat, & Peyrin, sous presse; cf. ANNEXE II), ont mis en évidence que le traitement des fréquences spatiales au niveau du cortex occipital respectait l'organisation rétinotopique de celui-ci. La partie postérieure du cortex occipital, liée aux représentations fovéales, était plus activée lors de la catégorisation de scènes en HFS qu'en BFS, suggérant une sélectivité pour les HFS dans cette partie du cortex. Il est donc possible que la sélectivité de l'OPA pour les HFS observée dans l'**Expérience 5** puisse être liée au fait que cette région se situe au niveau des aires rétinotopiques dédiées au traitement des HFS.

Dans l'ensemble ces études fournissent des données supplémentaires pour la cartographie du traitement des fréquences spatiales au sein du cortex occipito-temporal. Il serait intéressant d'approfondir ces travaux en étudiant plus spécifiquement comment l'extraction des fréquences spatiales au sein des régions sélectives aux scènes contribue directement à la construction de représentations de haut niveau et aux fonctions complexes qui leur sont attribuées lors de la perception de scènes. Par ailleurs, la contribution relative de ces régions, ainsi que leurs interactions permettant d'aboutir à la reconnaissance de scènes restent également à déterminer. Dans une étude récente, Korkmaz Hacialihafiz et Bartels (2015) ont examiné la réponse de la PPA, de l'OPA et du RSC au mouvement horizontal de

scènes et de « scrambles » de phase de scènes (aléatorisation du spectre de phase). Les auteurs ont tout d'abord observé que la PPA et l'OPA répondaient plus fortement aux stimuli en mouvement qu'aux stimuli statiques, qu'ils soient des scènes ou des « scrambles ». Cet effet était plus important pour les scènes que pour les « scrambles » au sein de la PPA. Aucun effet du mouvement n'était observé au sein du RSC. La sensibilité au mouvement des scènes au sein de la PPA est compatible avec les résultats de travaux précédents qui indiquent que cette région est sensible aux différents points de vue des scènes et les analyse de façon spécifique à leur point de vue (Epstein et al., 2003; Epstein & Higgins, 2007; Park & Chun, 2009). A l'inverse, le RSC intègrerait les différents points de vue des scènes dans une représentation plus globale (Park et al., 2010; Park & Chun, 2009), ce qui pourrait expliquer que cette région n'est pas sensible au mouvement des scènes. En ce qui concerne l'OPA, sa sensibilité au mouvement, et ce, quel que soit le type de stimulus (i.e. scène ou « scramble ») semble cohérente avec sa position le long de la voie dorsale, impliquée dans le traitement du mouvement, ainsi qu'avec l'implication de cette région dans la détection de changements fins dans l'agencement spatial des scènes (Dilks et al., 2013). Ces résultats, ainsi que l'ensemble des travaux précédents ayant examiné le rôle des régions sélectives aux scènes, ont conduit les auteurs à proposer une organisation hiérarchique des traitements effectués par ces régions. L'OPA serait impliquée dans les premières étapes du traitement des scènes, en encodant des caractéristiques simples telles que le mouvement ou l'organisation spatiale des éléments. La PPA jouerait un rôle dans la perception et l'identification rapide de scènes individuelles, tandis que le RSC serait impliqué dans les dernières étapes de traitement, en intégrant les représentations des scènes selon différents points de vue pour les envisager dans leur globalité et permettre la navigation. Au regard de cette interprétation, il apparaît important d'examiner plus précisément les interactions entre ces régions lors de la catégorisation de scènes, ainsi que leur rôle respectif au sein du réseau cérébral plus large impliqué dans le traitement des fréquences spatiales et la catégorisation (cf. **Expériences 3 et 4**).

Dans le prochain et dernier chapitre consacré à la Discussion générale, nous résumerons donc les principaux résultats des études effectuées pour ce travail de thèse, et tenterons de les articuler dans un modèle neurocognitif global de la perception et de la catégorisation de scènes.

RESUME

Ce chapitre portait sur le traitement des fréquences spatiales au sein des régions occipito-temporales sélectives aux scènes : la PPA, le RSC et l'OPA. Nous avons vu que ces régions sont impliquées dans des processus complexes lors de la perception de scènes, tels que l'encodage de l'agencement spatial des scènes, la navigation, la reconnaissance et la catégorisation des scènes ou encore le traitement d'informations sémantiques liées au contexte spatial ou non spatial de l'environnement visuel. Nous avons également présenté plusieurs travaux indiquant que ces régions seraient aussi impliquées dans l'extraction de caractéristiques physiques élémentaires des scènes, telles que les régularités dans la répartition du contraste en fonction des orientations et des fréquences spatiales (i.e. leur spectre d'amplitude) caractéristiques des scènes par rapport à d'autres types de stimuli, mais aussi entre différentes catégories de scènes. A travers deux études en IRMf, nous avons examiné si ces régions participaient également au traitement des fréquences spatiales dans les scènes, et si elles privilégiaient un mode de traitement « coarse-to-fine » des fréquences spatiales. Nous avons montré que ces trois régions présentent une sensibilité distincte aux fréquences spatiales, mais aussi au contraste de luminance dans les scènes (**Expérience 5**). Lorsque le contraste de luminance des scènes filtrées n'était pas égalisé, la PPA répondait plus fortement aux BFS qu'aux HFS. Cependant, cette région était sélective aux HFS lorsque le contraste de luminance était égalisé entre les scènes filtrées. Concernant le RSC, nous avons observé que cette région répondait plus fortement aux BFS qu'aux HFS lorsque le contraste de luminance n'était pas égalisé, mais aucun effet des fréquences spatiales n'était observé lorsque le contraste de luminance était égalisé. Enfin, l'OPA était sélective aux HFS quelle que soit la condition de contraste. De plus, nous avons observé qu'une stratégie d'analyse « coarse-to-fine » était privilégiée seulement au sein de la PPA (**Expérience 6**). Ces résultats suggèrent que dans des tâches de catégorisation rapide de scènes, la PPA privilégierait l'analyse d'une information très contrastées en BFS avant l'analyse d'une information plus détaillée en HFS, le RSC une information très contrastée en BFS et l'OPA les détails contenus dans les HFS. Ces nouveaux résultats permettent de préciser les modèles actuels de la catégorisation visuelle de scènes en démontrant l'implication des régions sélectives aux scènes au sein du réseau cérébral impliqué dans le traitement des fréquences spatiales et l'analyse « coarse-to-fine ».

CHAPITRE 5 - DISCUSSION GENERALE ET PERSPECTIVES

Ce travail de thèse a eu pour objectif d'étudier et de cartographier au niveau cérébral les mécanismes neurocognitifs impliqués dans le traitement des fréquences spatiales lors de la catégorisation de scènes naturelles, chez l'adulte sain. Ce travail s'est inscrit dans l'objectif plus général d'élaborer un modèle biologiquement plausible de la catégorisation de scènes et d'en chercher les bases neurales en utilisant de nouveaux paradigmes expérimentaux construits sur la base des propriétés neurophysiologiques des voies rétino-corticales.

Comme nous l'avons vu au début de ce manuscrit, l'analyse visuelle débute par l'extraction en parallèle de différentes caractéristiques élémentaires du champ visuel à différentes fréquences spatiales. Sur la base des données issues de la neurophysiologie, de la psychophysique et de modélisations computationnelles, les modèles actuels de la reconnaissance (Hegdé, 2008; Musel et al., 2012; Schyns & Oliva, 1994) postulent que l'analyse d'une scène visuelle se ferait selon un mode de traitement par défaut « coarse-to-fine ». L'analyse rapide de l'information grossière véhiculée par les BFS via la voie magnocellulaire fournirait un aperçu de la structure globale de la scène et permettrait une première catégorisation perceptive. Celle-ci serait ensuite affinée, validée ou rejetée sur la base de l'analyse de l'information détaillée véhiculée plus tardivement par les HFS via la voie parvocellulaire. Au niveau neurobiologique, des travaux en neurophysiologie chez le primate non-humain (Bullier, 2001) et en neuroimagerie chez l'humain (e.g., Bar et al., 2006; Kveraga et al., 2007; Peyrin et al., 2010) indiquent que l'information en BFS activerait le cortex visuel primaire et serait rapidement projetée le long des voies dorsale et ventrale vers des régions pariétales, frontales et temporales de haut niveau, afin de permettre un premier traitement grossier du stimulus visuel. Le résultat de ces traitements serait ensuite renvoyé, par de rapides « feedbacks » vers des régions de plus bas niveau, afin de guider et moduler le traitement plus tardif des HFS au niveau du cortex inférotemporal pour aboutir à la reconnaissance. Cependant, les mécanismes précis d'intégration des fréquences spatiales et leurs bases cérébrales au cours de l'analyse « coarse-to-fine » des scènes sont encore sujets à débat dans la littérature.

Les travaux sur la perception visuelle de stimuli complexes tels que les scènes indiquent également qu'il existe, au sein du cortex occipito-temporal, une multitude d'aires

sélectives à différentes catégories de stimuli. Plus particulièrement, trois régions ont été mises en évidence comme étant sélectivement impliquées dans le traitement des scènes : la PPA, (Epstein & Kanwisher, 1998; Epstein, 2005), le RSC (Epstein, Higgins, Jablonski, & Feiler, 2007) et l'OPA (Dilks, Julian, Paunov, & Kanwisher, 2013). De nombreux travaux ont démontré que ces régions sont impliquées dans des processus complexes tels que la reconnaissance de scènes (Dilks et al., 2013; Epstein, 2005, 2008; Epstein, Graham, & Downing, 2003; Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Higgins, 2007; Epstein & Kanwisher, 1998; Epstein & Ward, 2010), la navigation (Epstein et al., 2007; Vass & Epstein, 2013), ou dans le traitement des liens contextuels spatiaux ou non spatiaux entre les différents éléments de l'environnement visuel (Aminoff, Gronau, & Bar, 2007; Bar, 2004, 2007; Bar & Aminoff, 2003; Bar, Aminoff, & Ishai, 2008; Bar, Aminoff, & Schacter, 2009). Cependant, la participation de ces régions au traitement des fréquences spatiales lors de la perception et la catégorisation de scènes a peu été étudiée.

Ce travail de thèse visait donc (1) à étudier les mécanismes de traitement des fréquences spatiales lors de la catégorisation de scènes naturelles chez des adultes sains et (2) à préciser les bases cérébrales de l'analyse « coarse-to-fine » des scènes. Pour cela, nous avons mené deux études comportementales qui nous ont permis d'étudier l'influence des paramètres de présentation des fréquences spatiales sur la vitesse de catégorisation de scènes et d'élaborer des stimuli adaptés à l'étude des bases cérébrales de la catégorisation « coarse-to-fine » des scènes. Par la suite, nous avons réalisé quatre études en IRMf qui nous ont permis de préciser le réseau cortical sous-jacent à la catégorisation « coarse-to-fine » des scènes, ainsi que la connectivité au sein de ce réseau.

1. Apports méthodologiques

1.1 Apport de la manipulation de l'ordre de présentation des fréquences spatiales d'une scène

Dans les **Expériences 1 et 2**, nous avons étudié quels facteurs expérimentaux permettraient un traitement optimal des fréquences spatiales pour une catégorisation rapide de scènes. Sur la base de l'hypothèse qu'une analyse « coarse-to-fine » serait privilégiée et avantageuse pour le système visuel, nous avons spécifiquement étudié l'effet de l'ordre de présentation des fréquences spatiales. Nous avons utilisé comme stimuli des séquences dynamiques adaptées d'une étude de Musel et al. (2012). Ces stimuli ont l'avantage de

permettre une manipulation explicite des fréquences spatiales délivrées au système visuel au cours du temps, afin d'imposer un mode de traitement spécifique. Dans ces deux expériences, les stimuli étaient de courts films de 150 ms composés de six versions filtrées d'une même scène (filtrage passe-bande) assemblées soit des BFS aux HFS soit des HFS aux BFS. Ainsi, ces séquences imitaient et permettaient d'imposer au système visuel une analyse « coarse-to-fine » ou « fine-to-coarse » des scènes, respectivement. La nature de ces stimuli nous a amené à nous intéresser à l'influence du mode d'accumulation de l'information en fréquences spatiales (successive ou additive) au cours des séquences (**Expérience 1**), ainsi que du niveau de contraste de luminance associé à chaque bande de fréquences spatiales de la séquence (**Expérience 2**). D'un point de vue pragmatique, ces deux expériences avaient pour objectif de proposer les stimuli visuels les plus adaptés à l'étude des bases cérébrales de la catégorisation de scènes, et tout particulièrement les corrélats neurobiologiques d'une stratégie d'analyse « coarse-to-fine ».

Dans l'**Expérience 1**, nous avons montré que l'analyse des BFS avant les HFS était plus avantageuse que celle des HFS avant les BFS pour la catégorisation des scènes et ce, indépendamment du mode d'accumulation de l'information en fréquences spatiales. Ce résultat vient donc à l'appui de nombreux travaux démontrant qu'une analyse « coarse-to-fine » est une stratégie avantageuse pour la catégorisation rapide de scènes. De plus, ce résultat indique que la réinjection au système visuel d'une information précédemment traitée (i.e. pour les séquences additives) n'avantage pas la catégorisation. Cette première expérience montre également qu'en imposant ce mode de traitement, les scènes dynamiques successives et additives sont des stimuli visuels particulièrement adaptés à l'étude de l'analyse « coarse-to-fine ». Cependant, pour la suite de nos travaux, nous avons décidé d'utiliser exclusivement des séquences dynamiques successives qui présentent plusieurs avantages. Rappelons que ces séquences sont construites en utilisant des filtres passe-bande de largeur constante permettent de limiter et de mieux contrôler la quantité d'information en fréquences spatiales délivrée au cours de la séquence. De plus, contrairement aux séquences additives pour lesquelles différents filtres (et donc différentes images filtrées) sont utilisés pour construire les séquences CtF et FtC, les séquences successives permettent d'utiliser exactement les mêmes filtrages pour les deux types de séquences et de ne faire varier que l'ordre de présentation, sans introduire de variabilité supplémentaire entre les conditions. Ainsi, une différence de performances ou d'activation cérébrale entre des séquences CtF et FtC successives

s'expliquerait en termes d'ordre relatif du traitement des fréquences spatiales et non en termes de différence de signal entre les deux séquences.

1.2. Intérêt et conséquences de l'égalisation du contraste des scènes filtrées

Dans l'**Expérience 2**, nous avons examiné plus spécifiquement dans quelle mesure l'avantage d'une analyse « coarse-to-fine » résultait plus d'une différence de contraste entre les BFS et HFS que de l'information en fréquences spatiales *per se*. En effet, rappelons que les scènes naturelles ont la propriété d'avoir un contraste de luminance qui décroît avec les fréquences spatiales, de telle sorte que les BFS sont associées à un fort contraste de luminance et les HFS à un faible contraste de luminance. Or, un stimulus très contrasté est détecté plus rapidement qu'un stimulus peu contrasté. Ainsi, il est possible que l'effet des fréquences spatiales observé dans l'**Expérience 1** puisse être confondu avec un effet du contraste associé à ces fréquences spatiales. Dans l'**Expérience 2**, nous avons donc utilisé des séquences CtF et FtC nous permettant de tester, d'une part, l'effet des fréquences spatiales seules (condition SF) et, d'autre part, l'effet du contraste seul (condition CON) par rapport à l'effet combiné des fréquences spatiales et du contraste (condition SF-CON). Les résultats ont montré que les séquences CtF étaient catégorisées plus rapidement que les séquences FtC quelle que soit la condition expérimentale (FS-CON, FS ou CON). Ces résultats indiquent que l'avantage d'une catégorisation selon une séquence « coarse-to-fine » résulte bien d'un effet de des fréquences spatiales, indépendamment du contraste, mais peut être également expliqué par un effet du contraste, indépendamment des fréquences spatiales. Cependant, l'avantage du CtF sur le FtC était plus important lorsque les séquences variaient à la fois en termes de fréquences spatiales et de contraste (condition SF-CON) que lorsqu'elles variaient seulement en termes de contraste (condition CON), mais pas lorsqu'elles variaient seulement en termes de fréquences spatiales (condition FS). Ces résultats suggèrent que si la différence de contraste entre les fréquences spatiales explique en partie l'avantage d'une analyse « coarse-to-fine » pour la catégorisation rapide, celui-ci résulte de façon prédominante d'un effet des fréquences spatiales.

D'un point de vue méthodologique, ces résultats soulèvent la question de l'utilisation de l'égalisation du contraste RMS entre des scènes filtrées pour l'étude du traitement des fréquences spatiales. A première vue, ce contrôle peut sembler important dans la mesure où il permettrait de s'abstraire de l'effet du contraste confondu à celui des fréquences spatiales lors

de la catégorisation de scènes filtrées. Nous avons vu que cela semble d'autant plus nécessaire dans le cadre d'études en IRMf, en particulier lorsque l'on s'intéresse au cortex visuel primaire, dont l'activité est connue pour être très sensible au contraste (Boynton et al., 1996; Boynton, 2005; Dakin et al., 2002; Dean, 1981; Goodyear & Menon, 1998; Movshon et al., 1978; Olman et al., 2004; Rieger et al., 2013). Par ailleurs, l'utilisation de l'égalisation du contraste RMS de stimuli filtrés peut également apparaître pertinent lorsque l'objectif de recherche est d'étudier spécifiquement l'effet des fréquences spatiales et/ou du contraste, comme c'était le cas dans l'**Expérience 2** (voir aussi, par exemple, Vlamings et al., 2009). Cependant, dans le cadre de l'étude du traitement des fréquences spatiales par le système visuel en conditions naturelles, l'utilisation de l'égalisation du contraste RMS peut être contestable. En effet, comme nous l'avons vu, la décroissance du contraste de luminance avec les fréquences spatiales est une propriété intrinsèque des scènes et constitue l'une des régularités statistiques des scènes exploitées par le système visuel pour la catégorisation (Guyader et al., 2004; Hérault et al., 1997; Oliva et al., 1999). Une manipulation de ces caractéristiques pourrait donc s'avérer contre-nature aux propriétés fonctionnelles du système visuel et induire des modifications dans les mécanismes étudiés. D'ailleurs, les résultats de l'**Expérience 2** montrent que l'information liée aux fréquences spatiales et celle liée au contraste contribuent toutes deux à la catégorisation rapide de scènes. Au niveau cérébral, les résultats de l'**Expérience 5** montrent que les activations cérébrales en lien avec le traitement des fréquences spatiales dans les régions sélectives aux scènes diffèrent significativement en fonction de l'égalisation ou non du contraste RMS entre les BFS et les HFS. Dans l'ensemble, ces résultats indiquent clairement que cette manipulation du contraste des scènes filtrées influence le traitement des fréquences spatiales au niveau comportemental et cérébral, et dans ce dernier cas jusque dans des aires visuelles impliquées dans les dernières étapes de la reconnaissance des scènes. Il est alors difficile de savoir si les résultats obtenus reflètent vraiment l'effet isolé des fréquences spatiales, ou simplement une réponse à une manipulation non-naturelle des stimuli visuels.

Une autre problématique liée à l'égalisation du contraste RMS réside dans le choix des paramètres de cette égalisation. En effet, le choix de la valeur d'égalisation (i.e. écart-type des valeurs de luminance) reste arbitraire. On note également qu'elle est variable dans la littérature. Par exemple, dans les études de Goffaux et al. (2011) et Vlamings et al. (2009) utilisant des visages filtrés en BFS et en HFS, le contraste RMS des visages filtrés était égalisé afin de correspondre au contraste RMS des visages non filtrés (i.e. contraste RMS

d'environ 0.1 pour des valeurs de luminance comprises entre 0 et 1). Etant donné que celui-ci était proche du contraste RMS des visages filtrés en BFS, l'égalisation du contraste avait principalement pour effet de rehausser le contraste des HFS. Dans le cadre de nos études, l'égalisation du contraste RMS sur la base du contraste RMS des scènes non filtrées (i.e. contraste RMS d'environ 0.2), bien plus élevé que celui des visages, aurait été problématique car cela aurait eu pour effet de rehausser drastiquement le contraste des HFS (Figure 32). Nous avons donc égalisé le contraste RMS des scènes filtrées en appliquant à toutes les scènes filtrées un contraste RMS de 0.1. Cette valeur a été choisie afin de se situer entre les valeurs moyennes de contraste RMS correspondant aux BFS et HFS des scènes naturelles (environ 0.2 et 0.03 respectivement), et d'éviter ainsi d'altérer une gamme de fréquences spatiale plus que l'autre (cf. Figure 32). Nous nous sommes donc volontairement placés dans la situation expérimentale où l'égalisation du contraste avait pour effet de rehausser le contraste des HFS, mais aussi de diminuer le contraste des BFS.

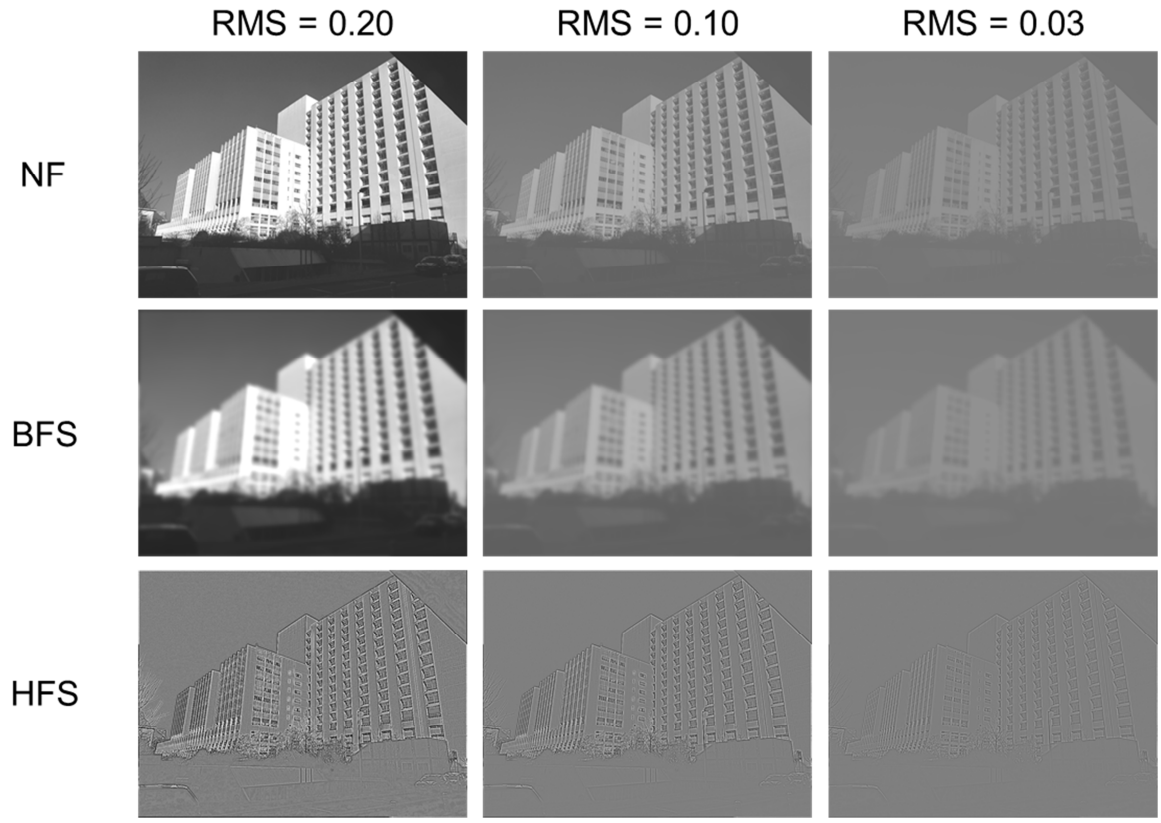


Figure 32 : Exemple de scènes non filtrées (NF), filtrées en basses fréquences spatiales (BFS) et en hautes fréquences spatiales (HFS) en fonction de différentes valeurs du contraste RMS (0.20, 0.10 et 0.03). A noter que la valeur de contraste RMS 0.20 correspond à la valeur de contraste d'une scène filtrée en BFS, que la valeur de contraste RMS 0.10 correspond à celle utilisée pour égaliser le contraste des scènes filtrées dans les Expériences 2 et 5 et que la valeur de contraste RMS 0.03 correspond à la valeur de contraste d'une scène filtrée en HFS.

Ainsi, la valeur d'égalisation du contraste RMS choisie peut avoir pour conséquence de rehausser/diminuer de façon plus importante le contraste des BFS/HFS, et d'influencer de façon différente le traitement des BFS et des HFS.

On peut par ailleurs noter que même une égalisation du contraste RMS à 0.1 semble affecter le spectre d'amplitude des scènes en HFS. Comme on peut le constater sur la Figure 33, l'égalisation du contraste RMS d'une scène en BFS a principalement pour conséquence de réduire l'énergie des basses et moyennes fréquences spatiales. En revanche, l'égalisation du contraste RMS d'une scène en HFS a pour conséquence de rehausser l'énergie des HFS mais également d'ajouter de l'énergie dans les BFS. Cet effet asymétrique de l'égalisation du contraste RMS sur le spectre des scènes en BFS et en HFS pourrait donc potentiellement biaiser le traitement des HFS par le système visuel. Afin de dissocier scrupuleusement l'effet du contraste de celui du contenu en fréquences spatiales, il serait judicieux dans le futur d'étudier l'effet de la modification du contraste RMS d'une scène filtrée en faisant systématiquement varier différentes valeurs de contraste RMS pour chaque bande de fréquences spatiales.

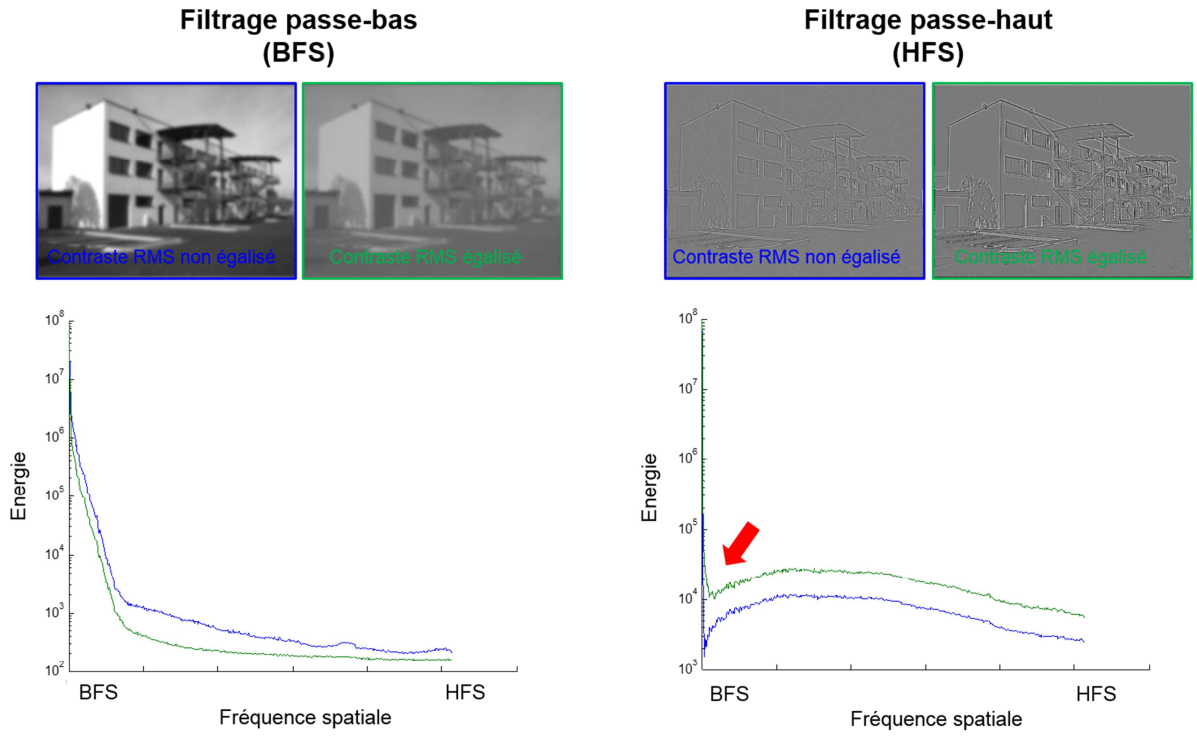


Figure 33 : Illustration de l'effet de l'égalisation RMS sur le spectre d'énergie d'une scène filtrée en BFS (à gauche) et d'une scène filtrée en HFS (à droite). On constate que pour les scènes en HFS, l'égalisation du contraste RMS (ligne verte) a pour effet d'ajouter de l'énergie dans les BFS (indiqué par la flèche rouge).

Ces considérations nous amènent donc à nous positionner en faveur d'une utilisation prudente de l'égalisation du contraste RMS pour l'étude du traitement des fréquences spatiales, et d'une prise en compte systématique de ce facteur méthodologique pour l'interprétation de résultats et la comparaison entre les études. Nous proposons que dans les cas où l'objectif de recherche n'impose pas l'étude séparée de l'effet du contraste et de l'effet des fréquences spatiales, il serait plus adapté d'interpréter les résultats en prenant explicitement en compte le fait que ces deux facteurs contribuent de façon conjointe aux mécanismes étudiés. Notons par ailleurs qu'un raisonnement similaire pourrait être appliqué en ce qui concerne le contrôle des orientations dominantes des scènes de différentes catégories utilisées comme stimuli (cf. **Expériences 1, 2, 3, 5 et 6**). En effet, l'information liée aux orientations dans les scènes est connue pour contribuer également à leur catégorisation (e.g., Guyader et al., 2004). Bien que dans le cadre de nos études les scènes ou leur spectre d'amplitude n'aient pas été directement altérés, il est cependant possible que le contrôle de la similitude des orientations dominantes entre les scènes des différentes catégories ait pu également affecter les stratégies d'analyses lors de la catégorisation.

2. Vers l'élaboration d'un modèle neurocognitif de la catégorisation de scènes

A travers quatre études en IRMf, nous avons tenté de préciser les bases cérébrales du traitement des fréquences spatiales et de l'analyse « coarse-to-fine » lors de la catégorisation de scènes. Ces quatre études ont permis de fournir des données complémentaires pour l'élaboration d'un modèle neurocognitif de la catégorisation de scènes, en soulignant notamment (1) le rôle du cortex visuel primaire (**Expérience 3**), (2) les fonctions intégratives et/ou modulatrices du cortex frontal inférieur (**Expérience 4**), et (3) l'implication différentielle des régions occipito-temporales sélectives aux scènes lors du traitement des fréquences spatiales (**Expériences 5 et 6**).

2.1. Le cortex visuel : région intégrative de l'information visuelle ?

Dans l'**Expérience 3**, nous avons utilisé comme stimuli les séquences dynamiques adaptées de l'**Expérience 1** afin d'imposer une analyse « coarse-to-fine » (séquence CtF) ou « fine-to-coarse » (séquence FtC) des scènes. Nous avons montré que les séquences CtF

(relativement aux séquences FtC) activaient plus fortement le cortex visuel primaire, les gyri fusiforme et parahippocampique, et le gyrus frontal inférieur dans sa partie orbitale. Ces activations sont très cohérentes avec le réseau cérébral du modèle proactif proposé par Bar et collaborateurs (Bar, 2003 ; Bar et al., 2006 ; Kveraga et al., 2007). Ce modèle postule que lors de l'analyse d'un stimulus visuel, l'information en BFS serait rapidement projetée vers le cortex orbitofrontal afin de générer des interprétations possibles de l'identité du stimulus. Le résultat de ces prédictions serait ensuite intégré, via des influences « top-down », au traitement de l'information en HFS acheminée plus lentement le long de la voie ventrale jusqu'au cortex inférotemporal, et faciliterait ainsi la reconnaissance du stimulus visuel.

Cependant, d'autres travaux suggèrent que le traitement des afférences magnocellulaires (comme les BFS) au sein du cortex préfrontal, pourrait influencer le traitement ultérieur des afférences parvocellulaires (comme les HFS) jusqu'au cortex visuel primaire (Bullier, 2001 ; Peyrin et al., 2010). Nous avons donc effectué une analyse de connectivité effective sur les données de l'**Expérience 3** afin de spécifier les interactions sous-jacentes au réseau cortical mis en évidence dans cette étude, et notamment celles impliquant le cortex visuel primaire. Les résultats de cette analyse ont montré que l'analyse « coarse-to-fine » des scènes avait pour effet d'augmenter la connectivité du cortex visuel vers le gyrus frontal inférieur et vers le gyrus fusiforme, et du gyrus frontal inférieur vers le gyrus fusiforme. Ces résultats sont bien compatibles avec le réseau de connectivité effective entre le cortex occipital, le cortex préfrontal et le cortex inférotemporal mis en évidence dans les travaux de Kveraga et al. (2007) dans le cadre de la reconnaissance d'objets. Mais un nouveau résultat apparaît dans notre étude. La catégorisation des séquences CtF avait également pour effet d'augmenter la connectivité du gyrus frontal inférieur vers le cortex visuel primaire. Ce dernier résultat va donc dans le sens d'une influence « top-down » du cortex frontal vers les aires visuelles primaires (Bullier, 2001 ; Peyrin et al., 2010) dans le but supposé de guider le traitement ultérieur de l'information visuelle. Ainsi, le cortex visuel primaire ne serait pas un simple « point d'entrée » sensoriel de l'information visuelle mais serait un relais important dans le réseau cérébral impliqué dans la catégorisation de scènes, en intégrant à la fois les afférences rétino-thalamiques et les influences « top-down » de régions de haut niveau.

Plusieurs travaux ont déjà démontré que l'activité du cortex visuel primaire pouvait être influencée par des processus de haut niveau, tels que l'attention visuo-spatiale (Brefczynski & Deyoe, 1999; Gandhi, Heeger, & Boynton, 1999; Martínez et al., 1999;

Sasaki et al., 2001; Saygin & Sereno, 2008; Silver et al., 2007; Tootell, Hadjikhani, Hall, et al., 1998; Watanabe et al., 1998), l'apprentissage perceptuel (Li, Piëch, & Gilbert, 2004; Pourtois, Rauss, Vuilleumier, & Schwartz, 2008; Schwartz, Maquet, & Frith, 2002) ou la valeur accordée à des stimuli visuels, sur la base des récompenses précédemment associées à ces stimuli (Serences, 2008). Ces influences « top-down » sur les aires visuelles primaires pourraient provenir de régions cérébrales supérieures des cortex pariétal, temporal et frontal (Bestmann et al., 2008; Koivisto et al., 2011; Ruff et al., 2006, 2009). Dans une revue récente, Rauss et al. (2011) ont proposé que ces effets « top-down » sur le cortex visuel primaire puissent être expliqués par un mécanisme de codage prédictif au sein du cortex visuel (« predictive coding» ; voir aussi Friston, 2005; Lee et al., 1998; Lee, 2003; Rao & Ballard, 1999; Rauss & Pourtois, 2013; Summerfield & Egner, 2009). Dans ce contexte, les réponses des aires visuelles seraient moins liées au stimulus visuel *per se* qu'au fait qu'il corresponde plus ou moins à des attentes et des prédictions formées sur la base d'expériences précédentes et des contingences apprises dans l'environnement. Selon ce modèle théorique, les prédictions générées par des aires corticales supérieures influenceraient le traitement du cortex visuel de telle sorte que le signal entrant serait comparé au signal prédit, et seul le signal ne correspondant pas aux prédictions (i.e. erreur de prédiction) serait transmis vers les aires de plus haut niveau, permettant de mettre à jour les prédictions et de réduire les coûts de traitement (cf. Figure 34a). Ces va-et-vient entre les prédictions et les signaux d'erreur de prédictions résulteraient en une augmentation de l'activité du cortex visuel primaire d'autant plus importante que les erreurs de prédictions seraient importantes (de-Wit, Machilsen, & Putzeys, 2010; Price & Devlin, 2011, cf. Figure 34b).

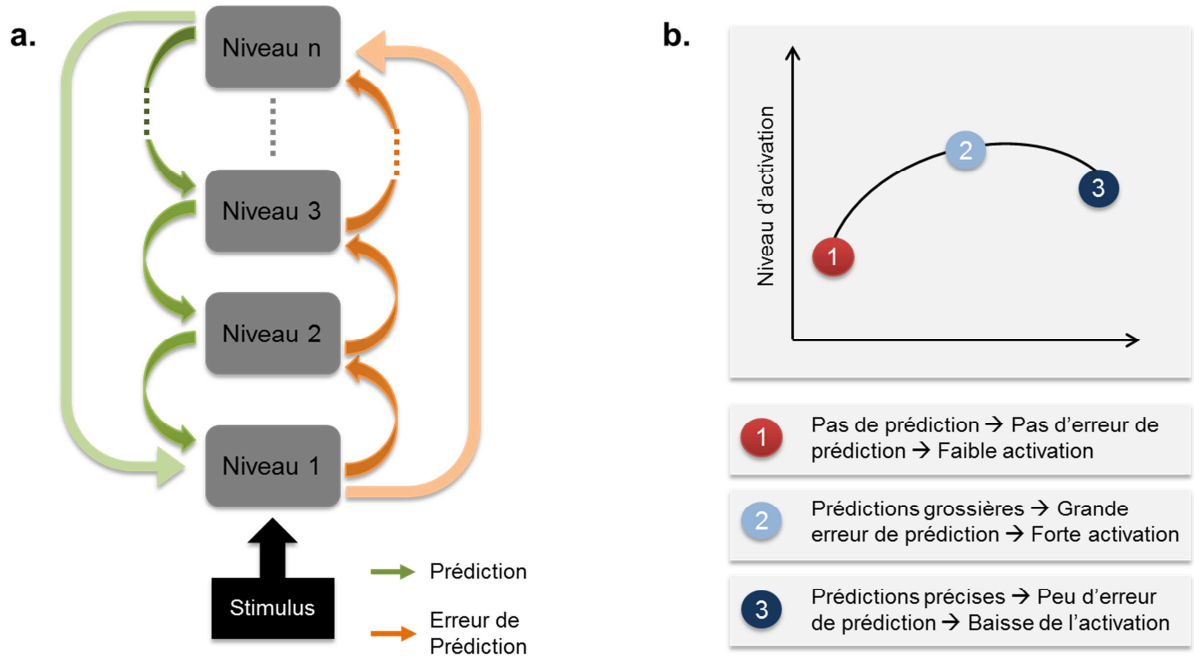


Figure 34 : Schématisation du mécanisme de codage prédictif. (a) Les prédictions générées dans les régions de haut niveau dans la hiérarchie influencent les traitements effectués dans les régions de plus bas niveau : le signal traité dans les régions de bas niveau est comparé aux prédictions générées par les régions de plus haut niveau dans la hiérarchie. Seule l'erreur de prédiction est transmise aux niveaux supérieurs. Ces interactions peuvent se faire entre niveaux voisins (flèches foncées) ou entre des niveaux plus éloignés dans la hiérarchie (flèches claires). Adapté de Rauss et Pourtois (2013). (b) Niveau d'activation en fonction de la nature des prédictions. Adapté de Price et Devlin (2011).

Dans le cadre de ce modèle, l'activité plus importante du cortex visuel lors de la catégorisation des séquences CtF pourrait donc refléter un codage prédictif de l'information par le cortex visuel primaire, grâce aux prédictions générées au niveau du cortex frontal inférieur sur la base des BFS. Les prédictions grossières formées sur la base des BFS seraient susceptibles d'entraîner des erreurs de prédiction importantes, résultant en une activité plus importante du cortex visuel. Des études plus approfondies manipulant explicitement la congruence entre les prédictions générées et le signal réel pourraient permettre de tester directement cette hypothèse. En effet, dans le cadre théorique du codage prédictif, on pourrait s'attendre à ce que l'activité du cortex visuel soit plus importante en cas de prédictions incongruences que congruentes. Les résultats de l'**Expérience 4,** relatifs à l'effet de la similarité physique des scènes composant les hybrides, apportent effectivement des arguments en faveur de cette hypothèse. Le contraste entre l'activation cérébrale induite par la catégorisation des hybrides dont l'information en BFS et en HFS était physiquement dissimilaire à celle induite par la catégorisation des hybrides dont l'information en BFS et en HFS était physiquement similaire, indépendamment de la similarité sémantique (contraste

[DP>SP] a révélé une activité plus importante du cortex occipital, incluant le cuneus, le gyrus lingual et les gyri occipitaux inférieur et moyen. Comme nous l'avons vu tout au long de ce manuscrit, ces régions sont impliquées dès les premières étapes de traitement de l'information visuelle, dans l'extraction d'éléments locaux à différents points du champ visuel. Dans le contexte d'un codage prédictif au sein du cortex occipital, il est possible que lors du traitement des hybrides, le traitement rapide de l'information en BFS ait conduit à former des prédictions sur des points particuliers du champ visuel (e.g., présence de contour) à traiter dans les HFS. Dans le cas où les hybrides étaient physiquement dissimilaires, ces prédictions auraient été erronées (i.e. absence de contours aux endroits prédits), conduisant à une erreur de prédiction plus importante, et donc à une activation plus importante au sein du cortex occipital.

L'interprétation des autres résultats de l'**Expérience 4** au regard de l'hypothèse d'un codage prédictif sera reprise plus détail dans les parties qui suivent.

2.2. Le cortex frontal inférieur : prédiction, intégration sémantique ou pilotage du traitement visuel ?

Dans l'**Expérience 3**, nous avons postulé que l'activation plus importante du gyrus frontal inférieur (cortex orbitofrontal, AB 47) lors de la catégorisation des séquences CtF que FtC reflétait son implication rapide dans la génération de prédictions quant à l'identité possible de la scène, et ce, sur la base des BFS en accord avec le modèle proactif de Bar et collaborateurs (Bar, 2003, 2007; Bar et al., 2006; Kveraga et al., 2007). Les résultats de l'analyse de connectivité suggèrent par ailleurs que cette région influencerait l'activité du cortex visuel primaire et du cortex parahippocampique. Nous avons précédemment proposé que ces influences puissent être expliquées en termes de codage prédictif de l'information visuelle, permettant de faciliter le traitement ultérieur de l'information et la catégorisation. Plusieurs travaux suggèrent en effet que la partie orbitale du gyrus frontal inférieur serait particulièrement impliquée lors de la génération de prédictions et d'attentes relatives à un stimulus (Bar et al., 2006; Carlsson et al., 2000; Chaumon et al., 2014; Fink, Halligan, et al., 1997; Petrides et al., 2002; Thorpe et al., 1983). Carlsson et al. (2000) ont par exemple montré que l'anticipation d'une stimulation tactile entraînait une activité plus importante du cortex orbitofrontal, par rapport à la stimulation elle-même. Dans une étude IRMf récente, Chaumon et al. (2014) ont tenté de préciser le rôle et les propriétés fonctionnelles de cette région lors de la reconnaissance d'objets. Ces auteurs ont présenté des objets et des réseaux sinusoïdaux

filtrés en BFS ou en HFS aux participants qui avaient pour tâche de détecter la répétition d'un de ces stimuli présentés deux essais auparavant. Les résultats ont révélé une activation plus importante du cortex orbitofrontal lors de la perception de stimuli en BFS qu'en HFS, mais seulement lorsque ceux-ci étaient des objets. Ces résultats suggèrent donc que l'implication du cortex orbitofrontal dans la perception visuelle n'est pas simplement liée au traitement d'une information en BFS, mais est spécifique à une information en BFS porteuse de sens et permettant de générer des prédictions (contrairement aux réseaux sinusoïdaux). Par ailleurs, une analyse de connectivité a montré que plus les objets étaient identifiables, plus la connectivité entre le cortex orbitofrontal et les régions inférotemporales augmentait. Ces résultats soutiennent donc l'idée d'un rôle prédictif du cortex orbitofrontal et soulignent également l'importance des propriétés sémantiques des stimuli pour cela.

Dans l'**Expérience 4**, nous avons montré que la catégorisation de scènes hybrides, sur la base de l'information en HFS, était associée à une activation plus importante du gyrus frontal inférieur (incluant la partie orbitale) lorsque l'information en BFS des hybrides était sémantiquement dissimilaire à l'information en HFS. Dans le cadre du modèle proactif de Bar et collaborateurs (Bar et al., 2006 ; Bar 2007), nous avons proposé que cette activation plus importante résultait du fait que le traitement rapide et automatique de l'information en BFS des hybrides aurait conduit à générer des prédictions erronées au sein de cette région. Cela suppose cependant que le cortex frontal inférieur ait également traité l'information sémantique contenue dans les HFS. Dans leur étude princeps, Mu et Li (2013) ont en effet proposé que le cortex frontal soit impliqué de façon précoce dans l'intégration de l'information sémantique issue des BFS, mais aussi des HFS. Cela suggère donc un rôle intégratif plutôt que prédictif du cortex frontal inférieur. Comme nous l'avons suggéré dans l'**Expérience 4**, l'intégration de l'information sémantique contenue dans les deux bandes de fréquences spatiales au sein du cortex frontal inférieur pourrait être permise par le biais d'interactions réciproques entre le cortex frontal inférieur et le cortex inférotemporal. C'est en effet ce que suggère une étude récente de Fintzi et Mahon (2013) qui ont examiné la réponse du cortex orbitofrontal, inférotemporal et occipital à des séquences d'objets filtrés passant progressivement des BFS aux HFS (cf. Figure 35a). Ces auteurs ont montré que la réponse du cortex orbitofrontal présentait deux pics au cours du temps : le premier lors de la présentation de la première image de la séquence (BFS) et le second lors de la présentation des dernières images de la séquence (MFS et HFS, cf. Figure 35b). En analysant séparément les réponses du cortex orbitofrontal de chaque hémisphère, les auteurs ont par ailleurs montré que le cortex

orbitofrontal droit répondait de façon maximale spécifiquement à l'image en BFS des séquences, tandis que le cortex orbitofrontal gauche présentait un pattern de réponses similaires à celles du cortex inférotemporal, répondant de façon maximale aux MFS et HFS (cf. Figure 35c).

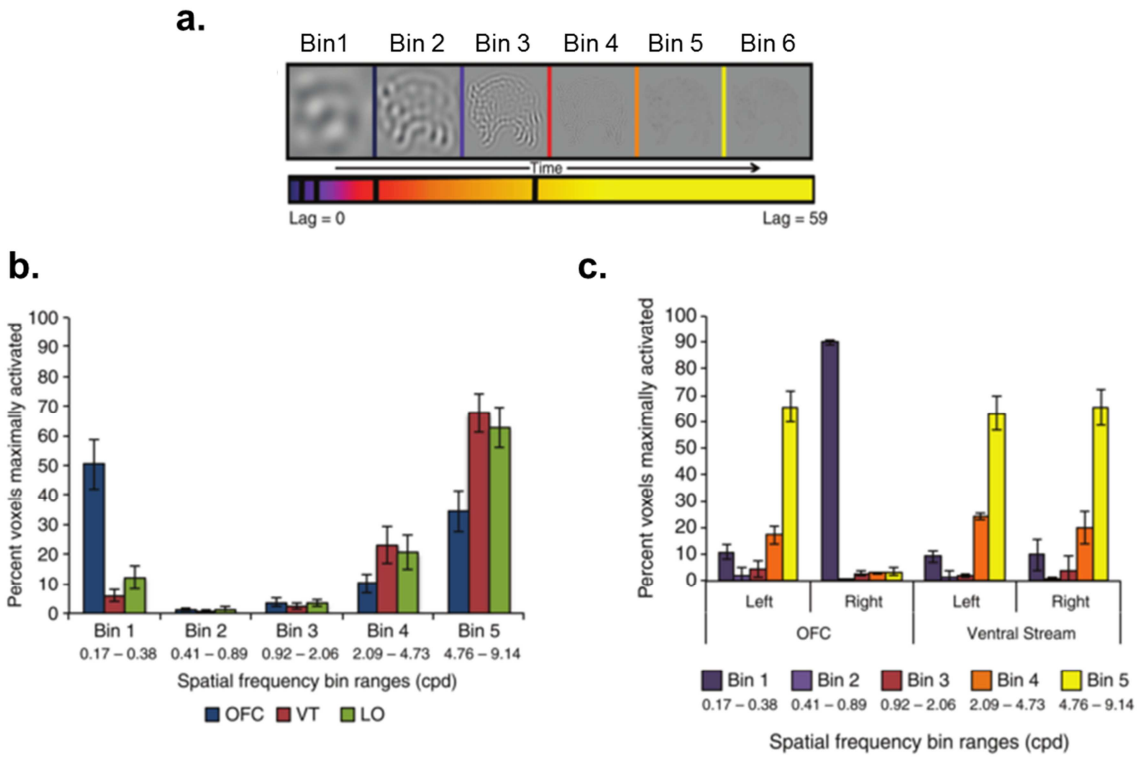


Figure 35 : Résumé des résultats obtenus par Fintzi et Mahon (2013). (a) Exemple de séquence d'objet passant des BFS (Bin1) aux HFS (Bin 6). (b) Pourcentage de voxels activés de façon maximale au sein du cortex orbitofrontal (OFC), inférotemporal (VT) et occipital (LO) en fonction du contenu en fréquences spatiales dans la séquence. (c) Pourcentage de voxels activés de façon maximale au sein du cortex orbitofrontal et du cortex inférotemporal en fonction de l'hémisphère et du contenu en fréquences spatiales dans la séquence. Adapté de Fintzi et Mahon (2013).

Ces auteurs ont proposé que le cortex orbitofrontal droit soit rapidement impliqué dans le traitement des BFS, afin de générer des prédictions sur l'identité possible du stimulus qui déclencheraient des influences « top-down » vers le cortex inférotemporal. Le cortex orbitofrontal gauche accèderait quant à lui à l'identité de l'objet sur la base d'une information plus fine en MFS et HFS. Selon les auteurs, le cortex orbitofrontal recevrait également des influences du cortex inférotemporal, qui viendraient valider ou infirmer les prédictions grossières effectuées sur la base des BFS, permettant de renforcer ou non, via des interactions inter-hémisphériques, ces prédictions en cas de traitement ultérieur de cette même information. Dans ce contexte, l'activité plus importante du cortex frontal inférieur lors de la

catégorisation d'hybrides sémantiquement dissimilaires dans l'**Expérience 4** pourrait donc résulter de l'incompatibilité entre les prédictions générées sur la base des BFS et les influences provenant du cortex inférotemporal relatives à l'information détaillée contenue dans les HFS. Cette interprétation apparaît également cohérente avec les résultats de notre analyse de connectivité fonctionnelle (analyse PPI) effectuée dans cette étude, et qui suggère des interactions plus importantes entre le gyrus fusiforme droit et le gyrus frontal inférieur droit en cas d'interférence sémantique des BFS sur les HFS.

Dans l'**Expérience 4**, nous avons également suggéré une interprétation alternative du rôle du cortex frontal inférieur dans le contexte de l'interférence sémantique. En effet, nous avons proposé que celui-ci puisse être impliqué dans le pilotage et le contrôle du traitement de l'information visuelle, par exemple en amplifiant/inhibant le traitement de l'information en fréquence spatiale pertinente/non-pertinente pour la tâche. En effet, plusieurs travaux ont mis en évidence l'implication du gyrus frontal inférieur (AB 44-45) lors de la gestion de l'interférence ou de tâche nécessitant une résolution de conflit (e.g., tâche de Stroop, tâche de Go/no-Go; voir pour une méta-analyse Nee, Wager, & Jonides, 2007). Des études récentes suggèrent que le gyrus frontal inférieur agirait comme un filtre, afin de privilégier le traitement des informations pertinentes pour une tâche donnée (Chong et al., 2008; DiQuattro & Geng, 2011; Hampshire, Chamberlain, Monti, Duncan, & Owen, 2010; Miller & Cohen, 2001; Nelson et al., 2009; Wais et al., 2010). Dans ce contexte, le gyrus frontal inférieur aurait donc un rôle plus large, incluant le contrôle cognitif et le pilotage du traitement de l'information.

Il semblerait que les divergences apparentes concernant les fonctions du cortex frontal inférieur puissent être cependant intégrées sur la base d'une dissociation entre la partie orbitale (AB 47) et la partie triangulaire (AB 45) du gyrus frontal inférieur. En effet, plusieurs travaux suggèrent que la partie orbitale du gyrus frontal inférieur serait principalement impliquée lors de la récupération d'informations sémantiques relatives à un stimulus, tandis que la partie triangulaire du gyrus frontal inférieur interviendrait plutôt lors de la sélection de l'information pertinente pour la tâche parmi les différentes informations récupérées (Badre et al., 2005; Badre & Wagner, 2007; Gold et al., 2006). Dans ce contexte, les activations du cortex orbitofrontal observées dans le cadre de nos travaux pourraient donc refléter la génération de prédictions relatives à la catégorie sémantique du stimulus visuel, sur la base des BFS (**Expériences 3 et 4**), et les activations de la partie triangulaire du gyrus frontal inférieur pourraient correspondre au processus de sélection de l'information pertinente pour

réaliser la tâche (**Expérience 4**). Des études plus approfondies, avec des conditions expérimentales permettant de mieux dissocier les activations liées aux prédictions (e.g. manipulation du caractère prédictible *vs*. non prédictible), des activations liées à la gestion de l'interférence (e.g., manipulation de la congruence *vs*. incongruence entre deux stimuli), ainsi que des analyses de connectivité entre les différentes parties du cortex frontal inférieur permettraient de préciser les fonctions de cette région dans le contexte de la catégorisation de scènes.

2.3. Les régions du cortex occipito-temporales : des relais majeurs pour l'analyse de scènes ?

Les résultats des **Expériences 3 et 4** soulignent également l'implication de régions inférotemporales, et notamment du gyrus parahippocampique au sein du réseau sous-jacent à la catégorisation « coarse-to-fine » des scènes. Dans l'**Expérience 3**, nous avons montré que ces régions recevaient des influences « top-down » du cortex frontal inférieur. Dans l'**Expérience 4**, nous avons observé une activation plus importante de ces régions lors de la catégorisation des hybrides dont l'information en BFS et en HFS était sémantiquement dissimilaire que similaire. Sur la base du modèle proactif de Bar et collaborateurs (Bar, 2003, 2007; Bar et al., 2006; Kveraga et al., 2007), nous avons proposé que ces régions, connues pour être impliquées lors des dernières étapes de la reconnaissance (Epstein & Kanwisher, 1998; Grill-Spector, 2003; Haxby et al., 2001; Ishai et al., 1999; Lerner et al., 2001; Tanaka, 1996), intègrent les prédictions grossières effectuées sur la base de BFS au sein du cortex orbitofrontal au traitement « bottom-up » de l'information en provenance du cortex visuel, acheminée le long de la voie ventrale, et ce, dans le but de faciliter la reconnaissance. Comme pour le cortex visuel primaire, il est possible que cette intégration se fasse également en termes de codage prédictif au sein du cortex inférotemporal, sur la base des prédictions grossières générées par le cortex orbitofrontal. L'erreur de prédiction serait ensuite transmise au cortex orbitofrontal (Fintzi & Mahon, 2013). Dans ce contexte, l'activation plus importante des gyri fusiforme et parahippocampique induite par les hybrides sémantiquement dissimilaires que similaires dans l'**Expérience 4** pourrait être expliquée par le fait que les prédictions générées sur la base des BFS aient été erronées, conduisant à une erreur de prédiction – et donc une activation – plus importante au sein du cortex inférotemporal (cf. Figure 34b).

Ces résultats nous ont amenés à nous interroger plus généralement sur le rôle des régions occipito-temporales sélectives aux scènes dans le traitement des fréquences spatiales des scènes. Dans les **Expériences 5 et 6**, nous avons spécifiquement examiné comment la PPA, le RSC et l'OPA traitaient l'information en fréquences spatiales dans les scènes. Les résultats de l'**Expérience 5** ont révélé des patterns de réponse aux BFS et HFS spécifiques à chaque région. Nous avons montré que la PPA et le RSC répondaient plus fortement aux BFS qu'aux HFS, tandis que l'OPA était sélective aux HFS. De plus, nous avons mis en évidence que la PPA et le RSC étaient sensibles au contraste associé aux BFS et aux HFS. Lorsque le contraste était égalisé entre les BFS et les HFS, le pattern de réponse au sein de la PPA s'inversait en faveur des HFS, et la différence de réponse entre les BFS et les HFS au sein du RSC n'était plus observée. Ces résultats suggèrent que la PPA est sensible au contraste associé aux fréquences spatiales et répond à des interactions particulières entre ces deux caractéristiques en répondant plus fortement aux BFS lorsque celles-ci sont associées à un contraste plus important que les HFS, mais en privilégiant les HFS lorsque leur contraste est rehaussé (alors que le contraste des BFS est diminué ; condition d'égalisation du contraste). Le RSC serait, quant à lui, principalement sensible à de grandes différences de contraste. Il répondrait donc plus fortement aux BSF dont le contraste est plus élevé que les HFS, mais ne privilégierait aucune de ces informations en l'absence de différence de contraste entre elles. Enfin, l'OPA serait sélective aux HFS, quel que soit le contraste qui y est associé.

Dans l'**Expérience 6**, nous avons examiné si un mode de traitement « coarse-to-fine » ou « fine-to-coarse » était privilégié au sein de ces régions. Nous avons montré que seule la PPA répondait de façon distincte aux deux types de séquences et privilégiait les séquences « coarse-to-fine ». En résumé, les résultats des **Expériences 5 et 6** montrent distinctement l'implication des régions sélectives aux scènes dans le traitement des fréquences spatiales et nous permet de faire des hypothèses sur l'information privilégiée par ces régions pour remplir les fonctions complexes qui leur sont attribuées (navigation, encodage de l'agencement spatial, reconnaissance de scènes, associations contextuelles ; voir pour des revues Aminoff et al., 2013; Bar & Aminoff, 2003; Epstein, 2005, 2008). Par exemple, le traitement de l'information grossière et très contrastée des BFS au sein de la PPA et du RSC pourrait permettre d'extraire les invariants pertinents pour avoir accès à la structure globale de l'environnement lors de la navigation (e.g., Auger et al., 2012; Epstein, 2008). L'analyse ultérieure des HFS au sein de la PPA permettrait d'affiner ces représentations et d'identifier la scène de façon spécifique à son point de vue (e.g., Epstein & Higgins, 2007; Rajimehr et al.,

2011). L'OPA semble au contraire sélective aux HFS, suggérant que cette région contribuerait à l'extraction des contours et de détails fins dans l'agencement spatial des scènes (Dilks et al., 2013).

Malheureusement nos données ne nous permettent pas de conclure quant à l'utilisation par ces régions de l'information en fréquences spatiales pour la construction de représentations de plus haut niveau lors de la perception de scènes. Par ailleurs, la contribution relative de ces régions ainsi que leurs interactions permettant d'aboutir à la reconnaissance de scènes restent également à déterminer. En effet, une meilleure compréhension des rôles relatifs et des interactions entre les régions sélectives aux scènes permettrait de les intégrer au sein du réseau plus large impliqué dans la catégorisation de scènes (cf. **Expériences 3 et 4**). Pour l'heure, certains travaux tendent à suggérer de fortes interactions entre la partie postérieure de la PPA et l'OPA d'une part, et entre la partie antérieure de la PPA et le RSC d'autre part, aux propriétés fonctionnelles différentes (Baldassano et al., 2013; Nasr et al., 2013). La partie postérieure de ce réseau serait principalement impliquée dans le traitement de caractéristiques physiques de bas niveau des scènes, tandis que la partie antérieure contribuerait plus à la construction de représentations plus globales et abstraites (Korkmaz Hacialihafiz & Bartels, 2015). Par ailleurs, les résultats des **Expériences 3 et 4** suggèrent de fortes interactions entre le cortex orbitofrontal et la PPA, lors de la catégorisation de scènes. Cependant, à notre connaissance, le rôle des autres régions sélectives aux scènes au sein du réseau cérébral impliqué dans la catégorisation de scènes n'a jamais été directement étudié.

En résumé, les résultats des travaux présentés dans ce travail de thèse nous permettent d'appuyer et de préciser les modèles neurocognitifs de la perception de scènes basés sur le traitement des fréquences spatiales (Bar, 2003, 2007; Bar et al., 2006; Bullier, 2001; Peyrin et al., 2010; Schyns & Oliva, 1994). Sur la base de ces précédents travaux et des résultats de nos études, nous proposons donc que l'analyse visuelle de scènes se ferait selon un mode de traitement par défaut « coarse-to-fine », avantageux pour la catégorisation rapide des scènes (**Expériences 1 et 2**). Lors de la catégorisation « coarse-to-fine » des scènes, l'information grossière en BFS activerait le cortex visuel primaire et serait rapidement projetée vers le cortex orbitofrontal, afin de générer des prédictions sur la catégorie de la scène. Ces prédictions permettraient d'influencer et de guider le traitement ultérieur de l'information en HFS au niveau du cortex inférotemporal (**Expériences 3 et 4**), mais aussi dès le cortex visuel

primaire (**Expérience 3**), possiblement par le biais d'un mécanisme de codage prédictif au sein de ces régions. Dans ce contexte, l'erreur de prédiction serait transmise au cortex frontal inférieur afin d'affiner et de sélectionner, de façon itérative, l'interprétation la plus pertinente permettant finalement de catégoriser et d'identifier correctement la scène (**Expérience 4**). Au sein de ce réseau cortical dédié à l'analyse « coarse-to-fine », les régions occipito-temporales sélectives aux scènes seraient différemment impliquées dans le traitement fréquentiel de l'information visuelle (**Expériences 5 et 6).**

Le modèle proposé est illustré dans la Figure 36, qui résume plus largement les bases cérébrales du traitement des fréquences spatiales et de la catégorisation de scènes que nous avons développées dans cette thèse. Ce modèle comprend donc le cortex occipital, au sein duquel le traitement des fréquences spatiales est représenté de façon rétinotopique et latéralisée entre les hémisphères (cf. Musel et al., 2013). Celui-ci est également impliqué dans l'intégration des influences « top-down » du cortex frontal inférieur aux afférences rétino-thalamiques, possiblement par le biais d'un mécanisme de codage prédictif. On y retrouve également les régions du cortex inférotemporal qui intègrent les influences « top-down » du cortex frontal inférieur au traitement « bottom-up » de l'information via la voie ventrale, et les aires occipito-temporales sélectives aux scènes, qui traitent de façon distincte l'information en fréquences spatiales des scènes. Enfin, ce modèle comprend le cortex frontal inférieur dont les différentes subdivisions (i.e. partie orbitale et triangulaire) pourraient être impliquées dans des processus distincts lors du traitement de l'information visuelle et de la catégorisation de scènes.

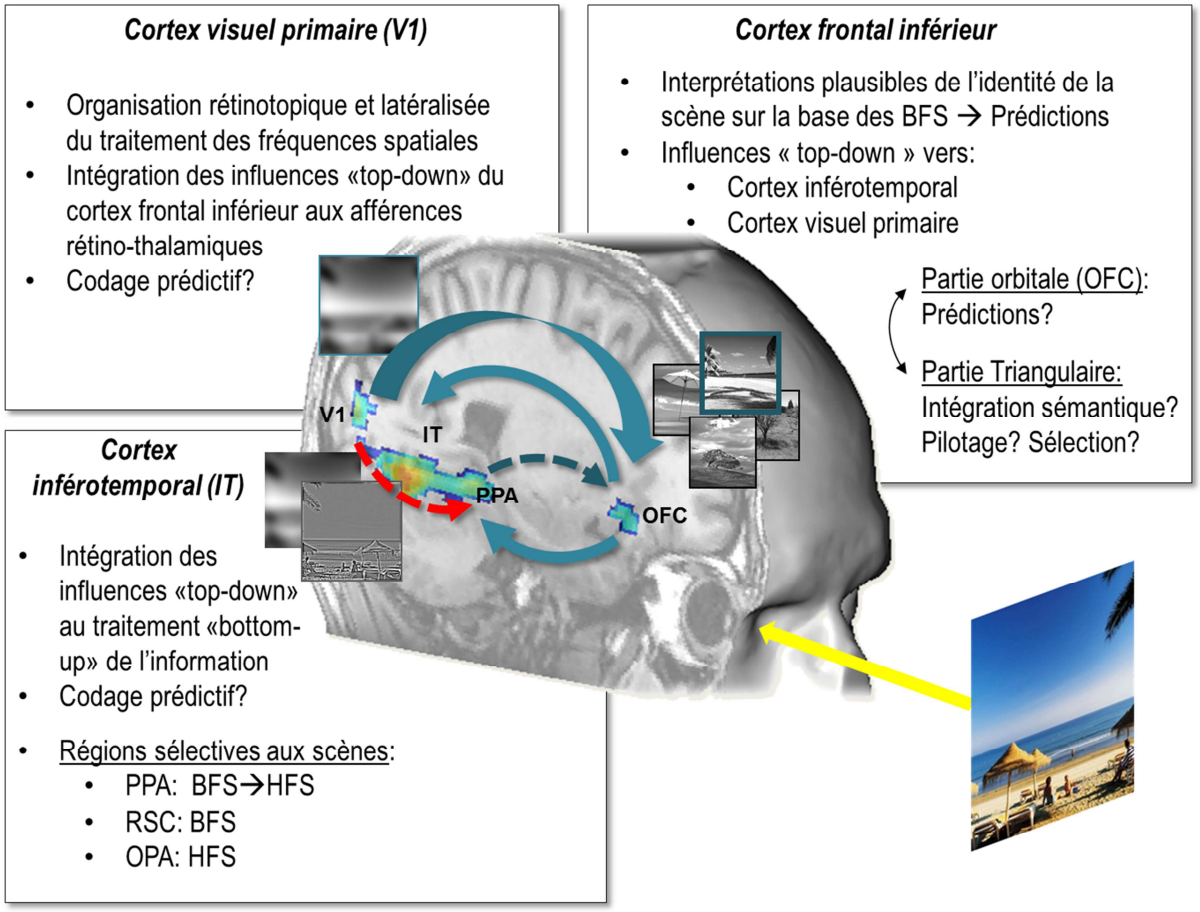


Figure 36 : Modèle neurocognitif de la catégorisation de scènes proposé sur la base des résultats obtenus dans ce travail de thèse.

3. Limites et perspectives

3.1. Précision du modèle via des aspects temporels et anatomiques

Les travaux présentés dans cette thèse nous ont permis de préciser et d'enrichir les données existantes concernant les bases cérébrales de la perception de scènes chez le jeune adulte sain. Ces travaux présentent cependant certaines limites que des études complémentaires pourraient permettre de pallier. L'une des principales limites de ces travaux vient du fait que ceux-ci se situent dans le cadre de modèles neurocognitifs pour lesquels la dimension temporelle est cruciale. En effet, les modèles sur lesquels nous nous sommes appuyés postulent, par exemple, l'implication précoce des aires frontales lors du traitement des BFS, suivie des aires occipito-temporales (Bar et al., 2006; Bullier, 2001; Peyrin et al., 2010). Or, tous les travaux présentés ici ont été réalisés dans le cadre d'études en IRMf qui présente l'inconvénient d'avoir une très mauvaise résolution temporelle. Nos mesures nous permettent donc difficilement de conclure quant au décours temporel des activations

observées. L'interprétation de nos résultats est largement inférée de précédents travaux précédents utilisant des techniques de haute résolution temporelle. Il semblerait donc très important pour la suite de ces travaux d'utiliser d'autres techniques de neuroimagerie ayant une meilleure résolution temporelle, telles que l'EEG ou la TMS. La TMS semblerait particulièrement adaptée pour tester directement l'influence du cortex frontal inférieur sur le cortex visuel primaire. En effet, si ces influences permettent de faciliter le traitement ultérieur de l'information visuelle, on pourrait s'attendre à ce qu'une stimulation inhibitrice sur le cortex frontal inférieur entraîne une perturbation lors de la catégorisation de scènes.

Il reste par ailleurs à déterminer quelle connectivité anatomique pourrait sous-tendre ces influences. Rappelons que la méthode d'analyse de connectivité utilisée dans ces travaux (DCM) permet seulement de rendre compte de l'influence qu'exerce une région sur l'activité d'une autre (i.e. connectivité effective). Cependant, la connectivité effective entre deux régions n'implique pas nécessairement l'existence d'une connexion anatomique directe entre ces régions. Dans l'**Expérience 3**, nous avons évoqué les liens de connectivité anatomiques susceptible de sous-tendre les interactions entre le cortex visuel primaire, le cortex inférotemporal et le cortex orbitofrontal. Si les connexions entre le cortex visuel et le cortex inférotemporal d'une part, et le cortex inférotemporal et le cortex orbitofrontal d'autre part sont bien connues (Catani et al., 2003; Catani & Thiebaut de Schotten, 2008; Cavada et al., 2000; Ghashghaei & Barbas, 2002; Powell et al., 2004), les connexions anatomiques qui permettraient l'activation rapide du cortex orbitofrontal et son influence « top-down » sur le cortex visuel restent obscurs. Dans leur étude, Kveraga et al. (2007) ont suggéré que l'implication rapide du cortex orbitofrontal lors de la reconnaissance d'objets se ferait par le biais d'afférences magnocellulaires. Ces auteurs ont proposé que la voie dorsale serait un chemin plausible pour supporter ce mécanisme, de par la prédominance de fibres de type magnocellulaire le long de cette voie. Il est cependant à noter que le cortex visuel et le cortex orbitofrontal ne sont pas connectés de façon directe le long de la voie dorsale, mais par l'intermédiaire du cortex pariétal (Cavada et al., 2000; Kravitz, Saleem, et al., 2011). Kveraga et al. (2007) proposent également une voie alternative sous-corticale, via le pulvinar ou le thalamus, puis l'amygdale, qui entretient des liens de connectivité anatomique importants avec le cortex préfrontal (Ghashghaei & Barbas, 2002; Ghashghaei, Hilgetag, & Barbas, 2007) et qui est fonctionnellement connue pour être impliquée dans le traitement des BFS (Vuilleumier, Armony, Driver, & Dolan, 2003). Il apparaît donc important de déterminer précisément les substrats et les circuits anatomiques, et éventuellement le rôle des structures

sous-corticales, qui pourraient permettre, d'une part, l'implication précoce du cortex orbitofrontal, et d'autre part, l'influence du cortex orbitofrontal sur le cortex visuel.

2.2 Influence de facteurs individuels, développementaux et pathologiques

Les travaux présentés dans cette thèse ouvrent par ailleurs de nombreuses perspectives de recherches. Tout d'abord, ils nous amènent à nous poser la question d'un potentiel effet des différences interindividuelles en termes de correction des troubles visuels ophtalmologiques. En effet, l'ensemble de nos études ont été menées sur des participants ayant une vision normale ou corrigée à la normale. Il semble cependant que la présence de myopie[3], même corrigée, influence les stratégies d'analyses visuelles. Par exemple, les myopes sont plus performants que les emmétropes (i.e. participants dans défaut visuel) pour s'adapter à des environnements floutés (George & Rosenfield, 2004; Mon-Williams, Tresilian, Strang, Kochhar, & Wann, 1998; Pesudovs & Brennan, 1993; Rosenfield & Abraham-Cohen, 1999). Une étude de Giraudet & Azavant (2006) a également montré que des participants myopes (dont la vision était corrigée) étaient plus performants que les emmétropes pour localiser un objet dans une scène filtrée en BFS, car ils avaient tendance à plus s'appuyer sur le contexte spatial de la scène pour effectuer la tâche. Cela suggère donc que les myopes, même après correction de l'erreur de réfraction, mettent en place des stratégies d'analyse visuelle différentes de celles des emmétropes leur permettant de mieux s'adapter à certaines contraintes visuelles. Cependant, très peu d'études ont directement testé si les myopes et les emmétropes avaient des performances similaires sur des tâches évaluant explicitement la catégorisation de scènes dont le contenu en fréquences spatiales est manipulé. Dans ce contexte, nous avons mené, en collaboration avec Guillaume Giraudet et Jocelyn Faubert (Laboratoire de Psychophysique et de Perception Visuelle, Ecole d'Optométrie de l'Université de Montréal), une étude visant à évaluer si le traitement des fréquences spatiales lors de la catégorisation de scènes pouvait être influencé par la présence d'une myopie. Dans cette étude, nous avons testé si le contenu en BFS et en HFS minimum nécessaire pour catégoriser une scène différait entre les participants myopes et emmétropes. Au vu des précédentes études suggérant que les myopes s'adaptent mieux à des conditions de vision floutée (George & Rosenfield, 2004; Giraudet & Azavant, 2006; Mon-Williams et al., 1998;

[3] La myopie provient d'une anomalie du cristallin, entraînant une erreur de réfraction des rayons lumineux, qui se projettent en avant de la rétine. En conséquence, un point éloigné du champ visuel se projette comme une tâche sur la rétine, induisant une perception floutée des éléments éloignés dans le champ visuel.

Pesudovs & Brennan, 1993; Rosenfield & Abraham-Cohen, 1999), nous nous attendions à ce que les participants myopes aient besoin de moins d'information en HFS pour catégoriser les scènes que les emmétropes. Nous avons testé 20 participants emmétropes (erreur de réfraction comprise entre -1.00 et +1.00 D) et 20 participants myopes (erreur de réfraction comprise entre -2.00 et -10.00 D) dont la vision était corrigée à la normale par lentilles lors de l'expérience. Les stimuli étaient des séquences composées de 18 versions filtrées d'une même scène présentées 250 ms et dont le contenu en fréquences spatiales augmentait d'une image à l'autre dans la séquence jusqu'à aboutir à une image intacte à la fin de la séquence. Les séquences commençaient soit par une image en BFS à laquelle de l'information en HFS était progressivement ajoutée (séquences BFS→HFS), soit par une image en HFS à laquelle de l'information en BFS était progressivement ajoutée (séquences HFS→BFS). Les participants avaient pour tâche d'appuyer sur un bouton dès qu'ils pensaient avoir assez d'information pour catégoriser correctement la scène comme étant une scène d'extérieur ou d'intérieur. L'appui sur le bouton entrainaît l'arrêt de la séquence et l'apparition d'un écran leur demandant de donner la catégorie. Contrairement à nos hypothèses, nous avons observé que les participants myopes catégorisaient les séquences BFS→HFS plus tardivement que les emmétropes, suggérant qu'ils attendaient d'avoir plus d'information en HFS que les emmétropes pour catégoriser les scènes. Ces résultats suggèrent donc des différences entre les myopes et les emmétropes lors de la catégorisation de scènes filtrées qui pourraient avoir des conséquences d'un point de vue méthodologique. En effet, la majorité des études sur les mécanismes de traitement de l'information visuelle ne prennent pas en compte ce type de différences interindividuelles, et les participants testés sont sélectionnés entre autres sur la base de leur acuité visuelle qui doit être normale ou corrigée à la normale. Cependant, nos résultats, ainsi que ceux de travaux précédents, laissent à penser que, même corrigée, une myopie est susceptible d'induire des différences dans les stratégies d'analyse visuelle et que ces populations ne sont pas tout à fait équivalentes en termes de traitement de l'information. Dans l'ensemble ces résultats préliminaires offrent donc des pistes de recherche intéressantes à approfondir sur ce sujet.

On peut par ailleurs noter que l'ensemble de nos études ont été menées sur des participants droitiers. Comme nous l'avons vu dans le Chapitre 3, de nombreux travaux ont mis en évidence une spécialisation hémisphérique pour le traitement des fréquences spatiales (e.g., Peyrin et al., 2005; Peyrin, Baciu, Segebarth, & Marendaz, 2004; Peyrin, Chauvin, Chokron, & Marendaz, 2003). Or, plusieurs études suggèrent une relation entre la

latéralisation manuelle et la latéralisation fonctionnelle de certains processus, avec une variabilité plus importante chez les participants gauchers que droitiers (e.g., Cherbuin & Brinkman, 2006; Iwabuchi & Kirk, 2009). Ainsi, lorsque l'on étudie des processus cognitifs latéralisés (cas de l'analyse « coarse-to-fine » qui prédominerait dans l'hémisphère droit, Peyrin et al., 2005), et qu'on demande une réponse manuelle (main dominante) aux participants, il est généralement préférable de n'inclure que des participants droitiers dont la latéralisation des processus est mieux connue et qui présentent moins de variabilité dans leurs réponses. Il serait donc intéressant d'étudier si la latéralisation manuelle peut également influencer les stratégies d'analyse visuelles des fréquences spatiales.

L'étude du traitement des fréquences spatiales lors de la perception de scènes chez le jeune adulte sain nous permet également de fournir une base pour l'étude de ces mécanismes au cours du développement (enfance et vieillissement). En effet, plusieurs travaux suggèrent que les mécanismes de traitement des fréquences spatiales changent au cours du développement. Par exemple, Poirel et collaborateurs (Poirel et al., 2011; Poirel, Leroux, Pineau, Houdé, & Simon, 2014; Poirel, Mellet, Houdé, & Pineau, 2008) ont étudié l'évolution des stratégies d'analyses visuelles chez l'enfant de 4 à 9 ans et ont montré une précédence locale chez l'enfant de 4 ans, s'inversant progressivement autour de 6 ans en faveur d'une précédence globale à 9 ans et maintenue jusqu'à l'âge adulte. Ces changements s'accompagnaient de modifications anatomiques au niveau cortical et ont été attribués à des changements dans les stratégies visuo-spatiales d'exploration visuelles des stimuli liées aux besoins et aux contraintes environnementales de chaque âge. Bien que le traitement des fréquences spatiales n'ai pas été directement l'objet de cette étude, on peut supposer que les enfants de 4 ans privilégient une stratégie d'analyse « fine-to-coarse », avant de privilégier à partir de 9 ans une analyse « coarse-to-fine », comme nous l'avons observé chez les adultes.

Au cours du vieillissement normal, la précédence globale s'inverserait à nouveau en faveur d'une précédence locale (Lux, Marshall, Thimm, & Fink, 2008; Staudinger, Fink, Mackay, & Lux, 2010). Dans une étude récente, Musel et al. (2012) ont spécifiquement étudié l'effet du vieillissement normal sur les stratégies d'analyse des fréquences spatiales lors de la catégorisation de scènes. Ces auteurs ont présenté des séquences dynamiques imitant une analyse « coarse-to-fine » ou « fine-to-coarse » de scènes d'extérieur et d'intérieur à des participants jeunes (moins de 24 ans) et âgés (plus de 60 ans) qui avaient pour tâche de les catégoriser. Les résultats ont révélé que les participants jeunes catégorisaient les séquences « coarse-to-fine » plus rapidement que les séquences « fine-to-coarse », en accord avec les

résultats obtenus dans nos études (e.g., **Expériences 1 et 2**). Cependant les auteurs ont observé que, chez les participants âgés, la stratégie d'analyse variait en fonction de la catégorie de la scène. Les séquences « coarse-to-fine » étaient catégorisées plus rapidement que les séquences « fine-to-coarse » lorsque celles-ci étaient des scènes d'extérieur, mais cette différence tendait à s'inverser en faveur des séquences « fine-to-coarse » pour les scènes d'intérieur. Ces résultats suggèrent que les mécanismes d'analyse des fréquences spatiales évoluent également au cours du vieillissement normal. Cependant, l'effet du vieillissement normal sur le traitement des fréquences spatiales et ses bases cérébrales a très peu été étudié.

Dans l'étude IRMf réalisée en collaboration avec Stephen Ramanoël dans le cadre de sa thèse, nous avons spécifiquement examiné l'effet du vieillissement sur la catégorisation de scènes filtrées en fréquences spatiales (Ramanoël, Kauffmann, Cousin, Dojat & Peyrin, Sous presse ; cf. ANNEXE II). Cette étude nous a permis de mettre en évidence un déficit spécifique au traitement des HFS chez les participants âgés, accompagné d'une hypoactivité du cortex occipital par rapport aux participants jeunes. Nous avons également observé que la catégorisation de scènes filtrées entraînait plus généralement une réorganisation fonctionnelle chez les participants âgés qui recrutaient de façon plus importante que les participants jeunes des structures attentionnelles pariétales et frontales, ainsi que de régions temporales impliquées dans le traitement sémantique de l'information visuelle. Enfin, nous avons démontré que le déficit spécifique au traitement des HFS chez les participants âgés affectait également le traitement des fréquences spatiales au sein de la PPA. Dans l'ensemble, ces résultats suggèrent que le vieillissement normal s'accompagne d'une modification des mécanismes impliqués dans le traitement des fréquences spatiales et la catégorisation de scènes.

L'étude de l'évolution des mécanismes d'analyse visuelle au cours du développement apparaît donc très importante pour déterminer les changements qui s'opèrent à différentes périodes de la vie, de l'enfance à la vieillesse. Mieux comprendre comment l'information visuelle est traitée au niveau cérébral au cours du vieillissement normal comporte de nombreuses implications en termes de santé publique et d'amélioration de la qualité de vie des personnes âgées, mais est également devenu indispensable pour mieux cerner la nature des déficits chez des patients atteints de pathologies visuelles liées au vieillissement. En effet, les pathologies visuelles dues au vieillissement sont en constante progression depuis plusieurs années. En particulier, on peut prendre l'exemple de la dégénérescence maculaire liée à l'âge (DMLA) qui se caractérise par une perte de vision centrale due à une lésion de la macula et

qui touche en majorité des personnes de plus de 50 ans. Une étude récente montre explicitement que les patients atteints de DMLA présentent un déficit spécifique au traitement des HFS lors de la catégorisation de scènes naturelles (Musel et al., 2011). Il est donc primordial de spécifier ces mécanismes en lien avec le traitement des fréquences spatiales et leurs substrats cérébraux au cours du vieillissement normal afin de mieux appréhender ensuite l'évolution de ceux-ci dans la pathologie.

4. Conclusion

Les études comportementales et en neuroimagerie présentées dans ce travail de thèse nous ont permis de préciser et de cartographier au niveau cérébral les mécanismes neurocognitifs impliqués dans le traitement des fréquences spatiales lors de la perception de scènes. Nous avons montré, à travers deux études comportementales, que la catégorisation rapide de scènes est facilitée par le traitement de l'information grossière véhiculée par les BFS avant celui de l'information fine véhiculée par les HFS. Le système visuel privilégierait une analyse « coarse-to-fine » pour catégoriser rapidement des scènes. Au cours de deux études en IRMf, nous avons spécifiquement examiné les bases cérébrales d'une analyse « coarse-to-fine » des scènes. Nous avons mis en évidence l'implication d'un large réseau d'aires cérébrales, incluant le cortex visuel primaire, ainsi des aires supérieures frontales et temporales. Une analyse de connectivité entre ces régions nous a permis de déterminer les interactions sous-jacentes aux activations observées, suggérant en particulier une influence « top-down » du cortex frontal inférieur sur le cortex inférotemporal, mais également sur le cortex visuel primaire. Enfin, nous avons directement étudié le traitement des fréquences spatiales au sein de régions occipito-temporales sélectives aux scènes. Nos résultats ont montré que la PPA, le RSC et l'OPA répondent de façon distincte aux différentes fréquences spatiales, et qu'une stratégie d'analyse « coarse-to-fine » serait privilégiée au sein de la PPA.

Dans l'ensemble, ces résultats nous permettent de conforter et de préciser les modèles actuels de la catégorisation de scènes basés sur un traitement en fréquences spatiales de l'information visuelle. Nos résultats indiquent que les traitements effectués par les régions frontales pourraient influencer le traitement ultérieur de l'information dès le cortex occipital. Nous proposons que l'intégration des influences « top-down » des aires frontales au traitement « bottom-up » de l'information puisse être résumée en termes de codage prédictif au sein du cortex occipital. Dans ce contexte, le traitement du signal visuel se ferait sur la base

de sa correspondance avec des attentes de des prédictions, afin de réduire les coûts de traitements de faciliter la perception. Par ailleurs, nos travaux suggèrent que le cortex frontal inférieur pourrait supporter des fonctions complexes de prédiction, d'intégration, et de pilotage du traitement de l'information visuelle lors de la perception de scènes. Nos données permettent de préciser les bases cérébrales du traitement des fréquences spatiales en démontrant également le rôle central du cortex inférotemporal dans l'intégration sémantique des prédictions grossières générées par les aires frontales au traitement plus fin de l'information visuelle en provenance des régions occipitales. Comme pour le cortex occipital, cette intégration pourrait se traduire par un codage prédictif au sein de ces régions. Enfin, nos résultats ont permis de mettre en évidence l'implication différentes des régions occipito-temporales sélectives aux scènes dans le traitement des fréquences spatiales, fournissant des bases pour étudier de façon plus approfondie leur rôle spécifique au sein du réseau plus large de la perception des scènes.

Ces études fournissent des données supplémentaires sur le traitement et la catégorisation de scènes chez le jeune adulte sain et peuvent servir de base pour étudier les modifications de la perception visuelle au cours du développement normal et pathologique. Plus spécifiquement, la compréhension des changements qui s'opèrent au cours du vieillissement normal permettrait de mieux cerner la nature des déficits et la façon dont l'environnement est perçu dans les pathologies visuelles liées à l'âge.

RÉFÉRENCES BIBLIOGRAPHIQUES

Abrams, J., Barbot, A., & Carrasco, M. (2010). Voluntary attention increases perceived spatial frequency. *Attention, perception & psychophysics*, *72*, 1510–21.

Aggleton, J. P., & Vann, S. D. (2004). Testing the importance of the retrosplenial navigation system: lesion size but not strain matters: a reply to Harker and Whishaw. *Neuroscience and biobehavioral reviews*, *28*, 525–31.

Allen, E. A, & Freeman, R. D. (2006). Dynamic spatial processing originates in early visual pathways. *The Journal of neuroscience*, *26*, 11763–74.

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral cortex*, *17*, 1493–503.

Aminoff, E., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in cognitive sciences*, *17*, 379–390.

Aminoff, E., & Tarr, M. (2014). Anterior to posterior parahippocampal organization of scene information. In *Journal of Vision - VSS meeting abstract* (Vol. 14, pp. 1079–1079).

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *NeuroImage*, *49*, 703–11.

Arcaro, M. J., McMains, S. A, Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *The Journal of neuroscience*, *29*, 10638–52.

Auger, S. D., Mullally, S. L., & Maguire, E. a. (2012). Retrosplenial cortex codes for permanent landmarks. *PloS one*, *7*, e43620.

Awasthi, B., Sowman, P. F., Friedman, J., & Williams, M. a. (2013). Distinct spatial scale sensitivities for early categorization of faces and places: neuromagnetic and behavioral findings. *Frontiers in human neuroscience*, *7*, 91.

Azzopardi, P., Jones, K. E., & Cowey, A. (1999). Uneven mapping of magnocellular and parvocellular projections from the lateral geniculate nucleus to the striate cortex in the macaque monkey. *Vision Research*, *39*, 2179–2189.

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*, 1459–1469.

Badcock, J. C., Whitworth, F. A., Badcock, D. R., & Lovegrove, W. J. (1990). Low-frequency filtering and the processing of local - global stimuli. *Perception*, *19*, 617–629.

Badre, D., Poldrack, R. a, Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*, 907–18.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*, 2883–2901.

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. *NeuroImage*, *75*, 228–37.

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

Bar, M. (2004). Visual objects in context. *Nature reviews. Neuroscience*, *5*, 617–29.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in cognitive sciences*, *11*, 280–9.

Bar, M., & Aminoff, E. (2003). Cortical Analysis of Visual Context. *Neuron*, *38*, 347–358.

Bar, M., Aminoff, E., & Ishai, A. (2008). Famous faces activate contextual associations in the parahippocampal cortex. *Cerebral cortex*, *18*, 1233–8.

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *The Journal of neuroscience*, *28*, 8539–44.

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*, 449–454.

Barbas, H. (1995). Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex. *Neuroscience and biobehavioral reviews*, *19*, 499–510.

Barbas, H. (2000). Connections underlying the synthesis of cognition , memory , and emotion in primate prefrontal cortices. *Brain Research Bulletin*, *52*, 319–330.

Baumgartner, F., Hanke, M., Geringswald, F., Zinke, W., Speck, O., & Pollmann, S. (2013). Evidence for feature binding in the superior parietal lobule. *NeuroImage*, *68*, 173–80.

Behrmann, M., Geng, J. J., & Shomstein, S. (2004). Parietal cortex and attention. *Current opinion in neurobiology*, *14*, 212–7.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289–300.

Berman, D., & Walther, D. B. (2014). Differential Selectivity for Spatial Frequencies in Anterior and Posterior PPA. In *Journal of Vision - VSS meeting abstract* (Vol. 14, pp. 1081–1081).

Bestmann, S., Ruff, C. C., Blankenburg, F., Weiskopf, N., Driver, J., & Rothwell, J. C. (2008). Mapping causal interregional influences with concurrent TMS-fMRI. *Experimental brain research*, *191*, 383–402.

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of Optical Society of America*, *19*, 1096–1106.

Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision*, *9*, 1–19.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of physiology*, *203*, 237–260.

Boynton, G. M. (2005). Contras gain in the brain. *Neuron*, *47*, 476–477.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of neuroscience*, *16*, 4207–21.

Bredfeldt, C. E., & Ringach, D. L. (2002). Dynamics of Spatial Frequency Tuning in Macaque V1. *The journal of neuroscience*, *22*, 1976–1984.

Brefczynski, J. A., & Deyoe, E. A. (1999). A physiological correlate of the “spotlight ” of visual attention. *Nature Neuroscience*, *2*, 370-374.

Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research*, *15*, 1411–1412.

Breitmeyer, B. G. (2014). Contributions of magno- and parvocellular channels to conscious and non-conscious vision. *Philosophical transactions of the Royal Society of London.*, *369*, 20130213.

Bullier, J. (1998). Architecture fonctionnelle du système visuel. *Vision: aspects perceptifs et cognitifs*, 11–42.

Bullier, J. (2001). Integrated model of visual processing. *Brain research*, *36*, 96–107.

Bullier, J., Hupé, J., James, A. C., & Girard, P. (1996). Functional interactions between areas Vl and V2 in the monkey. *Journal of neurophysiology*, *90*, 217–220.

Bullier, J., Hupé, J.-M., James, A. C., & Girard, P. (2001). The role of feedback connections in shaping the response of visual cortical neurons. *Progress in Brain Research*, *134*, 193–204.

Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, *26*, 157–170.

Cant, J. S., & Goodale, M. a. (2011). Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. *The Journal of neuroscience*, *31*, 8248–58.

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *The Journal of neuroscience*, *32*, 7685–700.

Cant, J. S., & Xu, Y. (2014). The Impact of Density and Ratio on Object-Ensemble Representation in Anterior-Medial Ventral Visual Cortex. *Cerebral cortex*, *12*, 585–585.

Capotosto, P., Tosoni, A., Spadone, S., Sestieri, C., Perrucci, M. G., Romani, G. L., … Corbetta, M. (2013). Anatomical segregation of visual selection mechanisms in human parietal cortex. *Journal of neuroscience*, *33*, 6225–9.

Carlsson, K., Petrovic, P., Skare, S., Petersson, K. M., & Ingvar, M. (2000). Tickling expectations: neural processing in anticipation of a sensory stimulus. *Journal of cognitive neuroscience*, *12*, 691–703.

Catani, M., Jones, D. K., Donato, R., & Ffytche, D. H. (2003). Occipito-temporal connections in the human brain. *Brain*, *126*, 2093–107.

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, *44*, 1105–32.

Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J., & Reinoso-Suárez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cerebral cortex*, *10*, 220–42.

Chaumon, M., Kveraga, K., Barrett, L. F., & Bar, M. (2014). Visual Predictions in the Orbitofrontal Cortex Rely on Associative Content. *Cerebral cortex*, 1–9.

Cherbuin, N., & Brinkman, C. (2006). Hemispheric interactions are different in left-handed individuals. *Neuropsychology*, *20*, 700–707.

Chong, T. T.-J., Williams, M. A, Cunnington, R., & Mattingley, J. B. (2008). Selective attention modulates inferior frontal gyrus activity during action observation. *NeuroImage*, *40*, 298–307.

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*, 354–64.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews. Neuroscience*, *3*, 201–15.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of vision*, *10*, 16.1–17.

Curcio, C., & Allen, K. (1990). Topography of ganglion cells in human retina. *The journal of comparative neurology*, *300*, 5–25.

Curcio, C., Sloan, K., Kalina, R., & Hendrickson, A. (1990). Human photoreceptor topography. *The journal of comparative neurology*, *292*, 497–523.

Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current biology* : CB, *12*, R476–7; author reply R478.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual review of neuroscience*, *13*, 89–109.

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

De Cesarei, A., & Codispoti, M. (2011). Scene identification and emotional response: which spatial frequencies are critical? *The Journal of neuroscience*, *31*, 17052–7.

De Cesarei, A., & Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychonomic bulletin & review*, *18*, 840–7.

De Cesarei, A., Mastria, S., & Codispoti, M. (2013). Early spatial frequency processing of natural images: an ERP study. *PloS one*, *8*, e65103.

De Haan, E. H. F., & Cowey, A. (2011). On the usefulness of "what" and "where" pathways in vision. *Trends in cognitive sciences*, *15*, 460–6.

De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559.

De Valois, R., & De Valois, K. (1988). Spatial Vision. *Annual review of psychology*, *31*, 309–341.

De Valois, R., Yund, W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 531–544.

Dean, A. F. (1981). The relationship between response amplitude and contrast for cat striate cortical neurones. *Journal of physiology*, *318*, 413–427.

Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *Journal of physiology*, *357*, 219–240.

Desseilles, M., Schwartz, S., Dang-Vu, T. T., Sterpenich, V., Ansseau, M., Maquet, P., & Phillips, C. (2011). Depression alters "top-down" visual attention: a dynamic causal modeling comparison between depressed and healthy subjects. *NeuroImage*, *54*, 1662–8.

de-Wit, L., Machilsen, B., & Putzeys, T. (2010). Predictive coding and the neural response to predictable stimuli. *The Journal of neuroscience*, *30*, 8702–8703.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of neuroscience*, *33*, 1331–1336.

Dima, D., Stephan, K. E., Roiser, J. P., Friston, K. J., & Frangou, S. (2011). Effective connectivity during processing of facial affect: evidence for multiple parallel pathways. *The Journal of neuroscience*, *31*, 14378–85.

DiQuattro, N. E., & Geng, J. J. (2011). Contextual knowledge configures attentional control networks. *The Journal of neuroscience*, *31*, 18026–35.

Dos Santos, N. a, Andrade, S. M., & Fernandez Calvo, B. (2013). Detection of spatial frequency in brain-damaged patients: influence of hemispheric asymmetries and hemineglect. *Frontiers in human neuroscience*, *7*, 92.

Einevoll, G. T., Jurkus, P., & Heggelund, P. (2011). Coarse-to-fine changes of receptive fields in lateral geniculate nucleus have a transient and a sustained component that depend on distinct mechanisms. *PloS one*, *6*, e24523.

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.

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, *187*, 517–552.

Enroth-Cugell, C., & Robson, J. G. (1984). Functional characteristics and diversity of cat retinal ganglion cells. Basic characteristics and quantitative description. *Investigative ophthalmology & visual science*, *25*, 250–67.

Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*, 954–978.

Epstein, R. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*, 388–396.

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-Specific Scene Representations in Human Parahippocampal Cortex. *Neuron*, *37*, 865–876.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The Parahippocampal Place Area : Recognition , Navigation , or Encoding ? *Neuron*, *23*, 115–125.

Epstein, R., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cerebral cortex*, *17*, 1680–93.

Epstein, R., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of neurophysiology*, *97*, 3670–83.

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

Epstein, R., & Ward, E. J. (2010). How reliable are visual context effects in the parahippocampal place area? *Cerebral cortex*, *20*, 294–303.

Everson, R. M., Prashanth, A. K., Gabbay, M., Knight, B. W., Sirovich, L., & Kaplan, E. (1998). Representation of spatial frequency and orientation in the. *Proceedings of the National Academy of Sciences*, *95*, 8334–8338.

Farell, B., Li, S., & Mckee, S. P. (2004). Coarse scales , fine scales , and their interactions in stereo vision. *Journal of Vision*, *4*, 488–499.

Felleman, D. J., & Essen, D. C. Van. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1047–3211.

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.*, *4*, 2379–94.

Field, D. J. (1989). What the statistics of natural images tell us about visual coding. In *OE/LASE'89* (Vol. 1, pp. 269–276). Los Angeles. CA: International Society for Optics and Photonics.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S., & Dolan. (1996). Where in the brain does visual attention select the forest and the trees? *Nature*, *382*, 626–628.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1997). Neural mechanisms involved in the processing of global and local aspects of hierarchically organized visual stimuli. *Brain*, *120*, 1779–1791.

Fink, G. R., Marshall, J. C., Halligan, P. W., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1997). Hemispheric specialization for global and local processing : the effect of stimulus category. *Proceedings of the Royal Society*, *264*, 487-494.

Fink, G. R., Marshall, J., Halligan, P., & Dolan, R. (2000). Neuronal activity in early visual areas during global and local processing: a comment. *Journal of cognitive neuroscience*, *12*, 355–356.

Fintzi, A. R., & Mahon, B. Z. (2013). A bimodal tuning curve for spatial frequency across left and right human orbital frontal cortex during object recognition. *Cerebral cortex*, *24*, 1311–8.

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory Insights from functional neuroimaging. *Brain*, *124*, 849–881.

Foster, K. H., Gaska, J., Nagler, M., & Pollen, D. A. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *Journal of physiology*, *365*, 331–363.

Frazor, R. a, Albrecht, D. G., Geisler, W. S., & Crane, A. M. (2004). Visual cortex neurons of monkeys and cats: temporal dynamics of the spatial frequency response function. *Journal of neurophysiology*, *91*, 2607–27.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*, 312–6.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of neuroscience*, *23*, 5235–5246.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *360*, 815–836.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*, 218–229.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1995). Statistical Parametric Maps in Functional Imaging : A General Linear Approach. *Human brain mapping*, *2*, 189–210.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, a M. (1999). Stochastic designs in event-related fMRI. *NeuroImage*, *10*, 607–19.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary. *Proceedings of the National Academy of Sciences*, *96*, 3314–3319.

Garavan, H., Ross, T. J., Kaufman, J., & Stein, E. A. (2003). A midline dissociation between error-processing and response-conflict monitoring. *NeuroImage*, *20*, 1132–9.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable Executive Functions in the Dynamic Control of Behavior: Inhibition, Error Detection, and Correction. *NeuroImage*, *17*, 1820–1829.

Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the Macaque. *The journal of comparative neurology*, *201*, 519–539.

Gattass, R., Sousa, a P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience*, *8*, 1831–1845.

Gegenfurtner, K. R., Kiper, D. C., & Levitt, J. B. (1997). Functional properties of neurons in macaque area V3. *Journal of neurophysiology*, *77*, 1906–1923.

George, S., & Rosenfield, M. (2004). Blur adaptation and myopia. *Optometry and vision science*, *81*, 543–7.

Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience*, *115*, 1261–79.

Ghashghaei, H. T., Hilgetag, C. C., & Barbas, H. (2007). Sequence of information processing for emotion based on the anatomic dialogue between prefrontal cortex and amygdala. *NeuroImage*, *34*, 905–923.

Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance.* (pp. 1–41). Oxford, England: John Wiley & Sons.

Giraudet, G., & Azavant, L. (2006). Object localization in blurred and jumbled scenes: differences between myopic and emmetropic observers. *Investigative ophthalmology & visual science*, *47*, 4146–51.

Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *NeuroImage*, *19*, 200–207.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*, 416–29.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral cortex*, *21*, 467–76.

Gold, B. T., Balota, D. a, Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *The Journal of neuroscience*, *26*, 6523–6532.

Gollisch, T., & Meister, M. (2010). Eye smarter than scientists believed: Neural computations in circuits of the retina. *Neuron*, *65*, 150–164.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neuroscience*, *15*, 20–25.

Goodyear, B. G., & Menon, R. S. (1998). Effect of luminance contrast on BOLD fMRI response in human primary visual areas. *Journal of neurophysiology*, *79*, 2204–7.

Grabowska, A. & Nowicka, A. (1996). Visual-spatial-frequency model of cerebral asymmetry: a critical survey of behavioral and electrophysiological studies. *Psychological Bulletin*, *120*, 434–449.

Greene, M. R., & Oliva, A. (2009). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychological science*, *20*, 464–472.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*, 159–166.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research*, *41*, 1409–22.

Guérin-dugué, A., & Oliva, A. (2000). Classifcation of scene photographs from local orientations features. *Pattern recognition letters*, *21*, 1135–1140.

Guyader, N., Chauvin, A., Peyrin, C., Hérault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, *327*, 313–318.

Guyader, N., & Hérault, J. (2001). Représentation espace-fréquence pour la catégorisation d'images. In *18° Colloque sur le traitement du signal et des images, FRA, 2001. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images.*

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., & Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *NeuroImage*, *50*, 1313–9.

Harker, K. T., & Whishaw, I. Q. (2004). A reaffirmation of the retrosplenial contribution to rodent navigation: reviewing the influences of lesion, strain, and task. *Neuroscience and biobehavioral reviews*, *28*, 485–496.

Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision research*, *18*, 1579–86.

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, *37*, 1027–1041.

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–30.

Haxby, J. V, Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., … Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, *88*, 1621–1625.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in neurobiology*, *84*, 405–39.

Henderson, J. M., Larson, C. L., & Zhu, D. C. (2007). Cortical activation to indoor versus outdoor scenes: an fMRI study. *Experimental brain research.*, *179*, 75–84.

Henderson, J. M., Larson, C. L., & Zhu, D. C. (2008). Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: an fMRI study. *Brain and cognition*, *66*, 40–9.

Henderson, J. M., Zhu, D. C., & Larson, C. L. (2011). Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: An fMRI study. *Visual Cognition*, *19*, 910–927.

Henriksson, L., Nurminen, L., & Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *Journal of Vision*, *8*, 1–13.

Hérault, J., Oliva, A., & Guérin-dugué, A. (1997). Scene categorisation by curvilinear component analysis of low frequency spectra. In *Proceedings of European Symposium on Artificial Neural Networks* (pp. 91–96).

Hicks, T., Lee, B., & Vidyasagar, T. (1983). The response of cells in macaque lateral geniculate nucleus to sinusoidal gratings. *Journal of physiology*, *337*, 183–200.

Holmes, G. (1918). Disturbance of visual orientation. *The British journal of ophtalmology*, *2*, 449–468.

Horton, J. C., & Hoyt, W. (1991). The Representation of the Visual Field in Human Striate Cortex. *Archives of Ophthalmology*, *109*, 816.

Hubel, D. H. (1994). L'oeil, le cerveau et la vision. *Pour la Science, Belin*.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of physiology*, *160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of physiology*, *195*, 215–243.

Hubel, D. H., & Wiesel, T. N. (1977). Functional Architecture of Macaque Monkey Visual Cortex. *Proceedings of the Royal Society B: Biological Sciences*, *198*, 1–59.

Hughes, H. C. (1986). Asymmetric interference between components of suprathreshold compound gratings. *Perception & psychophysics*, *40*, 241–50.

Hughes, H. C., Fendrich, R., & Reuter-Lorenz, P. A. (1990). Global versus local processing in the absence of low spatial frequencies. *Journal of cognitive neuroscience*, *2*, 272–82.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of cognitive neuroscience*, *8*, 197–230.

Hupé, J., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., Bullier, J. (2001). Feedback Connections Act on the Early Part of the Responses in Monkey Visual Cortex. *Journal of physiology*, *85*, 134–145.

Hupé, J., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*, 784–787.

Iidaka, T., Yamashita, K., Kashikura, K., & Yonekura, Y. (2004). Spatial frequency of visual image modulates neural responses in the temporo-occipital lobe. An investigation with event-related fMRI. *Cognitive Brain Research*, *18*, 196–204.

Ishai, A., Ungerleider, L., Martrin, A., Schouten, J., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*, 9379–9384.

Issa, N. P., Trepel, C., & Stryker, M. P. (2000). Spatial frequency maps in cat visual cortex. *The Journal of neuroscience*, *20*, 8504–14.

Iwabuchi, S. J., & Kirk, I. J. (2009). Atypical interhemispheric communication in left-handed individuals. *Neuroreport*, *20*, 166–169.

Johnson, J. S., & Olshausen, B. a. (2003). Timecourse of neural signatures of object recognition. *Journal of vision*, *3*, 499–512.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of neuroscience*, *17*, 4302–11.

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., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial Frequency Processing in Scene-Selective Cortical Regions. *NeuroImage*, *112*, 86–95.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 1–14.

Kihara, K., & Takeda, Y. (2010). Time course of the integration of spatial frequency-based information in natural scenes. *Vision research*, *50*, 2158–62.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, *112*, 24–38.

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

Kitterle, F. L., Christman, S., & Hellige, J. B. (1990). Hemispheric differences are found in the identification, but not the detection, of low versus high spatial frequencies. *Perception & psychophysics*, *48*, 297–306.

Kitterle, F. L., Hellige, J. B., & Christman, S. (1992). Visual hemispheric asymmetries depend on which spatial frequencies are task relevant. *Brain and cognition*, *20*, 308–314.

Kitterle, F. L., & Selig, L. M. (1991). Visual field effects in the discrimination of sine-wave gratings. *Perception & psychophysics*, *50*, 15–18.

Kohler, W. (1964). *Psychologie de la forme: introduction à de nouveaux concepts en psychologie* (Gallimard.).

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *The Journal of neuroscience*, *31*, 2488–92.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, *74*, 1114–24.

Korkmaz Hacialihafiz, D., & Bartels, A. (2015). Motion responses in scene selective regions. *NeuroImage*, *118*, 438–444.

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *The Journal of neuroscience*, *31*, 7322–33.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature reviews. Neuroscience*, *12*, 217–30.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in cognitive sciences*, *17*, 26–49.

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in neurobiology*, *72*, 341–72.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of neuroscience*, *27*, 13232–40.

Lagasse, L. L. (1993). Effects of good form and spatial frequency on global precedence. *Perception & Psychophysics*, *53*, 89–105.

Lamb, M. R. (1996). Spatial Frequency and Interference Between Global and Local Levels of Structure. *Visual Cognition*, *3*, 193–220.

Lamb, M. R., Robertson, L. C., & Knight, R. (1990). Component mechanisms underlying the processing of hierarchically organized patterns: inferences from patients with unilateral cortical lesions. *Journal of experimental psychology. Learning, memory, and cognition*, *16*, 471–483.

Lamb, M. R., & Yund, W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception & psychophysics*, *54*, 773–784.

Lamme, V. A., Super, H., & Speckreijse, H. (1998). Horizontal and Feedback Processing in the Visual Cortex. *Current Opinion in Neurobiology*, *8*, 529–535.

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of physiology, Paris*, *97*, 121–39.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision research*, *38*, 2429–54.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral cortex*, *11*, 287–97.

Li, W., Piëch, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature neuroscience*, *7*, 651–657.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic bulletin & review*, *12*, 43–65.

Lupp, U., Hauske, G., & Wolf, W. (1976). Perceptual latencies to sinusoidal gratings. *Vision Research*, *16*, 969–972.

Lux, S., Marshall, J. C., Thimm, M., & Fink, G. R. (2008). Differential processing of hierarchical visual stimuli in young and older healthy adults: implications for pathology. *Cortex*, *44*, 21–8.

Maffei, L., & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Research*, *13*, 1255–1267.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. a, … Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*, 8135–9.

Marendaz, C., Rousset, S., & Charnallet, A. (2003). La reconnaissance des scènes, des objets et des visages. In A. Delorme & M. Fluckiger (Eds.), *Perception et Réalité* (Gaëtan Mor.). Montréal.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Fransisco: Freeman.

Martínez, A, Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., … Hillyard, S. a. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature neuroscience*, *2*, 364–9.

Masland, R. H. (2012). The Neuronal Organization of the Retina. *Neuron*, *76*, 266–280.

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, 1–14.

Maunsell, J., & Van Essen, D. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. Selectivity for stimulus direction, speed, and orientation. *Journal of neurophysiology*, *49*, 1127–1147.

Maunsell, J., & Van Essen, D. (1983b). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of neuroscience*, *3*, 2563–2586.

Mazer, J. A, Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 1645–50.

Mechelli, A., Price, C. J., Noppeney, U., & Friston, K. J. (2003). A dynamic causal modeling study on category effects: bottom-up or top-down mediation? *Journal of cognitive neuroscience*, *15*, 925–34.

Mellet, E., Briscogne, S., Tzourio-Mazoyer, N., Ghaëm, O., Petit, L., Zago, L., … Denis, M. (2000). Neural correlates of topographic mental exploration: the impact of route versus survey perspective learning. *NeuroImage*, *12*, 588–600.

Merigan, W., & Maunsell, J. (1993). How parallel are the primate visual pathways? *Annual review of neuroscience*, *16*, 369–402.

Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling. *Brain and cognition*, *57*, 151–7.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*, 167–202.

Miyashita, Y., & Hayashi, T. (2000). Neural representation of visual objects: encoding and top-down activation. *Current Opinion in Neurobiology*, *10*, 187–194.

Mon-Williams, M., Tresilian, J. R., Strang, N. C., Kochhar, P., & Wann, J. P. (1998). Improving vision: neural compensation for optical defocus. *Proceedings of the Royal Society of London*, *265*, 71–7.

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

Movshon, J. A, Thompson, I. D., & Tolhurst, D. J. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *The Journal of physiology*, *283*, 101–120.

Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental brain research.*, *227*, 367–77.

Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Bas, J. Le, & Peyrin, C. (2013). Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. *Journal of cognitive neuroscience*, *25*, 1315–1331.

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PloS one*, *7*, e38493.

Musel, B., Hera, R., Chokron, S., Alleysson, D., Chiquet, C., Romanet, J.-P., … Peyrin, C. (2011). Residual abilities in age-related macular degeneration to process spatial frequencies during natural scene categorization. *Visual neuroscience*, *28*, 529–41.

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of cognitive neuroscience*, *26*, 2287–2297.

Nasr, S., Devaney, K. J., & Tootell, R. B. H. (2013). Spatial encoding and underlying circuitry in scene-selective cortex. *NeuroImage*, *83*, 892–900.

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-selective cortical regions in human and nonhuman primates. *The Journal of neuroscience*, *31*, 13771–85.

Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual cortex. *The Journal of neuroscience*, *32*, 14921–6.

Navon, D. (1977). Forest Before Trees: The Precedence of Global Features in Visual Perception. *Cognitive psychology*, 353–383.

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cognitive, affective & behavioral neuroscience*, *7*, 1–17.

Nelson, J. K., Reuter-Lorenz, P. a, Persson, J., Sylvester, C.-Y. C., & Jonides, J. (2009). Mapping interference resolution across task domains: a shared control process in left inferior frontal gyrus. *Brain research*, *1256*, 92–100.

Nirody, J. A. (2013). Development of spatial coarse-to-fine processing in the visual pathway. *Journal of computational neuroscience*, *36*, 401-414.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430.

Nowak, L., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual neuroscience*, *12*, 371–384.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of cognitive neuroscience*, *12*, 1013–1023.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*, 604–609.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges ? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive psychology*, *107*, 72–107.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, *42*, 145–175.

Oliva, A., Torralba, A. B., Guérin-dugué, A., & Hérault, J. (1999). Global Semantic Classifcation of Scenes using Power Spectrum Templates. In *Proceedings of the 1999 international conference on Challenge of Image Retrieval* (pp. 1–12).

Olman, C. A., Ugurbil, K., Schrater, P., & Kersten, D. (2004). BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Research*, *44*, 669–683.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Nyt Nordisk Forlag*.

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*, 1108–19.

Özgen, E., Sowden, P., Schyns, P., & Daoutis, C. (2005). Top-down attentional modulation of spatial frequency processing in scene perception. *Visual Cognition*, *12*, 925–937.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *The Journal of neuroscience*, *31*, 1333–40.

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage*, *47*, 1747–1756.

Park, S., Chun, M. M., & Johnson, M. K. (2010). Refreshing and integrating visual scenes in scene-selective cortex. *Journal of cognitive neuroscience*, *22*, 2813–2822.

Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: boundary extension in human scene-selective visual cortex. *Neuron*, *54*, 335–42.

Park, S., Konkle, T., & Oliva, A. (2014). Parametric Coding of the Size and Clutter of Natural Scenes in the Human Brain. *Cerebral cortex*, *3*, 1-14.

Parker, D. M. (1980). Simple reaction times to the onset, offset, and contrast reversal of sinusoidal grating stimuli. *Perception & psychophysics*, *28*, 365–368.

Parker, D. M., & Dutch, S. (1987). Perceptual latency and spatial frequency. *Vision Research*, *27*, 1279–1283.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–60.

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*, 1448–1466.

Parker, D. M., Lishman, J. R., & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception*, *26*, 1169–1180.

Parker, D. M., & Salzen, E. A. (1977). Latency changes in the human visual evoked response to sinusoidal gratings. *Vision Research*, *17*, 1201–1204.

Parker Jones, O., Seghier, M. L., Kawabata Duncan, K. J., Leff, A. P., Green, D. W., & Price, C. J. (2013). Auditory-motor interactions for the production of native and non-native speech. *The Journal of neuroscience*, *33*, 2376–87.

Penny, W. D. (2012). Comparing dynamic causal models using AIC, BIC and free energy. *NeuroImage*, *59*, 319–30.

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004). Comparing dynamic causal models. *NeuroImage*, *22*, 1157–72.

Pesudovs, K., & Brennan, N. a. (1993). Decreased uncorrected vision after a period of distance fixation with spectacle wear. *Optometry and vision science*, *70*, 528–531.

Petrides, M., Alivisatos, B., & Frey, S. (2002). Differential activation of the human orbital, mid-ventrolateral, and mid-dorsolateral prefrontal cortex during the processing of visual stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 5649–5654.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *NeuroImage*, *23*, 698–707.

Peyrin, C., Chauvin, A., Chokron, S., & Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain and Cognition*, *53*, 278–282.

Peyrin, C., Chokron, S., Guyader, N., Gout, O., Moret, J., & Marendaz, C. (2006). Neural correlates of spatial frequency processing: A neuropsychological approach. *Brain research*, *1073-1074*, 1–10.

Peyrin, C., Mermillod, M., Chokron, S., & Marendaz, C. (2006). Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. *Brain and cognition*, *62*, 214–20.

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*, 2768–80.

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., & Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *NeuroImage*, *28*, 464–73.

Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative ophthalmology*, *11*, 368–77.

Poirel, N., Leroux, E., Pineau, A., Houdé, O., & Simon, G. (2014). Changes in cortical thickness in 6-year-old children open their mind to a global vision of the world. *BioMed Research International*, *2014*. 1-7.

Poirel, N., Mellet, E., Houdé, O., & Pineau, A. (2008). First came the trees, then the forest: developmental changes during childhood in the processing of visual local-global patterns according to the meaningfulness of the stimuli. *Developmental psychology*, *44*, 245–253.

Poirel, N., Simon, G., Cassotti, M., Leroux, G., Perchey, G., Lanoë, C., … Houdé, O. (2011). The shift from local to global visual processing in 6-year-old children is associated with grey matter loss. *PLoS ONE*, *6*, 1–5.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*, 25–42.

Pourtois, G., Rauss, K. S., Vuilleumier, P., & Schwartz, S. (2008). Effects of perceptual learning on primary visual cortex activity in humans. *Vision Research*, *48*, 55–62.

Powell, H. W. R., Guye, M., Parker, G. J. M., Symms, M. R., Boulby, P., Koepp, M. J., … Duncan, J. S. (2004). Noninvasive in vivo demonstration of the connections of the human parahippocampal gyrus. *NeuroImage*, *22*, 740–7.

Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, *15*, 246–253.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS biology*, *9*, e1000608.

Ramanoël, S., Kauffmann, L., Cousin, E., Dojat, M., & Peyrin, C. (in press). Age-related differences in spatial frequency processing during scene categorization. *PloS one*.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*, 79–87.

Rauss, K., & Pourtois, G. (2013). What is bottom-up and what is top-down in predictive coding. *Frontiers in Psychology*, *4*, 1–8.

Rauss, K., Schwartz, S., & Pourtois, G. (2011). Top-down effects on early visual processing in humans: a predictive coding framework. *Neuroscience and biobehavioral reviews*, *35*, 1237–53.

Reinvang, I., Magnussen, S., & Greenlee, M. W. (2002). Hemispheric asymmetry in visual discrimination and memory: ERP evidence for the spatial frequency hypothesis. *Experimental brain research*, *144*, 483–95.

Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-Level Image Properties of Visual Objects Predict Patterns of Neural Response across Category-Selective Regions of the Ventral Visual Pathway. *The Journal of neuroscience*, *34*, 8837–44.

Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S. C., & Price, C. J. (2011). Multiple routes from occipital to temporal cortices during reading. *The Journal of neuroscience*, *31*, 8239–47.

Rieger, J. W., Gegenfurtner, K. R., Koechy, N., Heinze, H., & Grueschow, M. (2013). BOLD responses in human V1 to local structure in natural scenes : Implications for theories of visual coding. *Journal of Vision*, *13*, 1–15.

Robertson, L. C., & Lamb, M. R. (1991). Neuropsychological contributions to theories of part/whole organization. *Cognitive Psychology*, *23*, 299–330.

Rolls, E. T., & Baylis, L. (1994). Gustatory , Olfactory , Orbitofrontal Cortex and Visual Convergence within the Primate. *The Journal of neuroscience*, *14*, 5437–5452.

Rolls, E. T., Browning, A., Inoue, K., & Hernadi, I. (2005). Novel visual stimuli activate a population of neurons in the primate orbitofrontal cortex. *Neurobiology of learning and memory*, *84*, 111–23.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurphysiology of visual masking. *Proceedings of the Royal Society*, *257*, 9–15.

Rosenfield, M., & Abraham-Cohen, J. A. (1999). Blur sensitivity in myopes. *Optometry and vision science*, *76*, 303–307.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*, 1–22.

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*, 1–20.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., … Driver, J. (2006). Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. *Current biology*, *16*, 1479–88.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Weiskopf, N., & Driver, J. (2009). Hemispheric differences in frontal and parietal influences on human occipital cortex: direct confirmation with concurrent TMS-fMRI. *Journal of cognitive neuroscience*, *21*, 1146–61.

Sasaki, Y., Hadjikhani, N., Fischl, B., Liu, a K., Marrett, S., Dale, a M., … Marret, S. (2001). Local and global attention are mapped retinotopically in human occipital cortex. *Proceedings of the National Academy of Sciences*, *98*, 2077–82.

Saygin, A. P., & Sereno, M. I. (2008). Retinotopy and attention in human occipital, temporal, parietal, and frontal cortex. *Cerebral cortex*, *18*, 2158–68.

Schettino, A., Loeys, T., Bossi, M., & Pourtois, G. (2012). Valence-specific modulation in the accumulation of perceptual evidence prior to visual scene recognition. *PloS one*, *7*, e38064.

Schettino, A., Loeys, T., Delplanque, S., & Pourtois, G. (2011). Brain dynamics of upstream perceptual processes leading to visual object recognition: a high density ERP topographic mapping study. *NeuroImage*, *55*, 1227–41.

Schettino, A., Loeys, T., & Pourtois, G. (2013). Multiple synergistic effects of emotion and memory on proactive processes leading to scene recognition. *NeuroImage*, *81C*, 81–95.

Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences*, *99*, 17137–17142.

Schyns, P. G., & Oliva, A. (1994). Evidence for time- and spatial-scale-dependent scene recognition. *Psychological science*, *5*, 195–201.

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*, 243–65.

Seghier, M. L., Josse, G., Leff, A. P., & Price, C. J. (2011). Lateralization is predicted by reduced coupling from the left to right prefrontal cortex during semantic decisions on written words. *Cerebral cortex*, *21*, 1519–31.

Seghier, M. L., Zeidman, P., Neufeld, N. H., Leff, A. P., & Price, C. J. (2010). Identifying abnormal connectivity in patients using dynamic causal modeling of FMRI responses. *Frontiers in systems neuroscience*, *4*, 1–14.

Serences, J. T. (2008). Value-based modulations in human visual cortex. *Neuron*, *60*, 1169–81.

Sergent, J. (1982). The cerebral balance of power: confrontation or cooperation? *Journal of experimental psychology. Human perception and performance*, *8*, 253–272.

Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision research*, *42*, 2547–54.

Shapley, R., & Enroth-Cugell, C. (1984). Visual Adaptation and Retinal Gain Controls. *Progress in retinal research*, *3*, 263–346.

Shapley, R., Kaplan, E., & Soodak, R. (1981). Spatial summation and contrast sensitivity of X and Y cells in the lateral geniculate nucleus of the macaque. *Nature*, *292*, 543–545.

Shapley, R., & Lam, D. (Eds). (1993). *Contrast sensitivity*. (Vol.5). MIT Press.

Shapley, R., & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual review of neuroscience*, *8*, 547-583

Shomstein, S. (2012). Cognitive functions of the posterior parietal cortex: top-down and bottom-up attentional control. *Frontiers in integrative neuroscience*, *6*, 38.

Shulman, G. L., Sullivan, M. A., Gish, K., & Sakoda, W. J. (1986). The role of spatial-frequency channels in the perception of local and global structure. *Perception*, *15*, 259–273.

Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to spatial location. *Perception*, *16*, 103–111.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural Correlates of Sustained Spatial Attention in Human Early Visual Cortex. *Journal of neurophysiology*, *97*, 229–237.

Singh, K. D., Smith, a T., & Greenlee, M. W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *NeuroImage*, *12*, 550–64.

Skottun, B., De Valois, R., Grosof, D., Movshon, J. A., Albrecht, D. G., & A, B. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision research*, *31*, 1079–1086.

Sowden, P. T., Özgen, E., Schyns, P. G., & Daoutis, C. (2003). Expectancy effects on spatial frequency processing. *Vision Research*, *43*, 2759–2772.

Staudinger, M. R., Fink, G. R., Mackay, C. E., & Lux, S. (2010). Gestalt perception and the decline of global precedence in older subjects. *Cortex*, *47*, 854–62.

Stephan, K. E., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, *49*, 3099–109.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*, 1004–17.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, *13*, 403–9.

Sung, Y.-W., Kamba, M., & Ogawa, S. (2008). Building-specific categorical processing in the retrosplenial cortex. *Brain research*, *1234*, 87–93.

Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research*, *18*, 1393–1399.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, *262*, 685–688.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual review of neuroscience*, *19*, 109–39.

Tanaka, K., Saito, H.-A., Fukada, Y., & Moriya, M. (1991). Coding Visual Images of Objects in the Inferotemporal Cortex of the Macaque Monkey. *Journal of neurophysiology*, *66*, 170–189.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Thorpe, S. J., Rolls, E. T., & Maddison, S. (1983). The Orbitofrontal Cortex : Neuronal Activity in the Behaving Monkey. *Experimental brain research*, *49*, 93–115.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic & physiological optics*, *12*, 229–32.

Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The retinotopy of visual spatial attention. *Neuron*, *21*, 1409–22.

Tootell, R. B., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, a M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in cognitive sciences*, *2*, 174–83.

Torralba, A., & Oliva, A. (1999). Semantic Organization of Scenes using Discriminant Structural Templates. *Proceedings of International Conference on Computer Vision*, 1253–1258.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, *14*, 391–412.

Torrisi, S. J., Lieberman, M. D., Bookheimer, S. Y., & Altshuler, L. L. (2013). Advancing understanding of affect labeling with dynamic causal modeling. *NeuroImage*, *82*, 481–8.

Ungerleider, L. G., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology*, *4*, 157–165.

Van Der Schaaf, A., & Van Hateren, J. H. (1996). Modelling the power spectra of natural images : statistics and information. *Vision Research*, *36*, 2759–2770.

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.

Van Essen, D., & Maunsell, J. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375.

Vann, S. D., Aggleton, J. P., & Maguire, E. a. (2009). What does the retrosplenial cortex do? *Nature reviews. Neuroscience*, *10*, 792–802.

Vannucci, M., Viggiano, M. P., & Argenti, F. (2001). Identification of spatially filtered stimuli as function of the semantic category. *Brain research. Cognitive brain research*, *12*, 475–8.

VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*, 655–668.

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

Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *The Journal of neuroscience*, *33*, 6133–42.

Vassilev, A., & Mitov, D. (1976). Perception time and spatial frequency. *Vision Research*, *16*, 89–92.

Viggiano, M. P., Righi, S., & Galli, G. (2006). Category-specific visual recognition as affected by aging and expertise. *Archives of gerontology and geriatrics*, *42*, 329–38.

Vlamings, P. H. J. M., Goffaux, V., & Kemner, C. (2009). Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information ? *Journal of vision*, *9(5)*, 1–13.

Volz, K. G., Schubotz, R. I., & von Cramon, D. Y. (2005). Variants of uncertainty in decision-making and their neural correlates. *Brain research bulletin*, *67*, 403–12.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature neuroscience*, *6*, 624–31.

Wagner, A. D., Pare, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering Meaning : Left Prefrontal Cortex Guides Controlled Semantic Retrieval. *Neuron*, *31*, 329–338.

Wais, P. E., Rubens, M. T., Boccanfuso, J., & Gazzaley, A. (2010). Neural mechanisms underlying the impact of visual distraction on retrieval of long-term memory. *The Journal of neuroscience*, *30*, 8541–50.

Wakana, S., Jiang, H., Zijl, P. C. M. Van, & Mori, S. (2004). Fiber tract-based Atlas of human white matter anatomy. *Radiology*, *230*, 77–87.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *The Journal of Neuroscience*, *29*, 10573–81.

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9661–6.

Wandell, B. a, Dumoulin, S. O., & Brewer, A. a. (2007). Visual field maps in human cortex. *Neuron*, *56*, 366–83.

Watanabe, T., Sasaki, Y., Miyauchi, S., Putz, B., Fujimaki, N., Nielsen, M., … Miyakawa, S. (1998). Attention-regulated activity in human primary visual cortex. *Journal of neurophysiology*, *79*, 2218–2221.

Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410.

Westheimer, G. (2001). The Fourier theory of vision. *Perception*, *30*, 531–541.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., … Gosselin, F. (2010). Does face inversion change spatial frequency tuning? *Journal of experimental psychology. Human perception and performance*, *36*, 122–35.

Xu, X., Anderson, T., & Casagrande, V. A. (2007). How do functional maps in primary visual cortex vary with eccentricity? *The Journal of comparative neurology*, *504*, 287–297.

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A, Pekar, J. J., & Courtney, S. M. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature neuroscience*, *5*, 995–1002.

Yvert, G., Perrone-Bertolotti, M., Baciu, M., & David, O. (2012). Dynamic causal modeling of spatiotemporal integration of phonological and semantic processes: an electroencephalographic study. *The Journal of neuroscience*, *32*, 4297–306.

Zeki, S. (1990). A century of cerebral achromatopsia. *Brain*, *113*, 1721–1777.

Zeki, S. (1991). Cerebral akinetopsia (visual motion blindess). *Brain*, *114*, 811–824.

LISTE DES FIGURES

Figure 1 : (a) Organisation rétinotopique des aires visuelles V1, V2 et V3. Les couleurs indiquent la correspondance entre la position dans le champ visuel (en termes d'excentricité par rapport à la fovéa - au premier plan, ou d'angle polaire – au deuxième plan) et la projection au niveau cortex visuel. Par exemple, on peut remarquer que la partie centrale du champ visuel (en rouge) est projetée sur la partie postérieure du cortex visuel, tandis que la périphérie du champ visuel (en violet) est projetée sur la partie antérieure du cortex visuel. Figure adaptée de Wandell, Dumoulin, & Brewer (2007). (b) Illustration de la magnification corticale dans V1. La surface correspondant au champ visuel central (à droite) est surreprésentée au niveau cortical (à gauche). Figure adaptée de Hubel (1994). 16

Figure 2 : Latences de réponse visuelle de neurones de différentes aires corticales. Pour chaque aire, les barres verticales centrales représentent la latence de réponse médiane (en millisecondes) et les barres latérales représentent les centiles de 10 à 90%. Les chiffres entre parenthèse renvoient aux références citées dans Bullier et al. (2001). On peut noter que les latences de réponse les plus courtes ne correspondent pas forcément aux niveaux les plus bas de la hiérarchie. En particulier, les aires MT, MST, ou FEF ont des latences de réponse comparables à celles de V1, bien qu'elles se situent plus haut dans la hiérarchie. Figure adaptée de Bullier (2001). 21

Figure 3 : Modèle intégré du traitement de l'information visuelle de Bullier (2001). Les flèches jaunes représentent la première vague d'activation véhiculée par la voie M, propagée des aires visuelle de bas niveau vers des aires visuelles de haut niveau pariétales et temporales. Les flèches orange représentent la vague d'activation descendante (rétro-injection), propagée des aires pariétales et temporales vers les aires visuelle de plus bas niveau. Cette rétro-injection a lieu au moment de l'arrivée plus tardive de l'information portée par la voie P (flèche rouge) sur le cortex visuel primaire, qui constitue la troisième vague d'activation. Figure adaptée de Bullier (2001). 22

Figure 4 : Exemple de la représentation d'une scène de ville dans le domaine spatial (à gauche) et de son spectre d'amplitude dans le domaine de Fourier (à droite). Sur le spectre d'amplitude, l'axe *x* correspond aux orientations verticales et l'axe *y* correspond aux orientations horizontales. Les basses fréquences sont codées au centre et les hautes fréquences sont codées en périphérie. Le spectre d'amplitude est symétrique selon les axes *x* et *y*. Sur cet exemple, on peut constater qu'une scène de ville se caractérise par une répartition de l'information en basses et hautes fréquences spatiales en majorité sur des orientations verticales et horizontales. 24

Figure 5 : Illustration de la procédure de filtrage d'une scène. Le spectre d'amplitude d'une scène est obtenu par la transformée de Fourier. On peut ainsi effectuer sur le spectre d'amplitude un filtrage passe-bas, passe-bande ou passe-haut pour ne conserver dans le domaine spatial (via la transformée de Fourier inverse) que l'information en basses, moyennes ou hautes fréquences spatiales, respectivement. Comme on peut le constater les BFS d'une scène contiennent une information grossière, floue, permettant d'avoir accès à sa structure globale. Les HFS contiennent quant à elles une information fine, permettant de distinguer les contours des objets et les détails de la scène. 25

Figure 6 : Exemples de réseaux sinusoïdaux à différentes fréquences spatiales. 26

Figure 7 : Exemple de figures hiérarchiques. A gauche, la forme globale (grand « H ») est congruente avec la forme locale (petit « H »). A droite, la forme globale (grand « H ») est non congruente avec la forme locale (petit « E »). 28

Figure 8 : Fonction de sensibilité au contraste chez l'adulte pour un réseau statique à différentes fréquences spatiales présenté en vision centrale. L'axe des abscisses représente la fréquence spatiale en cpd. L'axe des ordonnées correspond à la sensibilité au contraste (inverse du contraste : 1/c). L'échelle des axes est logarithmique. La surface sous la courbe correspond à la perception d'une alternance de bandes sombres et claires et la surface au-dessus de la courbe correspond à la perception d'une surface uniforme. Adapté de De Valois & De Valois (1988). 31

Figure 9 : Exemples d'image hybrides. A gauche, l'image hybride est composée des BFS d'une autoroute et des HFS d'une ville. A droite, l'image hybride est composée des BFS d'une ville et des HFS d'une autoroute. Adapté de Schyns et Oliva (1994). 35

Figure 10 : Exemples de visages hybrides utilisés dans l'étude de Schyns et Oliva (1999). A gauche, un visage de femme avec une expression neutre en BFS superposé à un visage d'homme en colère en HFS. A droite, un visage d'homme en colère en BFS superposé à un visage de femme neutre en HFS. 40

Figure 11 : Résultat du moyennage de 100 exemplaires de visages et de scènes appartenant à différentes catégories. On peut constater que les visages contiennent plus de régularités dans le domaine spatial que les scènes, qui sont difficilement reconnaissables une fois moyennées. Adapté de Torralba et Oliva (2003). 42
Figure 12 : Spectres d'amplitude moyens d'images de scènes de 14 catégories. L'axe horizontal dans le domaine de Fourier représente les orientations verticales dans le domaine spatial et l'axe vertical dans le domaine de Fourier représente des orientations horizontales dans le domaine spatial. Les trois contours correspondent, de l'extérieur vers l'intérieur, à 90, 80 et 60% de l'énergie du spectre. Adapté de Torralba et Oliva (2003). 44
Figure 13 : Exemple de stimuli utilisés par De Cesarei et al. (2011, 2013). Les séquences sont composées de 16 versions filtrées d'une même scène. Les séquences débutent soit par une image en BFS, à laquelle de l'information en HFS est progressivement ajoutée, soit par une image en HFS à laquelle de l'information en BFS est progressivement ajoutée. Adapté de De Cesarei et al. (2013). 51
Figure 14 : Résumé des résultats de l'Expérience 1. (a) Exemples de séquences utilisées dans l'Expérience 1. Pour les séquences Successives, l'information en fréquences spatiales change d'une image à l'autre dans la séquence et passe des BFS aux HFS (séquence CtF) ou des HFS aux BFS (séquence FtC). Pour les séquences Additives, de l'information en fréquences spatiales est ajoutée d'une image à l'autre dans la séquence et passe soit des BFS à une scène intacte (séquence CtF), soit des HFS à une image intacte (séquence FtC). (b) Temps de réponse corrects (en millisecondes) pour catégoriser les séquences CtF et FtC dans les conditions Successives et Additives. 53
Figure 15 : Résumé des résultats de l'Expérience 2. (a) Exemple de stimuli utilisés dans l'Expérience 2. Dans la condition FS-CON, le contraste des scènes filtrées n'est pas modifié. Les séquences passent des BFS avec un fort contraste aux HFS avec un faible contraste (séquences CtF) ou inversement (séquences FtC). Dans la condition FS, le contraste est égalisé entre les scènes filtrées. Les séquences passent des BFS aux HFS (séquences CtF) ou des HFS aux BFS (séquences FtC). Dans la condition CON les scènes composant les séquences ne sont pas filtrées mais le contraste est modifié pour correspondre à celui des scènes de la condition SF-CON. Les séquences passent d'un fort à un faible contraste (séquences CtF) ou d'un faible à un fort contraste (séquence FtC). (b) Temps de réponse corrects (en millisecondes) pour catégoriser les séquences CtF et FtC des conditions SF-CON, SF et CON. 56
Figure 16 : Cartographie de la représentation des fréquences spatiales et de l'excentricité dans le champ visuel obtenue par Sasaki et al. (2001). (a) Cartographie de la représentation des fréquences spatiales (BFS en vert et HFS en rouge) chez un participant. (b) Cartographie rétinotopique de l'excentricité dans le champ visuel (fovéa en rouge et périphérie en vert) chez le même participant. Adapté de Sasaki et al. (2001). 80
Figure 17 : Résultats obtenus par Musel et al. (2013) en fonction de la méthode d'analyse utilisée. La comparaison entre les deux conditions de fréquences spatiales révèle une organisation rétinotopique du traitement des fréquences spatiales tandis que la comparaison inter-hémisphérique (HD : hémisphère droit ; HG : hémisphère gauche) révèle une spécialisation hémisphérique pour le traitement des fréquences spatiales. Adapté de Musel et al. (2013). 82
Figure 18 : Illustration du mécanisme de facilitation de la reconnaissance d'objet proposé par Bar (2003). L'information en BFS du stimulus visuel est rapidement projetée du cortex visuel vers le cortex orbitofrontal (OFC). En parallèle, l'information visuelle (BFS et HFS) se propage plus lentement le long de la voie ventrale, jusqu'au gyrus fusiforme dans le cortex inférotemporal. Au niveau du cortex orbitofrontal, l'information en BFS permet d'activer des interprétations possibles du stimulus visuel, qui seront ensuite projetées vers le cortex inférotemporal et intégrées à l'analyse « bottom-up » de l'information visuelle, afin de faciliter sa reconnaissance. Adapté de Bar et al. (2006). 85
Figure 19 : Résumé des résultats obtenus par Bar et al. (2006). (a) Les résultats de l'Expérience 1 révèlent que la reconnaissance d'objets est associée à une activité plus importante du cortex orbitofrontal 130 ms après la présentation des stimuli, puis des gyri fusiformes droit et gauche 180 et 215 ms après présentation des stimuli, respectivement. (b) Les résultats de l'Expérience 2 (pourcentage de changement de signal dans le cortex orbitofrontal en fonction du contenu en fréquences spatiales de stimuli) montrent que les objets filtrés en BFS et non filtrés (NF) activent plus fortement le cortex orbitofrontal que les objets filtrés en HFS. Adapté de Bar et al. (2006). 87
Figure 20 : Résumé des résultats obtenus par Kveraga et al. (2007). (a) Activations induites par les stimuli M-biased (en jaune) et P-biased (en bleu : le cortex orbitofrontal (OFC) et le gyrus fusiforme (FG). (b) Résultats de l'analyse de connectivité. En haut : connectivité effective (i.e. influence qu'une région exerce sur l'activité d'une autre, indépendamment des conditions expérimentales) entre le gyrus occipital moyen (MOG), le cortex orbitofrontal (OFC) et le gyrus fusiforme (FG). Les chiffres correspondent aux paramètres de connectivité effective (exprimés en Hertz). En bas : effet de la condition P-biased (en rouge) sur la connectivité entre le MOG et le FG et de la condition M-biased (en gris) sur la connectivité entre le MOG

et l'OFC et entre l'OFC et le FG. Les chiffres correspondent à l'augmentation de la connectivité (en Hertz), par rapport aux paramètres de connectivité effective. 88
Figure 21 : Résumé des résultats obtenus par Peyrin et al. (2010). (a) Résultats des potentiels évoqués pour les séquences CtF et FtC. (b) Neuf cartes topographiques correspondant à la succession des différentes configurations du champ électrique pour chaque type de séquences. Les flèches colorées indiquent les périodes pour lesquelles l'analyse révèle des différences de topographie entre les deux types de séquences. Les cartes d'activations correspondent à la différence d'activation des sources correspondant à ces périodes. Les séquences CtF (relativement aux séquences FtC) induisent des potentiels évoqués de sources frontales et temporo-pariétales entre 150 et 194 ms après apparition de la première scène de la séquence, et entre 12 et 96 ms après apparition de la deuxième scène, et de source occipital entre 176 et 204 ms après apparition de la deuxième scène. 90
Figure 22 : Activations cérébrales induites par la catégorisation des scènes selon une séquence « coarse-to-fine » relativement à une séquence « fine-to-coarse ». (1) partie orbitale du gyrus frontal inférieur, (2) cuneus, (3) gyri fusiforme et parahippocampique, (4) gyrus frontal supérieur droit. 93
Figure 23 : Synthèse illustrée des résultats de l'analyse de connectivité. Les flèches en gras indiquent les liens de connectivité modulés par les séquences CtF (en noir) et FtC (en gris). La flèche en pointillé correspond à un lien de connectivité effective (i.e. non modulé par les conditions expérimentales). La flèche noire en bas correspond à l'input (stimulation visuelle) au niveau du cuneus. Les symboles « ++ » indiquent que la modulation par les séquences CtF a pour effet d'augmenter la connectivité entre les régions concernées. 94
Figure 24 : Exemple de scènes hybrides utilisées dans l'étude de Mu et Li (2013). Les scènes en BFS et en HFS composant les hybrides étaient sémantiquement et physiquement similaires (en haut à gauche), sémantiquement similaires et physiquement dissimilaires (en haut à droite), sémantiquement dissimilaires et physiquement similaires (en bas à gauche) ou sémantiquement en physiquement dissimilaires (en bas à droite). Adapté de Mu et Li (2013). 117
Figure 25 : Localisation anatomique de la PPA sur des coupes sagittale, axiale et coronale. 143
Figure 26 : Localisation anatomique du RSC sur des coupes sagittale, axiale et coronale. 144
Figure 27 : Localisation anatomique de l'OPA sur des coupes sagittale, axiale et coronale. 145
Figure 28 : Exemples de stimuli utilisés dans l'étude d'Andrews et al. (2010) 148
Figure 29 : Résumé des résultats obtenus par Goffaux et al. (2011). Différence de signal entre le traitement des visages intacts et « scrambles » en fonction du contenu en fréquences spatiales et du temps de présentation. On constate que les visages en BFS et en MFS induisent une activité plus importante que les visages en HFS pour des temps de présentation courts (75 ms) et ce pattern s'inverse en faveur des MFS et HFS avec l'augmentation du temps de présentation (dès 150 ms). Adapté de Goffaux et al. (2011). 150
Figure 30 : Résumé des résultats obtenus dans l'Expérience 5. Un « localizer » indépendant nous a permis de localiser fonctionnellement la PPA, le RSC et l'OPA chez chaque participant, et d'extraire au sein de ces régions les paramètres d'activation correspondant à la catégorisation des scènes non filtrées (NF) et filtrées en BFS et HFS des conditions LUM et RMS. 153
Figure 31 : Résumé des résultats obtenus dans l'Expérience 6. Le « localizer » nous a permis de localiser fonctionnellement la PPA, le RSC et l'OPA chez chaque participant et d'extraire au sein de ces régions les paramètres d'activation correspondant à la catégorisation des scènes selon une séquence CtF ou FtC. ... 176
Figure 32 : Exemple de scènes non filtrées (NF), filtrées en basses fréquences spatiales (BFS) et en hautes fréquences spatiales (HFS) en fonction de différentes valeurs du contraste RMS (0.20, 0.10 et 0.03). A noter que la valeur de contraste RMS 0.20 correspond à la valeur de contraste d'une scène filtrée en BFS, que la valeur de contraste RMS 0.10 correspond à celle utilisée pour égaliser le contraste des scènes filtrées dans les Expériences 2 et 5 et que la valeur de contraste RMS 0.03 correspond à la valeur de contraste d'une scène filtrée en HFS. 203
Figure 33 : Illustration de l'effet de l'égalisation RMS sur le spectre d'énergie d'une scène filtrée en BFS (à gauche) et d'une scène filtrée en HFS (à droite). On constate que pour les scènes en HFS, l'égalisation du contraste RMS (ligne verte) a pour effet d'ajouter de l'énergie dans les BFS (indiqué par la flèche rouge). 204
Figure 34 : Schématisation du mécanisme de codage prédictif. (a) Les prédictions générées dans les régions de haut niveau dans la hiérarchie influencent les traitements effectués dans les régions de plus bas niveau : le signal traité dans les régions de bas niveau est comparé aux prédictions générées par les régions de plus haut niveau dans la hiérarchie. Seule l'erreur de prédiction est transmise aux niveaux supérieurs. Ces interactions peuvent se faire entre niveaux voisins (flèches foncées) ou entre des niveaux plus éloignés dans la hiérarchie (flèches claires). Adapté de Rauss et Pourtois (2013). (b) Niveau d'activation en fonction de la nature des prédictions. Adapté de Price et Devlin (2011). 208
Figure 35 : Résumé des résultats obtenus par Fintzi et Mahon (2013). (a) Exemple de séquence d'objet passant des BFS (Bin1) aux HFS (Bin 6). (b) Pourcentage de voxels activés de façon maximale au sein du cortex orbitofrontal (OFC), inférotemporal (VT) et occipital (LO) en fonction du contenu en fréquences spatiales

dans la séquence. (c) Pourcentage de voxels activés de façon maximale au sein du cortex orbitofrontal et du cortex inférotemporal en fonction de l'hémisphère et du contenu en fréquences spatiales dans la séquence. Adapté de Fintzi et Mahon (2013). 211
Figure 36 : Modèle neurocognitif de la catégorisation de scènes proposé sur la base des résultats obtenus dans ce travail de thèse. 217

LISTE DES ANNEXES

ANNEXE I : Revue de littérature sur les bases cérébrales du traitement des fréquences spatiales publiée dans la revue « Frontiers in Integrative Neuroscience »……..ii

ANNEXE II : Article relatif à l'étude des effets de l'âge sur le traitement des fréquences spatiales. En révision dans la revue « PloS One »……..xx

ANNEXE III : Curriculum Vitae…….xlv

frontiers in
INTEGRATIVE NEUROSCIENCE

REVIEW ARTICLE
published: 07 May 2014
doi: 10.3389/fnint.2014.00037

The neural bases of spatial frequency processing during scene perception

Louise Kauffmann*[1,2], *Stephen Ramanoël*[1,2] *and Carole Peyrin*[1,2]*

[1] *University Grenoble Alpes, LPNC, Grenoble, France*
[2] *CNRS, LPNC, Université Pierre Mendès France, Grenoble, France*

Edited by:
Olivier A. Coubard, CNS-Fed, France
Reviewed by:
Summer Sheremata, George Washington University, USA
Michelle R. Greene, Stanford University, USA
****Correspondence:***
Carole Peyrin, Centre National de la Recherche Scientifique, Laboratoire de Psychologie et NeuroCognition, UMR 5105, Université Pierre Mendès France, BP 47, 38040 Grenoble Cedex 09, France e-mail: carole.peyrin@ upmf-grenoble.fr

Theories on visual perception agree that scenes are processed in terms of spatial frequencies. Low spatial frequencies (LSF) carry coarse information whereas high spatial frequencies (HSF) carry fine details of the scene. However, how and where spatial frequencies are processed within the brain remain unresolved questions. The present review addresses these issues and aims to identify the cerebral regions differentially involved in low and high spatial frequency processing, and to clarify their attributes during scene perception. Results from a number of behavioral and neuroimaging studies suggest that spatial frequency processing is lateralized in both hemispheres, with the right and left hemispheres predominantly involved in the categorization of LSF and HSF scenes, respectively. There is also evidence that spatial frequency processing is retinotopically mapped in the visual cortex. HSF scenes (as opposed to LSF) activate occipital areas in relation to foveal representations, while categorization of LSF scenes (as opposed to HSF) activates occipital areas in relation to more peripheral representations. Concomitantly, a number of studies have demonstrated that LSF information may reach high-order areas rapidly, allowing an initial coarse parsing of the visual scene, which could then be sent back through feedback into the occipito-temporal cortex to guide finer HSF-based analysis. Finally, the review addresses spatial frequency processing within scene-selective regions areas of the occipito-temporal cortex.

Keywords: natural scene, spatial frequencies, coarse-to-fine, hemispheric specialization, retinotopy, parahippocampal place area

INTRODUCTION

Scenes containing more realistic and more natural stimuli have increasingly become the object of scientific interest over the last 20 years, as they involve the perception of stimuli which are more complex and more realistic than simple objects or drawings. It is now widely agreed that visual recognition of scenes is a fast, automatic and reliable process. Experimental studies have shown that complex natural scenes can be categorized very rapidly (under 150 ms; Thorpe et al., 1996), indicating that a simple and efficient coding process is involved. There is considerable evidence suggesting the importance of the spatial frequency contents of images during scene recognition (Ginsburg, 1986; Field, 1987; Tolhurst et al., 1992; Hughes et al., 1996). On one hand, the primary visual cortex is mainly dominated by complex cells which respond preferentially to spatial frequencies (Poggio, 1972; De Valois et al., 1982a,b). On the other hand, findings from simulation and psychophysical experiments indicate that information from low/medium frequencies of the amplitude spectrum suffices to enable scene categorization (Torralba and Oliva, 2003; Guyader et al., 2004). Supported by convergent data from the functional neuroanatomy of magnocellular and parvocellular visual pathways (Van Essen and Deyoe, 1995), neurophysiological recordings in primates (for a review, see Bullier, 2001), and psychophysical results in humans (Ginsburg, 1986; Hughes et al., 1996), current influential models of visual perception (Schyns and Oliva, 1994; Bullier, 2001; Bar, 2003; Hegde, 2008) suggest that the first stage of visual analysis consists of the extraction of visual elementary features at different spatial frequencies. Low spatial frequencies (LSF), conveyed by fast magnocellular pathways, provide a coarse information about a visual stimulus (e.g., the global shape and structure of a scene), whereas high spatial frequencies (HSF), conveyed more slowly by the parvocellular pathways, provide finer information about the stimulus (e.g., the edges and borders of an object in the scene).

However, exactly how and where spatial frequencies are processed within the brain remain unsettled questions. The debate on retinotopic organization and/or the existence of cerebral asymmetries in the occipital cortex in spatial frequency processing is still ongoing in the literature. A number of studies demonstrated a retinotopic mapping of spatial frequency processing in the occipital cortex and have

for example showed that the perception of HSF sinusoidal gratings activated the foveal representation in all retinotopic areas of the occipital cortex, and LSF sinusoidal gratings activated more peripheral representations in the same cortical areas (Sasaki et al., 2001; Henriksson et al., 2007). However, other authors argue in favor of the hemispheric specialization for spatial frequency processing at the level of visual retinotopic areas, with the right hemisphere preferentially specialized in the processing of LSF information and the left hemisphere preferentially specialized in HSF information processing (Iidaka et al., 2004; Peyrin et al., 2004, 2006). It appears therefore important to investigate both retinotopic processing and hemispheric specialization on the same visual stimuli.

In addition, there is considerable evidence suggesting that spatial frequency processing takes place in a predominantly and default coarse-to-fine sequence (**Figure 1**) However, the cerebral circuit of the coarse-to-fine perception of scenes has never been investigated in humans. On the basis of neurophysiological recordings in nonhuman primates, Bullier (2001) suggested that during perception of a scene, LSF which are conveyed more rapidly than HSF by fast magnocellular pathways, access the occipital cortex and high-order cortical areas in the dorsal cortical stream (parietal and frontal cortices) allow a coarse perceptual parsing of the visual input, prior to their complete propagation along the ventral cortical stream (inferotemporal cortex) which ultimately mediates the input recognition. The initial low-pass analysis would serve to refine the subsequent processing of HSF, conveyed more slowly by parvocellular pathways through the ventral cortical stream. It is now essential to match data from non human primates with human data. Finally, the ventral visual stream contains a mosaic of different areas that respond selectively to different categories of visual stimuli (Haxby et al., 2001; Lerner et al., 2001; Spiridon and Kanwisher, 2002). While several studies (Epstein and Kanwisher, 1998; Bar and Aminoff, 2003; Bar, 2004; Epstein, 2005, 2008; Aminoff et al., 2007; Epstein and Higgins, 2007; Dilks et al., 2013) agree that a prominent region in the inferotemporal cortex known as the parahippocampal place area (PPA), the retrosplenial cortex (RSC) and a region around the transverse occipital sulcus called the occipital place area (OPA) all play a major role in the perception of scenes in humans, the specific functions supported by scene-selective regions during the spatial frequency processing in scenes remain unclear.

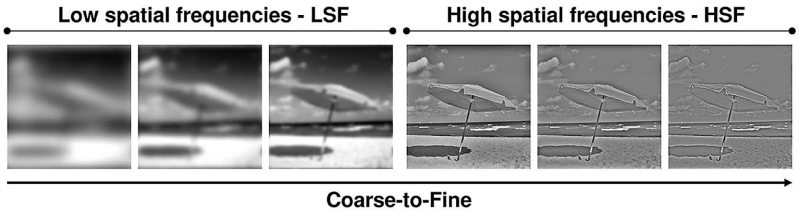
The present review addresses these issues and aims to identify the cerebral regions differentially involved in low and high spatial frequency processing and to clarify their attributes during scene perception.

NEURAL CORRELATES OF SPATIAL FREQUENCY PROCESSING DURING SCENE PERCEPTION

Many authors postulate that the two cerebral hemispheres are differently involved in spatial frequency processing, the right hemisphere predominating in the processing of LSF, and the left hemisphere predominating in the processing of HSF. Cerebral asymmetries have been observed in behavioral studies conducted on healthy participants (Sergent, 1982, 1983; Sergent and Hellige, 1986; Kitterle et al., 1990, 1992; Chokron et al., 2003; Peyrin et al.,2003), in neurological patients (Robertson et al.,1988; Lamb et al., 1990; Robertson and Lamb, 1991; Peyrin et al., 2006; Dos Santos et al., 2013), and from functional neuroimaging studies (Fink et al., 1996, 2000; Martinez et al., 1997, 2001; Heinze et al., 1998; Kenemans et al.,2000; Mangun et al.,2000;Yamaguchi et al.,2000; Wilkinsonet al.,2001; Hanet al.,2002; Iidakaet al.,2004; Luxet al., 2004; Peyrin et al., 2004; Weissman and Woldorff, 2005; Musel et al., 2013). However, the hemispheric specialization for spatial frequency processing was largely inferred from studies assessing cerebral asymmetries during the processing of global and local information.

PSYCHOPHYSICAL ARGUMENTS FOR HEMISPHERIC SPECIALIZATION IN SPATIAL FREQUENCY PROCESSING

The first experimental evidence in support of hemispheric specialization for global and local processing comes from psychophysical studies using hierarchical forms as visual stimuli (i.e., in general a large global letter made up of small local letters; Navon, 1977; Kinchla and Wolfe, 1979; **Figure 2A**).



iii

FIGURE 1 | Coarse-to-fine sequence of spatial frequency processing (from low-to-high spatial frequencies) during scene perception.

Using hierarchical visual stimuli displayed in either the left visual field (projecting directly to the right hemisphere) or the right visual field (projecting directly to the left hemisphere), Sergent (1982) demonstrated that the identification of the global letter was faster when displayed in the left visual hemifield/right hemisphere, and that the identification of local letters occurred more rapidly when they were displayed in the right visual hemifield/ left hemisphere. These results suggest a right hemispheric specialization for the processing of global information, and a left hemispheric specialization for the processing of local information. Based on evidence that global information is predominantly conveyed by LSF, and that local information is predominantly conveyed by HSF (Schulman et al., 1986; Badcock et al., 1990; Lamb and Yund, 1993), the cerebral asymmetries observed during global and local processing have been interpreted as reflecting the hemispheric specialization for LSF and HSF processing, respectively (Sergent, 1982). However, the relationship between global and local information, and LSF and HSF, respectively, is far from univocal within hierarchical visual forms (Palmer, 1993). It is for example possible that global information is conveyed by both LSF and HSF. Hemispheric specialization for spatial frequency processing was therefore subsequently

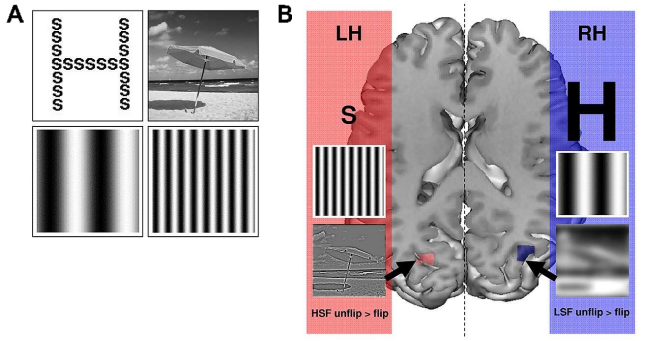


FIGURE 2 | **(A)** Example of stimuli used to assess cerebral asymmetries for spatial frequencies. Top-left: Hierarchical visual forms, consisting of a large global letter made up of small local letters; Top-right: Scene; Bottom-left: low spatial frequency (LSF) sinusoidal grating; Bottom-right: high spatial frequency (HSF) sinusoidal grating. **(B)** Hemispheric specialization: the left hemisphere (LH) is predominantly involved in the local letter identification, HSF grating identification and HSF categorization; the right hemisphere (RH) is predominantly involved in the global letter identification, LSF grating identification and LSF categorization. Activations reported showed stronger activation in the left than the right occipital cortex for HSF categorization [(HSF unflip > flip) contrast] and stronger activation in the right than the left occipital cortex for LSF categorization [(LSF unflip > flip) contrast]. Figure adapted from Peyrin et al. (2004) .

tested by directly manipulating the spatial frequency content of visual stimuli, using either sinusoidal gratings (Kitterle et al.,1990,1992; Kitterle and Selig,1991; **Figure2A**) or scene images (Peyrin et al., 2003, 2006; **Figure2A**). It should be noted that this type of manipulation is not feasible with hierarchical forms because low-pass filtering cancels out the local form and renders the task impossible. We evaluated hemispheric asymmetry in healthy participants in a series of psychophysical studies (Peyrin et al., 2003, 2006), by making explicit changes in the spatial frequency spectrum of scene images, which were displayed in either the left or the right visual fields. In the initial study, participants were asked to recognize a target scene (a city or a highway) filtered in either LSF or HSF (Peyrin et al., 2003). Results showed more rapid recognition of LSF scenes when they were displayed in the left visual hemifield/right hemisphere than when they were presented in the right visual field/left hemisphere. Conversely, recognition of HSF scenes occurred more rapidly in the right visual hemifield/left hemisphere than the left visual hemifield/right hemisphere. This study demonstrated a right hemispheric predominance for LSF and a left hemispheric predominance for HSF processing. It should be noted that the hemispheric specialization in question has been observed in males, but not in females (Peyrin et al., 2006). These results are consistent with studies showing a lesser degree of lateralization in female functional cerebral organization compared to males (McGlone and Kertesz, 1973; Voyer, 1996). Certain factors of interference, which may affect processing speed may render detection of hemispheric specialization in healthy females more difficult. For example, the hormonal level fluctuations over the menstrual cycle has been evidenced to modulate hemispheric asymmetries in visual, attentional, and language processes (Hausmann and Güntürkün, 2000; Hausmann et al., 2002; Hausmann, 2005), and to affect interhemispheric transfer time (Hausmann et al., 2013).

NEURAL CORRELATES OF HEMISPHERIC SPECIALIZATION IN SPATIAL FREQUENCY PROCESSING

Neuropsychological and neuroimaging studies that use hierarchical visual forms provide conflicting evidence on which cortical structures present hemispheric specialization. Robertson et al. (1988) showed impairment in the performance of tasks involving the perception of hierarchical visual form in patients with unilateral damage to the temporo-parietal junction. Performance of patients with a lesion situated in the left superior temporal cortex was impaired during the identification of local elements, whereas patients suffering from lesions in the right temporo-parietal

junction exhibited poor performance during the identification of the global form. These data suggests the right temporo-parietal junction specialization for global processing, and the left temporo-parietal junction specialization for local processing. However, using positron emission tomography, Fink et al. (1996; see also Fink et al., 1997, 2000) reported cerebral asymmetries at a lower level of visual cortical processing, with a right lingual gyrus activation during the identification of the global form and a left inferior occipital gyrus activation during the identification of local elements. Using event-related brain potentials (ERPs), Heinze et al. (1998; see also Mangun et al., 2000) failed to show hemispheric specialization in the first-stage of the visual analysis. Instead, their results show long latency asymmetries (260–360 latency range) for global and local processing, suggesting that cerebral asymmetries was rather present at the higher-stage of the visual analysis. Some functional imaging data have, furthermore, revealed an attentional cortical mechanism located in the temporo-parietal junction which controls the attentional selection of information presented either at global or the local level depending on the visual task demands (Robertson et al., 1988; Robertson and Lamb, 1991; Fink et al., 1996; Yamaguchi et al., 2000; Wilkinson et al., 2001; Weissman and Woldorff, 2005). For example, Yamaguchi et al. (2000) recorded ERPs while participants shifted their attention to the global or local level of hierarchical visual forms (the shift direction was controlled by a cue preceding the stimulus). Cerebral asymmetries were observed during the global and local processing of hierarchical forms, but also during the time interval of attention directed toward global or local levels by the cues. ERP responses indicated greater right-hemisphere amplitudes located in the right temporo-parietal junction when attention was directed at global level, and greater left-hemisphere amplitudes located in the left temporo-parietal junction when it was directed at local level. This study provided a neural basis for a “top-down” mechanism of allocation of attention to global and local information, and revealed the asymmetrical involvement of the temporal-parietal regions.

Neuroimaging studies previously mentioned have provided conflicting results concerning hemispheric specialization for spatial frequency processing using hierarchical visual forms as stimuli. Subsequent studies, including those of our own team, which involved the direct manipulation of the spatial frequency content of stimuli, provided evidence of hemispheric specialization involving occipital areas (Iidaka et al., 2004; Peyrin et al., 2004). In an fMRI study, Peyrin et al. (2004) investigated the hemispheric specialization for spatial frequency processing during the recognition of LSF and HSF scenes (city vs. highway scenes at a visual angle of 4◦). Comparison of LSF to HSF scene recognition, revealed significant activation in regions which are known to be involved in scene processing: the right anterior temporal region which is particularly sensitive to familiar versus unfamiliar scenes (Nakamura et al., 2000), and the right parahippocampal gyrus which is known to be involved in tasks requiring the retrieval of topographical information in scenes (Maguire et al., 1998; it should be noted that right-side parahippocampal gyrus activation did not correspond to PPA activation reported by EpsteinandKanwisher,1998). These results suggest that in Peyrin et al. (2004), scene perception was based mainly on LSF extraction and analysis, and they support the models proposing the prevalence of LSF information in scene categorization (coarse-to-fine strategy; Schyns and Oliva, 1994). Significant activation also occurred in the right inferior parietal lobule near the temporo-parietal junction. This activation was interpreted as reflecting an attentional control mechanism during spatial frequency selection. Yamaguchi et al. (2000) had previously shown cerebral activity in the right temporo-parietal area for a global attention shift during the perception of hierarchical letter forms (i.e., allocation of attention to global information). Finally, LSF scene recognition (as opposed to HSF) activated the superior temporal cortex bilaterally. This particular result concerned us, because it contradicted neuropsychological studies (Robertson et al., 1988; Lamb et al., 1990; Robertson and Lamb, 1991), showing specialization of the right superior temporal cortex in the perceptual processing of global information (supposed to be preferentially conveyed by LSF), and specialization of the left superior temporal cortex in the perceptual processing of local information (supposed to be preferentially conveyed by HSF). It should be noted that HSF scene recognition (as opposed to LSF) failed to show significant activation, suggesting a processing bias toward LSF.

Based on behavioral studies in which performances between the two visual hemifields are directly compared (see our abovementioned original psychophysical experiments; Sergent, 1982; Kitterle et al., 1990, 1992; Kitterle and Selig, 1991; Peyrin et al., 2003, 2006), we suggested to directly compare activation between the two hemispheres in order to assess cerebral asymmetries in fMRI study. For this purpose, we created an fMRI method of direct inter-hemispheric comparison. Two sets of functional volumes, obtained from functional scans, are compared at individual level. One set is represented by functional volumes in accordance with neurological convention (the left hemisphere appears on the left side of images) and the other set is represented by the same functional volumes this time in accordance with radiological convention (the right hemisphere appears on the left side of images). Images from the second set are "flipped" by 180◦ in the midsagital plane, thus providing "mirror" images of the first set. Contrasts between "unflipped" and "left-right flipped" functional volumes from the same experimental condition allow to compare activity in one hemisphere with activity in homologous regions of the other hemisphere (Iidaka et al., 2004; Peyrin et al., 2004, 2005; Musel et al., 2013; see also Cousin et al., 2006 for an application of this method on language processes; **Figure3**). This method revealed greater activation in the right than the left middle occipital gyrus for LSF scene recognition, and greater activation in the left than the right middle occipital gyrus for HSF scene recognition (**Figure 2B**). This study provided new evidence for hemispheric specialization at the first cortical level of visual analysis. Analyzing fMRI data with a more traditional approach which contrasts spatial frequencies to one another, we observed a higher degree of activation for LSF scenes (as opposed to HSF), while the reverse contrast did not reveal any significant activation. This study suggests that the results considerably differ according to the method applied to analysis fMRI data. Inter-hemispheric comparison seems more appropriate for the investigations of cerebral asymmetries, since it allows any main effect deriving from spatial frequency bias to be canceled out.

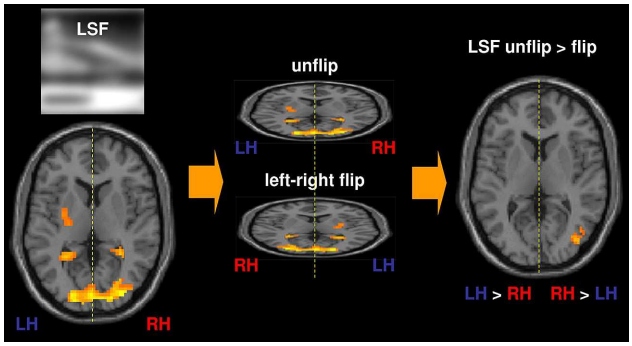


FIGURE 3 | Method of direct inter-hemispheric comparison. Two sets of functional volumes, obtained from functional scans, are compared at individual level. One set is represented by functional volumes in accordance with neurological convention (the left hemisphere – LH appears on the left side of images) and the other set is represented by the same functional volumes this time in accordance with radiological convention (the right hemisphere – RH appears on the left side of images). Images from the second set were "flipped" by 180◦ in the midsagital plane, thus providing "mirror" images of the first set. Contrasts between "unflipped" and "left-right flipped" images were then calculated for each of the spatial frequency components of natural scenes. In order to assess hemispheric predominance during the perception of LSF scenes, for instance, the following contrast was calculated: LSF unflip > flip. Regions which were more highly activated in the left hemisphere than in the right hemisphere appear on the left side, and regions which were statistically more highly activated in the right hemisphere than in the left hemisphere appear on the right side.

We proceeded to investigate the role of the occipital cortex in spatial frequency processing using a neuropsychological approach (Peyrin et al., 2006). We studied the categorization of LSF and HSF scenes in a female neurological patient who suffered from a focal lesion in the right occipito-temporal cortex following the embolization of an arterioveinous malformation. This lesion had induced a left homonymous hemianopsia. Two evaluations were conducted, the first 1 week prior to surgical intervention and the second 6 months afterward. As expected, the performance of the patient was more severely impaired for LSF than HSF scene recognition following embolization. This result suggests again the right occipital cortex specialization for LSF, and on a more general level suggests that hemispheric specialization could occur in women, although this is difficult to demonstrate behaviorally in the healthy population. This finding highlights the necessity of studying males and females together and both normal and brain-damaged patients' performance in order to establish the neural correlates of visual functions.

The extent to which hemispheric asymmetries during spatial frequency processing result from perceptual or attentional processes remains to be determined. While some studies have clearly demonstrated that attentional processes exert control on hemispheric specialization in the processing of global and local information at high-level stages of visual processing (e.g., via the temporo-parietal junction; Robertson et al., 1988; Robertson and Lamb, 1991; Fink et al., 1996; Heinze et al., 1998;

Yamaguchi et al., 2000; Wilkinson et al., 2001; Weissman and Woldorff, 2005), other studies have evidenced hemispheric asymmetries at lowerlevel stages, in the occipital cortex (Fink et al., 1997, 2000; Peyrin et al.,2004; Musel et al.,2013). However, activation of the occipital cortex was frequently associated with activation of the temporo-parietal junction in these studies. This cortical structure may have exerted attentional influence on lower-level areas. Furthermore, a number of neuroimaging studies have evidenced attentional modulation of activity in early visual areas (Tootell et al., 1998; Watanabe et al., 1998; Brefczynski and DeYoe, 1999; Gandhi et al., 1999; Martinez et al., 1999; Sasaki et al., 2001; Silver et al., 2007; Saygin and Sereno, 2008). For example, Martinez et al. (1999) showed that attending to a target whose location was cued by an arrow enhanced the amplitude of activation in striate and extrastriate visual areas. Cerebral asymmetries observed at low-level stages of visual processing, such as the occipital cortex, may not, therefore, necessarily result from strictly perceptual processes.

However, despite the considerable body of research in favor of the hemispheric specialization for spatial frequency processing in the occipital cortex, other authors postulate that a spatial frequency processing mapping according to the retinotopic organization of the visual cortex.

RETINOTOPIC PROCESSING OF SPATIAL FREQUENCIES

Imaging data obtained from patients with cerebral lesions (Holmes, 1918; Horton and Hoyt, 1991) and from healthy participants (Engel et al., 1994, 1997) show that the human primary visual cortex is retinotopically organized. The central (foveal) part of the visual field is represented at the very back of the visual cortex and more peripheral regions of the visual field are represented further forward (**Figure4A**). Importantly, the distribution of retinal photoreceptors and retinal ganglion cells is nonhomogeneous throughout the retina (Curcio and Allen, 1990; Curcio et al., 1990). The density of cones and midget ganglion cells from which the parvocellular pathway originates and which are used to process HSF information, is greatest in the fovea, while the density of rods and parasol ganglion cells from which the magnocellular pathway originates and which are used to process LSF information, increases with foveal eccentricity. Therefore HSF information could be predominantly processed in the areas dedicated to foveal vision. Similarly, LSF information might well be predominantly processed in the areas devoted to peripheral vision.

Neurophysiological studies performed on cats (Everson et al., 1998; Issa et al., 2000), primates (De Valois et al., 1982a; Foster et al., 1985; Tootell et al., 1988; Gegenfurtner et al., 1997; Xu et al., 2007) and humans (Singh et al., 2000; Sasaki et al., 2001; Henriksson et al., 2007) have mapped the representation of spatial frequencies in retinotopic areas. In an fMRI study, using retinotopic encoding with achromatic

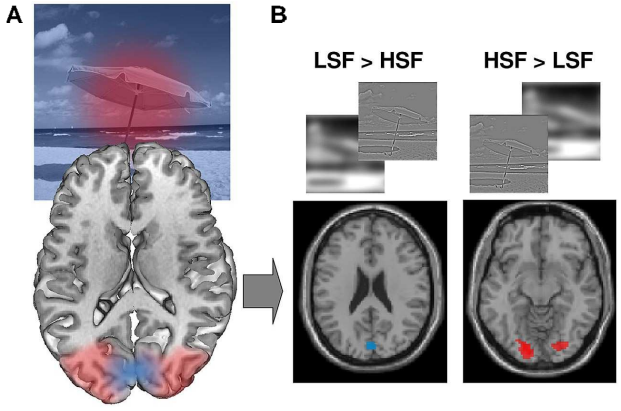


FIGURE 4| **(A)** Retinotopic mapping of the visual field on the visual cortex. The central (foveal) part of the visual field is represented at the very back of the visual cortex and laterally. More peripheral regions of the visual field are represented further forward in the medial part of the visual cortex. **(B)** Retinotopic organization of spatial frequency processing during scene perception: LSF [as opposed to HSF, (LSF > HSF) contrast] scene categorization recruits areas dedicated to peripheral vision, while HSF [as opposed to LSF, (LSF > HSF) contrast] scene categorization recruits areas dedicated to foveal vision. Figure adapted from Musel et al. (2013) .

sinusoidal gratings, Sasaki et al. (2001) showed that LSF were mapped on the peripheral visual field representation of the occipital cortex, whereas HSF were mapped on the central visual field representation. More recently, Henriksson et al. (2007) evidenced that in the retinotopic area of the occipital cortex, lower spatial frequencies selectivity was observed as eccentricity of the achromatic sinusoidal grating increased. Other studies have provided evidence of consistent cortical retinotopic mapping of more complex cognitive functions, such as visual spatial attention (Tootell et al., 1998; Watanabe et al., 1998; Brefczynski and DeYoe,1999;Gandhi et al.,1999; Martinezet al.,1999; Sasaki et al., 2001; Silver et al., 2007) and working memory (Pratte and Tong, 2014) in the early visual areas, as well as in the higher cortical areas, such as in the temporal, parietal, and frontal cortices (Silver et al., 2005; Hagler and Sereno, 2006; Wandell et al., 2007; Saygin and Sereno,2008;Arcaro et al.,2009,2011; Sheremata et al., 2010). Tootell et al.

(1998), for example, showed that paying attention to a specific location in the visual field increased activity in the corresponding retinotopic location of the extrastriate visual areas. Attentional modulation which was similar, albeit to a lesser degree, was also observed in the primary visual cortex (V1). Saygin and Sereno (2008) subsequently investigated the independent modulation of retinotopic responses by visual stimulus properties and attention in a number of areas exhibiting retinotopic organization (in the occipital cortex, the precuneus, the motion-sensitive temporal cortex, the intraparietal sulcus, and the frontal eye fields in the frontal cortex). These authors used retinotopically rotating polar angle mapping with point-light biological motion figures as complex visual stimuli. Participants fixated and viewed a rotating pie-shaped wedge containing biological motion figures. In the background, biological motion figures were either surrounded by either scrambled figures (stimulus contrast) or similar figures (no stimulus contrast). Participants were asked to perform one of two tasks while fixating – they were asked to attend to either the wedge (attention) or to the center of gaze (no attention). The authors demonstrated that the retinotopy of early visual areas was mainly driven by visual stimuli contrast, that the retinotopy of classical attentional control areas in the parietal and frontal cortices was mainly driven by attention, and that the retinotopy of lateral temporal regions was driven by both. In a recent study, Bressler et al. (2013) measured the effects of endogenous visual spatial attention (i.e., attention directed voluntarily by the participant) on the amplitude of retinotopic responses in the occipital and parietal cortices. Participants were asked to direct their attention toward a target of different eccentricities and to detect a target during retinotopic mapping. The authors showed that attending to the target in the visual field enhanced the amplitude of activations in corresponding retinotopic cortical locations for all the areas investigated, but that the modulation of retinotopic responses depended on target eccentricity. In occipital areas (V1, V2, V3, and hV4), directed attention elicited greater activation in cortical locations which corresponded to target eccentricities closer to the center than those which were farther out. Conversely, in parietal areas, directed attention elicited greater activation in target eccentricities which were farther away than in those which were closer. The authors suggest that endogenous attention potentially plays a role in processing the fine details of an object in central vision and in detecting relevant objects in the periphery during motor planning. Interestingly, Sasaki et al. (2001) provided direct evidence of retinotopic modulation of response resulting from global and local attentional demands in the occipital cortex. The authors used very large hierarchical arithmetic symbols (for example, a global "x" form composed of several local "+" elements). During "attention to global" periods, participants focused their attention on the global symbol (the "x") involving their peripheral vision, and during "attention to local" periods, they were instructed to focus attention on local symbols (the "+"), involving foveal vision. Results showed that when attention was directed at local (as opposed to global) level, activation occurred in the visual areas in relation to the foveal representation. When attention was directed at global (as opposed to local) level, activation was consistent with peripheral cortical representation. Since it can be assumed that global processing is mediated by low-pass spatial analysis, and local processing is mediated by high-pass spatial analysis (Schulman et al., 1986; Badcock et al., 1990; Lamb and Yund, 1993), the retinotopic organization observed in global and local attentional processing may constitute an argument in favor of a retinotopic organization for the attentional selection of spatial frequencies.

On the whole, the neuroimaging studies mentioned previously either highlight retinotopic mapping of spatial frequency processing (Sasaki et al., 2001), or reveal hemispheric specialization for spatial frequency processing (Iidaka et al., 2004; Peyrin et al., 2004). A recent fMRI study showed that spatial frequency processing could be both retinotopically mapped and lateralized between the two hemispheres (Musel et al., 2013).

RETINOTOPIC AND LATERALIZED PROCESSING OF SPATIAL FREQUENCIES DURING SCENE CATEGORIZATION

After demonstrating retinotopic organization of spatial frequency processing, Sasaki et al. (2001) concluded that neither global nor local processing was lateralized in the occipital cortex. However, the authors compared activation elicited by global and local conditions to one another (traditional method of fMRI data analysis), rather than activation between hemispheres (direct inter-hemispheric comparison method used in Peyrinet al.,2004). Musel et al. (2013) evaluated both the retinotopy and the functional lateralization of spatial frequency processing using a categorization task of scenes (indoors vs. outdoors) filtered in HSF and LSF. They used larger scene images (with a visual angle of $24^{\circ} \times 18^{\circ}$) than in Peyrin et al. (2004) in which the visual angle was $4^{\circ} \times 4^{\circ}$, thus covering the same breadth

of visual field as Sasaki et al. (2001). Results provided firstly evidence of retinotopic processing of spatial frequencies. At group level, the comparison between the spatial frequency content revealed that LSF scene categorization (as opposed to HSF) elicited activation in the anterior half of the calcarine fissures linked to the peripheral visual field, whereas HSF scene categorization (as opposed to LSF) elicited activation in the posterior part of the occipital lobes which are linked to the fovea, according to the retinotopic property of visual areas (**Figure 4B**). The retinotopic organization of spatial frequencies was also assessed at individual level by projecting LSF and HSF related activations onto retinotopic maps established for a number of participants. Functional activations projected onto individual retinotopic maps revealed that LSF processing is mapped in the anterior part of V1, whereas HSF processing is mapped in the posterior and ventral part of V2, V3, and V4. Furthermore, at the group level, the direct inter-hemispheric comparisons performed on the same fMRI data revealed a right-sided occipito-temporal predominance for LSF scene categorization and a left-sided temporal cortex predominance for HSF scene categorization, according to the hemispheric specialization theories. By using suitable method of fMRI analysis on the same data, as well as visual stimuli filtered in spatial frequencies covering a large part of the visual field, Musel et al. (2013) demonstrated for the first time retinotopic and lateralized spatial frequency processing in the human occipito-temporal cortex. It should be noted that hemispheric asymmetries were also highlighted within retinotopically defined parietal and frontal cortices during spatial working memory tasks (Sheremata et al., 2010; Szczepanski et al., 2010; Szczepanski and Kastner, 2013).

However, results from certain neurophysiological, computational, and behavioral studies indicate that the totality of spatial frequency information is not immediately conveyed through the brain, but that analysis follows a predominantly coarse-tofine processing sequence. If LSF extraction and analysis occurs first, followed by that of HSF, why should there be any hemispheric lateralization for the processing of LSF or HSF? Identification of the neural basis of the coarse-to-fine analysis in scene perception is the first step toward responding to this question.

COARSE-TO-FINE PROCESSING DURING SCENE PERCEPTION

PSYCHOPHYSICAL ARGUMENTS OF COARSE-TO-FINE PROCESSING Data from the functional neuroanatomy of magnocellular and parvocellular visual pathways indicate that the totality of visual information is not conveyed immediately, but that LSF reach the visual cortex before HSF (Van Essen and Deyoe, 1995; Bullier, 2001), although some controversies still remain (Merigan and Maunsell, 1993; Kaplan, 2004). A temporal precedence of LSF processing over HSF has been observed in psychophysical studies using sinusoidal gratings (Breitmeyer, 1975; Ginsburg, 1986; Hughes et al., 1996). Studies manipulating spatial frequency content of faces and scenes have provided further evidence of a coarse-to-fine processing sequence (Schyns and Oliva, 1994, 1997, 1999; Parker et al., 1996; Oliva and Schyns, 1997; Musel et al., 2012). Schyns and Oliva (1994) used hybrid images made of two superimposed scenes belonging to different categories and containing different spatial frequency bands (e.g., a city scene in LSF superimposed on a highway scene in HSF). When presentation time of hybrids was very short (30 ms), categorization of the hybrid was dominated by LSF information. However, categorization was dominated by HSF information for longer presentation times (150 ms). This suggests that LSF take precedence over HSF during scene perception. Furthermore, when the authors displayed two successive hybrids depicting simultaneously a coarse-to-fine sequence for a given scene (a LSF city in the first hybrid follows by a HSF city in the second hybrid) and a fine-to-coarse sequence for another scene (a HSF highway in the first hybrid follows by a LSF highway in the second hybrid), scene categorization was more frequently based on the coarse-to-fine than the fine-to-coarse sequence.

Although LSF information may be perceptually available before HSF, it is important to note that it does not necessarily follow that it is always used first to support visual recognition in all tasks. In Schyns and Oliva (1994), scene categorization in hybrid sequences was in fact based on a fine-to-coarse rather than a coarse-to-fine sequence in a substantial proportion of sequences (29%). Despite the apparent predominance of coarse-to-fine processing, certain flexibility in the processing sequence of spatial scale information has emerged, and it has also been seen to be sensitive to the demands of the task or the visual characteristics available in the stimuli (Parker et al., 1996; Schyns and Oliva, 1997, 1999; Morrison and Schyns, 2001;

Mermillod et al., 2005; Ozgen et al., 2005, 2006; Rotshtein et al., 2010; Awasthi et al., 2013). A study by Schyns and Oliva (1999) showed that it was possible to constrain the spatial frequency band preferentially processed in hybrids by imposing a sensitization phase which implicitly "primes" visual processing in favor of a particular scale (coarse or fine). When participants were initially exposed to LSF information, subsequent categorization of hybrid images was preferentially performed following LSF cues, whereas it was biased toward HSF information after priming by HSF. The use of hybrid faces allowed Schyns and Oliva (1999) to show preferential recourse to HSF information to determine whether a face was expressive or not, and preferential recourse to LSF information to determine the nature of the emotion (e.g., happy, angry). It is therefore possible that the demands of a visual task determine which scale must be processed in hybrids (even using very short presentation). Overall, these studies suggest that all spatial frequencies are available at the beginning of categorization, and that their selection may depend on interactions between the perceptual information available and the demands of a given visual task.

Importantly, results from Schyns and Oliva (1994) studies suggest that coarse-to-fine processing constitutes a predominant and default strategy that seems advantageous for scene recognition (in the absence of task demands which constrain the use of a particular spatial frequency band). A recent study also evidenced a coarse-to-fine preference in the very early stages of development, in 7- to 8-months-old infants (Otsuka et al., 2014). Furthermore, a considerable number of recent studies have provided behavioral evidence of an LSF-based processing during rapid scene recognition (Kihara and Takeda, 2010; De Cesarei and Loftus, 2011; Musel et al., 2012; Mu and Li, 2013) and object categorization (Loftus and Harley, 2004). Using dynamic scenes composed of six filtered images of the same scene, from LSF to HSF or from HSF to LSF, allowing to experimentally mimic a coarse-to-fine or a reverse fine-to-coarse sequence, Musel et al. (2012) showed that coarse-to-fine sequences were categorized more rapidly than fine-to-coarse sequences in young adults. This provided new arguments in favor of a predominantly coarse-to-fine categorization of natural scenes, and a new experimental tool which imposes a coarse-to-fine processing and allows investigations of the neural substrates of coarse-to-fine processing.

NEURAL BASIS OF COARSE-TO-FINE ANALYSIS

We do not as yet know exactly how and where in the brain LSF and HSF information is differentially analyzed and eventually merged during visual processing. Traditional models generally maintain that incoming visual cues are combined at successive stages along the cortical hierarchy (Biederman, 1995; Riesenhuber and Poggio, 1999), and suggest that LSF and HSF converge only in higher-level visual areas of the inferior temporal cortex (such as the fusiform or parahippocampal cortex; Bar et al., 2006). However, drawing on evidence obtained from neurophysiological recordings in nonhuman primates (Hupe et al., 2001), Bullier (2001) postulated that a rapid LSF analysis takes place predominantly in the dorsal cortical stream. Information is then sent-back through feedback signals into low-level areas (e.g., the primary visual cortex, V1), where it influences subsequent slower HSF analysis and guides subsequent processing through the ventral cortical stream. The occipital cortex might therefore serve as an "active blackboard" integrating computations made by higher-order cortical areas.

Bar et al. (2006) later investigated the neural correlates and time course of spatial frequency processing during object recognition in a combined fMRI and MEG study. They found evidence that stimuli containing LSF information elicited rapid activation in the orbitofrontal cortex, 50 ms before the involvement of recognition-related areas in the temporal cortex (fusiform gyrus). Activation of the orbitofrontal cortex was not observed with stimuli containing only HSF information. These authors suggested that the orbitofrontal cortex – mediated by LSF information – acts as the trigger of top-down facilitation during object recognition. Using dynamic causal modeling to investigate the interaction between the orbitofrontal cortex and the fusiform gyrus during the perception of LSF and HSF objects, Kveraga et al. (2007) showed reciprocal connections between these two cortical structures, with LSF modulating feedback connections from the orbitofrontal cortex to the fusiform gyrus. LSF may therefore reach the orbitofrontal cortex rapidly, in order to trigger plausible interpretations of any given visual input. The result of these computations would then be projected, via feedback connections, to the fusiform gyrus, and would guide subsequent analysis of HSF information. It is worth noting that in a recent study, Patai et al. (2013) presented LSF or HSF scenes as memory-cues (i.e., contextual information) and then asked participants to detect a target (e.g., an object) in the non-filtered

version of the cued scene. These authors evidenced that LSF and HSF memory-cues were equally effective as triggers of contextual memory information, and facilitated target detection. This challenges Bar's proposal of LSF-based facilitation in object recognition. However, their target detection task may have involved fine-grained perception, thus favoring HSF processing.

However, to date, the neural architecture and temporal dynamics of such top-down mechanisms have never been systematically investigated via direct testing of the preferential coarse-to-fine processing sequence during visual scene perception in humans. Peyrin et al. (2010) combined fMRI and ERPs on the same participants to identify the neural substrates underlying the coarse-to-fine processing sequence. To constrain the order of spatial frequency processing, the authors displayed sequences of two spatial frequency-filtered scenes in rapid succession, with either a coarse-to-fine sequence (LSF scene followed by a HSF scene), or a fine-to-coarse sequence (HSF scene followed by an LSF scene). Participants' task was to decide whether the two scenes belonged to a same category (city, beach, or indoor). FMRI examination revealed selective increased activation in early stage occipital areas, and in frontal and tempo-parietal areas for coarse-to-fine sequences (compared to fine-to-coarse sequences). ERP topography and source analyses revealed a similar cortical network, but could additionally determine the time-course of activation in these areas. Frontal and temporo-parietal areas responded more to LSF scenes when these were presented first, whereas the occipital areas responded more to HSF scenes when these were presented after LSF scenes. More specifically, results demonstrated that low-pass signals (conveyed by fast magnocellular pathways) could rapidly activate high-order areas, providing semantic information (via the left prefrontal cortex and temporal areas) and spatial information (via the frontal eye fields), as well as attentional controls (via the temporo-parietal junction), all of which may promote the ongoing categorization and perceptual organization of the scene. This low-pass or coarse analysis is perhaps refined by further processing of high-pass signals (conveyed more slowly by the parvocellular pathways). To enable this, feedback from the low-pass analysis, which take place in frontal and temporo-parietal areas, might be sent back into lower level visual areas, such as the primary visual cortex, and would then guide the high-pass analysis and assist in the selection of the relevant finer details necessary for the recognition and categorization of scenes. These results are consistent with the LSF-based top-down facilitation of recognition, as proposed by Bar et al. (2006; see also Bar, 2003) in the context of object recognition, with the exception of the cortical site for feedback projections (occipital cortex in Peyrin et al., 2010; fusiform gyrus in Bar et al., 2006).

The influential models of visual perception assume a predominantly coarse-to-fine sequence of spatial frequency processing in the whole brain, based on the functional properties of the visual pathways. However, as mentioned previously, many studies have also shown that it is possible that the two hemispheres of the human brain may complement one another in the processing of LSF and HSF. The critical issue here is how to reconcile hemispheric specialization of spatial frequency processing with coarse-to-fine analysis of scenes.

CEREBRAL ASYMMETRIES FOR COARSE-TO-FINE PROCESSING

Thehemisphericspecializationobservedforspatialfrequencyprocessing raise the crucial question of the legitimacy of suggesting that coarse-to-fine sequencing is applied throughout brain. Peyrin et al. (2005) conducted an fMRI experiment in order to investigate whether coarse-to-fine processing predominates in only one hemisphere. They displayed sequences of two spatial frequency filtered scenes in rapid succession, with either a coarse-to-fine sequence (LSF scene followed by HSF scene), or a fine-to-coarse sequence (HSF scene followed by LSF scene). Participants' task was to decide whether the two scenes belonged to a same category (city, beach, or indoor). Cerebral asymmetries were identified using inter-hemispheric method of comparison (i.e., contrast between "unflipped" and "left-right flipped" functional images for each sequence). Results showed greater activation in the right than the left occipito-temporal cortex for the coarse-to-fine sequence, and greater activation in the left than the right occipito-temporal cortex for the fine-to-coarse sequence. These fMRI results suggest that the initial spatial frequency-band appearing in the sequence could determine which of the two hemispheres is preferentially involved in the sequential processing of spatial frequencies. According to input sequences or task demands, the right occipital cortex would give priority to LSF analysis for a coarse-to-fine processing and the left occipital cortex would give priority to HSF analysis for a fine-to-coarse analysis.

As far as the higher-level stages of visual scene processing are concerned, several studies have highlighted the sensitivity of scene-selective areas to low-level features, such as spatial frequencies and

amplitude spectrum properties, in scenes (Andrews et al., 2010; Rajimehr et al., 2011; Zeidman et al., 2012). However, we still lack evidence of coarse-to-fine processing within the scene-selective cortical regions.

SPATIAL FREQUENCY PROCESSING WITHIN SCENE-SELECTIVE AREAS

There is considerable evidence suggesting that the occipitotemporal cortex contains a mosaic of different areas that respond selectively to different category of stimuli (Haxby et al., 2001; Lerner et al., 2001; Spiridon and Kanwisher, 2002). More specifically, three regions were evidenced as scene-selective regions: the PPA, the RSC, and the OPA. These regions are known to be involved in high-level functions such as navigation (Epstein et al., 2007; Vass and Epstein, 2013), spatial layout processing and scene recognition (Epstein and Kanwisher, 1998; Epstein et al., 1999, 2003; Epstein, 2005, 2008; Epstein and Higgins, 2007; Epstein and Ward, 2010; Dilks et al., 2013), and contextual associations (Bar and Aminoff, 2003; Bar, 2004; Aminoff et al., 2007; Bar et al., 2008a,b). However, only a few studies investigated whether these regions are sensitive to scenes low-level properties such as spatial frequencies. For example, Peyrin et al. (2004) showed that the parahipopcampal gyrus was more strongly activated by LSF than HSF scenes. Conversely, Rajimehr et al. (2011) observed that in human and macaques, the PPA responded more strongly to HSF than LSF stimuli. This was also the main findings of Zeidman et al. (2012). In their study, they depicted three-dimensional spaces by positioning small dots following an exponential distribution and filtered them in either LSF or HSF. They showed stronger activation of the PPA when participants had to detect de disappearance of a small proportion of dots in HSF than LSF spaces. It should be noted that these studies differed in many methodological aspects such as the task demands or the duration of stimuli, that may have influenced spatial frequency selectivity within the PPA. However, whether coarse-to-fine processing of scenes occurs within scene-selective regions is still unclear.

Coarse-to-fine processing of faces in high level visual cortex was the central focus of a recent study by Goffaux et al. (2011) who showed an intriguing effect of spatial frequencies in a face-selective region, the fusiform face area (FFA; Kanwisher et al., 1997). By manipulating duration of exposure and the spatial frequency content of faces, these authors observed higher levels of FFA response to LSF when duration of exposure to faces was short, and higher levels of response to HSF for longer exposure durations. These results suggest that coarse-to-fine processing is the predominant strategy in the most prominent regions of the ventral visual stream (inferotemporal cortex). In an evoked potential study, Schettino et al. (2011) used sequences of filtered scenes (with blank screens occurring between scenes) in order to investigate the neural correlates of the accumulation of visual information during object recognition and the time course of these correlates. For this purpose, the authors used sequences in which the first scene was always in LSF and the scene was gradually revealed in six successive images by progressively adding HSF information. The authors observed that activation in the parahippocampal cortex decreases when the spatial frequency content of scenes increases, suggesting that this region is sensitive to the primary processing of LSF information, even if this study did not investigate explicit coarse-to-fine processing of scenes.

A recent fMRI study (Musel et al., 2014) tested whether such processing occurs in three scene-selective cortical regions: the PPA, the RSC, and the OPA. We measured activation in these scene-preferring regions during the categorization of dynamic scene stimuli (Musel et al., 2012). Dynamic scenes were composed of six filtered images of the same scene, from LSF to HSF or from HSF to LSF, allowing us to mimic either a coarse-to-fine or a fine-to-coarse sequence. We first identified scene-selective regions using a localizer adapted from previous studies (Epstein and Kanwisher, 1998; Epstein et al., 2003; Bar et al., 2008b; Walther et al., 2009) in which participants viewed gray-scale photographs of scenes, faces and common objects. The contrast between scenes and other categories was intended to enable localization of the regions involved in the perception of scenes. Once localized, we compared activation elicited by coarse-to-fine and fine-to-coarse dynamic scenes within the areas defined as the PPA, RSC, and OPA. Results showed greater activation of only the PPA for coarse-to-fine compared to fine-to-coarse sequences (**Figure 5**). Equivalent activations were observed for both types of sequence in the RSC and OPA. This study therefore suggests that coarse-to-fine sequence processing constitutes the predominant strategy for scene categorization in the PPA. It should be noted that evidence of spatial frequency sensitivity within other scene-selective areas, such as the RSC and the OPA, is still lacking.

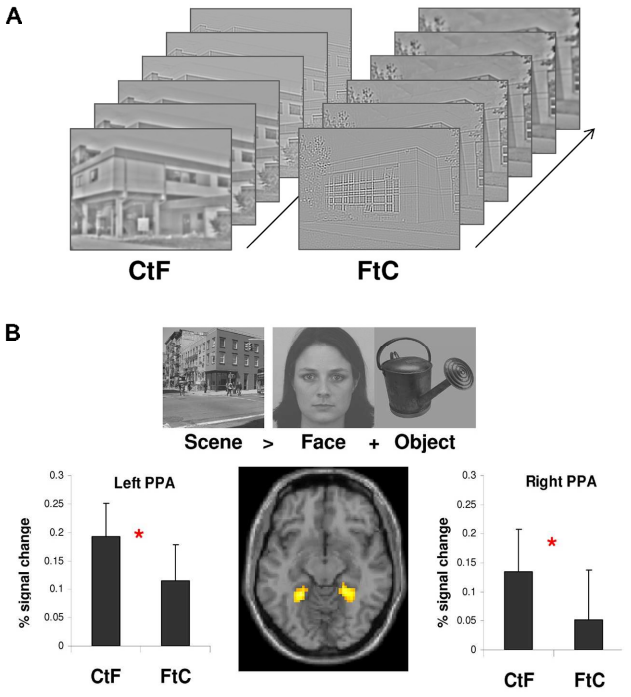


FIGURE 5 | (A) Six spatial frequency filtered images of scenes that depict a coarse-to-fine (CtF) and fine-to-coarse (FtC) sequences. **(B)** Parahippocampal place area (PPA) localized by contrasting the activation induced by the perception of scenes to those induced by the perception of faces and objects. Signal changes relative to the global mean intensity of signal were then extracted from the PPAs for each sequence (CtF and FtC). PPA showed stronger activation during the CtF than FtC categorization of sequences. Error bars indicate 95% Confidence Intervals. * Indicate significant differences. Figure adapted from Musel et al. (2014).

CONCLUSION

The present review aimed to identify cerebral regions differentially involved in low and high spatial frequency processing and to clarify their attributes during scene perception. Several neuroimaging studies suggest that spatial frequency processing could be retinotopically mapped and lateralized in both hemispheres. Right occipital areas are more activated than the left ones during the processing of LSF scenes, while left occipital areas are more activated than the right ones during the processing of HSF scenes. Concomitantly, the processing of HSF scenes (as opposed to LSF) activates the foveal representation in retinotopic areas of the occipital cortex, and LSF scenes (as opposed to HSF) activate more peripheral representations in retinotopic areas.

The present review also studied the neural bases of coarse-to-fine analysis as a default and predominant processing strategy. According to influential models (Bullier, 2001; Bar, 2003; Bar et al., 2006; Kveraga et al., 2007; Peyrin et al., 2010), LSF information may reach high-order areas rapidly, enabling coarse initial parsing of the visual scene, which can then be sent back through feedback connections into lower level visual areas to guide a finer analysis based on HSF. Studies also indicate that in scene perception, coarse-to-fine processing seems to be preferentially performed in the right hemisphere, from the occipital to the inferior temporal cortex. Overall, results from neuroimaging studies are consistent with the idea that explicit vision advances in a reverse hierarchical direction, as hypothesized by Hochstein and Ahissar (2002) and Ahissar and Hochstein (2004; see The Reverse Hierarchy Theory). According to this theory, rapid visual perception is not purely feedforward, it is also strongly mediated by top-down influences by high-level areas on lower-level areas. Finally, the present review addressed spatial frequency processing within scene-selective cortical areas. We reported results demonstrated that the coarse-to-fine strategy is a plausible modus operandi in the PPA.

Overall, these results obviously raised the question of the connectivity between the PPA and the cortical network specifically involved in coarse-to-fine processing. Baldassano et al. (2013) recently demonstrated that the PPA exhibits a gradient in connectivity with other scene-specific regions along the anterior-posterior axis in a way that suggests that the posterior part of the PPA is more closely connected to occipital areas and therefore contributes more to the processing of low level visual features (possibly to spatial frequencies and spatial envelope properties) while the anterior part of the PPA is more closely connected to the RSC and therefore contributes to the construction of a global scene representation. In Musel et al. (2014), the contrast between coarse-to-fine and fine-to-coarse processing revealed significant activation within the orbitofrontal cortex and the primary visual cortex (**Figure 6**). These two regions might play a predominant role during the coarse-to-fine categorization of scenes. Involvement of the orbitofrontal cortex was previously evidenced in rapid LSF-based categorical inferences (Bar et al., 2006; Peyrin et al., 2010) and the primary visual cortex was evidenced to be one of the cortical sites in which the first LSF computation could be "retro-injected" to guide the subsequent finer analysis of HSF (Bullier, 2001; Peyrin et al., 2010). In a proactive brain model, Bar (2007) attempts to clarify the functional role of the parahippocampal cortex (including the PPA) in object recognition. According to this model, LSF information in an object is projected from early stage visual areas to

the orbitofrontal cortex. Based on the global appearance of the object, this region then triggers activation of the most probable object identities. Parallel projection of LSF information to the parahippocampal cortex and the PPA also occurs to extract the context in which this object appears and activates its contextual associations. The intersection of possible object identities (from the orbitofrontal cortex) and the objects that typically appear in such contexts (from the parahippocampal cortex) provides fast and coarse recognition of the current view of the object. This assumption is supported by studies on the macaque brain which indicate that the orbitofrontal cortex has strong and reciprocal links with the temporal cortex, notably medial regions including parahippocampal areas (Cavada et al., 2000).

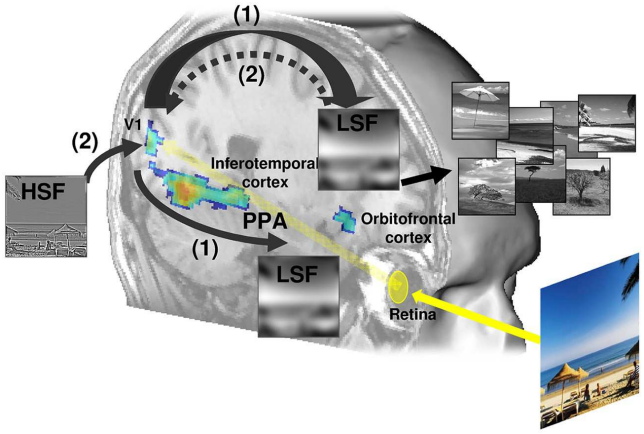


FIGURE 6 | A schematic illustration of the proposed coarse-to-fine cortical model. (1) Los spatial frequency (LSF) information reaches high-order areas of the dorsal visual stream rapidly, enabling coarse initial parsing of the visual scene (providing the spatial organization of the scene through the frontal eye fields and possible interpretations of the category of the scene through the orbitofrontal cortex), prior to its complete propagation along the ventral visual stream (inferotemporal cortex) that ultimately mediates the scene recognition. (2) This initial low-pass analysis might be then "retro-injected" through feedback into lower level areas (including the primary visual cortex, V1) to guide a slower analysis of high spatial frequency (HSF) information through the ventral visual stream and select the relevant finer details necessary for the recognition and identification. In this model, the coarse-to-fine analysis is preferentially performed in the right ventral visual stream, from the occipital to the inferior temporal cortex and the parahippocampal place area (PPA).

In humans, studies using diffusion tensor MRI have evidenced structural connectivity between the parahippocampal cortex and orbitofrontal areas (Powell et al., 2004). Bar et al. (2006) also demonstrated strong synchrony between the orbitofrontal cortex and the temporal cortex during the recognition of LSF-filtered objects, suggesting important functional interactions between these regions. Unfortunately, to our knowledge, the functional connectivity or direct influence between the orbitofrontal cortex and PPA has not been demonstrated yet.

To conclude, the results reported in the present review provide critical support for influential models of visual perception mainly based on a spatial frequency analysis which follows a coarse-to-fine strategy (Schyns and Oliva, 1994; Bar, 2003; Hegde, 2008; Peyrin et al., 2010).

ACKNOWLEDGMENTS

Our works were supported by the National Centre for Scientific Research in France, by a doctoral fellowship from the Région Rhone-Alpes to Louise Kauffman, by the RECOR ANR Grant (ANR-12-JHS2-0002-01 RECOR) and by a research grant from the Fyssen Foundation. The authors extend warm thanks to the UMS IRMaGe 017,the "Délégation à la Recherche Clinique" of the University Hospital of Grenoble and the "Cellule de Neuroimagerie Fonctionnelle" of the Laboratory of Psychology and NeuroCognition for sponsoring their works. We thank Catherine Dal Molin for the English revision of the manuscript.

REFERENCES

Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464. doi: 10.1016/j.tics.2004.08.011

Aminoff, E., Gronau, N., and Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb. Cortex* 17, 1493–1503. doi: 10.1093/cercor/bhl078

Andrews, T. J., Clarke, A., Pell, P., and Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *Neuroimage* 49, 703–711. 7:91. doi: 10.1016/j.neuroimage.2009.08.046

Arcaro, M. J., McMains, S. A., Singer, B. D., and Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *J. Neurosci.* 29, 10638–10652. doi: 10.1523/JNEUROSCI.2807-09.2009

Arcaro, M. J., Pinsk, M. A., Li, X., and Kastner, S. (2011). Visuotopic organization of macaque posterior parietal cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 31, 2064–2078. doi: 10.1523/JNEUROSCI.333410.2011

Awasthi, B., Sowman, P. F., Friedman, J., and Williams, M. A. (2013). Distinct spatial scale sensitivities for early categorization of faces and places: neuromagnetic and behavioral findings. *Front. Hum. Neurosci.* 7:91. doi: 10.3389/fnhum.2013. 00091

Badcock, J. C., Whitworth, F. A., Badcock, D. R., and Lovegrove, W. J. (1990). Low frequency filtering and the processing of local-global stimuli. *Perception* 19, 617–629. doi: 10.1068/p190617

Baldassano, C., Beck, D. M., and Fei-Fei, L. (2013). Differential connectivity within the parahippocampal place area. *Neuroimage* 75C, 236–245. doi: 10.1016/j.neuroimage.2013.02.073

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629. doi: 10.1038/nrn1476

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289. doi: 10.1016/j.tics.2007.05.005

Bar, M., and Aminoff, E. (2003). Cortical analysis of visual context. *Neuron* 38, 347–358. doi: 10.1016/S0896-6273(03)00167-3

Bar, M., Aminoff, E., and Ishai, A. (2008a). Famous faces activate contextual associations in the parahippocampal cortex. *Cereb. Cortex* 18, 1233–1238. doi: 10.1093/cercor/bhm170

Bar,M.,Aminoff,E.,and Schacter,D. L. (2008b). Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *J. Neurosci.* 28, 8539–8544. doi: 10.1523/JNEUROSCI.0987-08.2008

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454. doi: 10.1073/pnas.0507062103

Biederman, I. (1995). "Visual object recognition," in *An Invitation to Cognitive Science*, eds S. F. Kosslyn and D. N. Osherson (Cambridge, MA: MIT Press), 121–165.

Brefczynski, J.A., and DeYoe, E.A. (1999). A physiological correlate of the'spotlight' of visual attention. *Nat. Neurosci.* 2, 370–374. doi: 10.1038/7280

Breitmeyer,B.G.(1975). Simple reaction time as a measure of the temporal response properties of the transient and sustained channels. *Vision Res.* 15,1411–1412. doi: 10.1016/0042-6989(75)90200-X

Bressler, D. W., Fortenbaugh, F. C., Robertson, L. C., and Silver, M. A. (2013). Visual spatial attention enhances the amplitude of positive and negative fMRI responses to visual stimulation in an eccentricity-dependent manner. *Vision Res.* 85, 104–112. doi: 10.1016/j.visres.2013.03.009

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Rev.* 36, 96–107. doi: 10.1016/S0165-0173(01)00085-6

Cavada, C., Company, T., Tejedor, J., Cruz-Rizzolo, R. J., and Reinoso-Suarez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cereb. Cortex* 10, 220–242. doi: 10.1093/cercor/10.3.220

Chokron, S., Bartolomeo, P., Colliot, P., Brickman, A. M., Tabert, M., Wei, T., et al. (2003). Selective attention, inhibition for repeated events and hemispheric specialization. *Brain Cogn.* 53, 158–161. doi: 10.1016/S0278-2626(03) 00100-3

Cousin,E., Peyrin,C., and Baciu,M.(2006). Hemispheric predominance assessment of phonology and semantics: a divided visual field experiment. *Brain Cogn.* 61, 298–304. doi: 10.1016/j.bandc.2006.02.005

Curcio,C.A.,andAllen,K.A. (1990). Topography of ganglion cells in human retina. *J. Comp. Neurol.* 300, 5–25. doi: 10.1002/cne.903000103

Curcio, C. A., Sloan, K. R., Kalina, R. E., and Hendrickson, A. E. (1990). Human photoreceptor topography. *J. Comp. Neurol.* 292, 497–523. doi: 10.1002/cne.902920402

De Cesarei, A., and Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychon. Bull. Rev.* 18, 840–847. doi: 10.3758/s13423-0110133-6

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982a). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* 22, 545–559. doi: 10.1016/0042-6989(82)90113-4

De Valois, R. L., Yund, E. W., and Hepler, N. (1982b). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res.* 22, 531–544. doi: 10.1016/0042-6989(82)90112-2

Dilks, D. D., Julian, J. B., Paunov, A. M., and Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *J. Neurosci.* 33, 1331–1336a. doi: 10.1523/JNEUROSCI.408112.2013

Dos Santos, N. A., Andrade, S. M., and Fernandez Calvo, B. (2013). Detection of spatial frequency in brain-damaged patients: influence of hemispheric asymmetries and hemineglect. *Front. Hum. Neurosci.* 7:92. doi: 10.3389/fnhum.2013. 00092

Engel, S. A., Glover, G. H., and Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* 7, 181–192. doi: 10.1093/cercor/7.2.181

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., et al. (1994). fMRI of human visual cortex. *Nature* 369, 525. doi: 10.1038/369525a0

Epstein, R. A. (2005). The cortical basis of visual scene processing. *Vis. Cogn.* 12, 954–978. doi: 10.1080/13506280444000607

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* 12, 388–396. doi: 10.1016/j.tics.2008. 07.004

Epstein, R. A., Graham, K. S., and Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* 37, 865–876. doi: 10.1016/S0896-6273(03)00117-X

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23, 115–125. doi: 10.1016/S0896-6273(00)80758-8

Epstein, R. A., and Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb. Cortex* 17, 1680–1693. doi: 10.1093/cercor/bhl079

Epstein, R. A., Higgins, J. S., Jablonski, K., and Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *J. Neurophysiol.* 97, 3670– 3683. doi: 10.1152/jn.00003.2007

Epstein, R.A., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598–601. doi: 10.1038/33402

Epstein, R. A., and Ward, E. J. (2010). How reliable are visual context effects in the parahippocampal place area? *Cereb. cortex* 20, 294–303. doi: 10.1093/cercor/bhp099

Everson, R. M., Prashanth, A. K., Gabbay, M., Knight, B. W., Sirovich, L., and Kaplan, E. (1998). Representation of spatial frequency and orientation in the visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8334–8338. doi: 10.1073/pnas.95. 14.8334

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 2379–2394. doi: 10.1364/JOSAA.4.002379

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S., and Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature* 382, 626–628. doi: 10.1038/382626a0

Fink,G. R.,Halligan,P.W.,Marshall,J. C.,Frith,C. D.,Frackowiak,R. S.,and Dolan, R. J. (1997). Neural mechanisms involved in the processing of global and local aspects of hierarchically organized visual stimuli. *Brain* 120(Pt 10), 1779–1791. doi: 10.1093/brain/120.10.1779

Fink, R., Marshall, J. C., Halligan, P. W., and Dolan, R. J. (2000). Neuronal activity in early visual areas during global and local processing: a comment on Heinze, Hinrichs, Scholz, Burchert and Mangun. *J. Cogn. Neurosci.* 12, 355–356; discussion 357–359. doi: 10.1162/089892900562039

Foster, K. H., Gaska, J. P., Nagler, M., and Pollen, D. A. (1985). Spatial and temporal frequency selectivity of neurons in visual cortical areasV1 and V2 of the macaque monkey. *J. Physiol.* 365, 331–363.

Gandhi, S. P., Heeger, D. J., and Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3314–3319. doi: 10.1073/pnas.96.6.3314

Gegenfurtner, K. R., Kiper, D. C., and Levitt, J. B. (1997). Functional properties of neurons in macaque area V3. *J. Neurophysiol.* 77, 1906–1923.

Ginsburg, A. P. (1986).“Spatial filtering and visual form perception,” in *Hanbook of Perception and Human Performance*, ed. L. K. J. T. K. Boff (New York, NY: Wiley), 1–41.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., and Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cereb. Cortex* 21, 467–476. doi: 10.1093/cercor/bhq112

Guyader, N., Chauvin, A., Peyrin, C., Herault, J., and Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *C. R. Biol.* 327, 313–318. doi: 10.1016/j.crvi.2004.02.006

Hagler, D. J., and Sereno, M. I. (2006). Spatial maps in frontal and prefrontal cortex. *Neuroimage* 29, 567–577. doi: 10.1016/j.neuroimage.2005.08.058

Han, S., Weaver, J. A., Murray, S. O., Kang, X., Yund, E. W., and Woods, D. L. (2002). Hemispheric asymmetry in global/local processing: effects of stimulus position and spatial frequency. *Neuroimage* 17, 1290–1299. doi: 10.1006/nimg. 2002.1255

Hausmann, M. (2005). Hemispheric asymmetry in spatial attention across the menstrual cycle. *Neuropsychologia* 43, 1559–1567. doi: 10.1016/j.neuropsychologia.2005.01.017

Hausmann, M., Becker, C., Gather, U., and Güntürkün, O. (2002). Functional cerebral asymmetries during the menstrual cycle: a cross-sectional and longitudinal analysis. *Neuropsychologia* 40, 808–816. doi: 10.1016/S0028-3932(01) 00179-8

Hausmann, M., and Güntürkün, O. (2000). Steroid fluctuations modify functional cerebral asymmetries: the hypothesis of progesterone-mediated interhemispheric decoupling. *Neuropsychologia* 38, 1362–1374. doi: 10.1016/S00283932(00)00045-2

Hausmann, M., Hamm, J. P., Waldie, K. E., and Kirk, I. J. (2013). Sex hormonal modulation of interhemispheric transfer time. *Neuropsychologia* 51, 1734–1741. doi: 10.1016/j.neuropsychologia.2013.05.017

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430. doi: 10.1126/science. 1063736

Hegde, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Prog. Neurobiol.* 84, 405–439. doi: 10.1016/j.pneurobio.2007. 09.001

Heinze, H. J., Hinrichs, H., Scholz, M., Burchert, W., and Mangun, G. R. (1998). Neural mechanisms of global and local processing. A combined PET and ERP study. *J. Cogn. Neurosci.* 10, 485–498. doi: 10.1162/089892998562898

Henriksson, L., Raninen, A., Nasanen, R., Hyvarinen, L., and Vanni, S. (2007). Training-induced cortical representation of a hemianopic hemifield. *J. Neurol. Neurosurg. Psychiatry* 78, 74–81. doi: 10.1136/jnnp.2006.099374

Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804. doi: 10.1016/S08966273(02)01091-7

Holmes, G. (1918). Disturbances of vision by cerebral lesions. *Br. J. Ophthalmol.* 2, 353–384. doi: 10.1136/bjo.2.7.353

Horton, J. C., and Hoyt, W. F. (1991). The representation of the visual field in human striate cortex. A revision of the classic Holmes map. *Arch. Ophthalmol.* 109, 816–824. doi: 10.1001/archopht.1991.01080060080030

Hughes, H. C., Nozawa, G., and Kitterle, F. L. (1996). Global precedence, spatial frequency channels, and the statistic of the natural image. *J. Cogn. Neurosci.* 8, 197–230. doi: 10.1162/jocn.1996.8.3.197

Hupe, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., and Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *J. Neurophysiol.* 85, 134–145.

Iidaka, T., Yamashita, K., Kashikura, K., and Yonekura, Y. (2004). Spatial frequency of visual image modulates neural responses in the temporo-occipital lobe. An investigation with event-related fMRI. *Brain Res. Cogn. Brain Res.* 18, 196–204. doi: 10.1016/j.cogbrainres.2003.10.005

Issa, N. P., Trepel, C., and Stryker, M. P. (2000). Spatial frequency maps in cat visual cortex. *J. Neurosci.* 20, 8504–8514.

Kanwisher, N., Mcdermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.

Kaplan, E. (2004). “The M, P and K pathways of the primate visual system,” in *The Visual Neurosciences*, eds L. M. Chalupa and J. M. Werner (Cambridge, MA: MIT Press), 481–493.

Kenemans, J. L., Baas, J. M., Mangun, G. R., Lijffijt, M., and Verbaten, M. N. (2000). On the processing of spatial frequencies as revealed by evoked-potential source modeling. *Clin. Neurophysiol.* 111, 1113–1123. doi: 10.1016/S13882457(00)00270-4

Kihara, K., and Takeda, Y. (2010). Time course of the integration of spatial frequency-based information in natural scenes. *Vision Res.* 50, 2158–2162. doi: 10.1016/j.visres.2010.08.012

Kinchla, R. A., and Wolfe, J. M. (1979). The order of visual processing: “Topdown,” “bottom-up,” or “middle-out.” *Percept. Psychophys.* 25, 225–231. doi: 10.3758/BF03202991

Kitterle, F. L., Christman, S., and Hellige, J. B. (1990). Hemispheric differences are found in the identification, but not the detection, of low versus high spatial frequencies. *Percept. Psychophys.* 48, 297–306. doi: 10.3758/BF03 206680

Kitterle, F. L., Hellige, J. B., and Christman, S. (1992). Visual hemispheric asymmetries depend on which spatial frequencies are task relevant. *Brain Cogn.* 20, 308–314. doi: 10.1016/0278-2626(92)90023-F

Kitterle, F. L., and Selig, L. M. (1991). Visual field effects in the discrimination of sine-wave gratings. *Percept. Psychophys.* 50, 15–18. doi: 10.3758/BF03 212201

Kveraga, K., Boshyan, J., and Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *J. Neurosci.* 27, 13232–13240. doi: 10.1523/JNEUROSCI.3481-07.2007

Lamb, M. R., Robertson, L. C., and Knight, R. T. (1990). Component mechanisms underlying the processing of hierarchically organized patterns: inferences from patients with unilateral cortical lesions. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 471–483. doi: 10.1037/0278-7393.16.3.471

Lamb, M. R., and Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Percept. Psychophys.* 54, 773–784. doi: 10.3758/BF03211802

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., and Malach, R. (2001). A hierarchical axis of object processing stages in the

human visual cortex. *Cereb. Cortex* 11, 287–297. doi: 10.1093/cercor/11.4.287

Loftus, G. R., and Harley, E. M. (2004). How different spatial-frequency components contribute to visual information acquisition. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 104–118. doi: 10.1037/0096-1523.30.1.104

Lux, S., Marshall, J. C., Ritzl, A., Weiss, P. H., Pietrzyk, U., Shah, N. J., et al. (2004). A functional magnetic resonance imaging study of local/global processing with stimulus presentation in the peripheral visual hemifields. *Neuroscience* 124, 113–120. doi: 10.1016/j.neuroscience.2003.10.044

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S., Frith, C. D., and O'Keefe, J. (1998). Knowing where and getting there: a human navigation network. *Science* 280, 921–924. doi: 10.1126/science.280.5365.921

Mangun, G. R., Heinze, H. J., Scholz, M., and Hinrichs, H. (2000). Neural activity in early visual areas during global and local processing: a reply to Fink, Marshall, Halligan and Dolan. *J. Cogn. Neurosci.* 12, 357–359. doi: 10.1162/089892900562048

Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.* 2, 364–369. doi: 10.1038/7274

Martinez, A., Dirusso, F., Anllo-Vento, L., Sereno, M. I., Buxton, R. B., and Hillyard, S. A. (2001). Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Res.* 41, 1437–1457. doi: 10.1016/S0042-6989(00)00267-4

Martinez, A., Moses, P., Frank, L., Buxton, R., Wong, E., and Stiles, J. (1997). Hemispheric asymmetries in global and local processing: evidence from fMRI. *Neuroreport* 8, 1685–1689. doi: 10.1097/00001756-19970506000025

McGlone, J., and Kertesz, A. (1973). Sex differences in cerebral processing of visuospatial tasks. *Cortex* 9, 313–320. doi: 10.1016/S0010-9452(73)80009-7 Merigan, W. H., and Maunsell, J. H. (1993). How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* 16, 369–402. doi: 10.1146/annurev.ne.16.030193.002101

Mermillod, M., Guyader, N., and Chauvin,A. (2005). The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling. *Brain Cogn.* 57, 151– 157. doi: 10.1016/j.bandc.2004.08.035

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

Mu, T., and Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Exp. Brain Res.* 227, 367–377. doi: 10.1007/s00221-013-3517-1

Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Le Bas, J. F., et al. (2013). Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. *J. Cogn. Neurosci.* 25, 1315–1331. doi: 10.1162/jocn_a_00397

Musel, B., Chauvin, A., Guyader, N., Chokron, S., and Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PLoS ONE* 7:e38493. doi: 10.1371/journal.pone.0038493

Musel,B.,Kauffmann,L.,Ramanoel,S.,Giavarini,C.,Guyader,N., Chauvin,A.,et al. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *J. Cogn. Neurosci.* doi:10.1162/jocn_a_00643 [Epub ahead of print].

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., et al. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing. A PET study. *Brain* 123(Pt 9), 1903–1912. doi: 10.1093/brain/123.9.1903

Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *Cogn. Psychol.* 9, 353–383. doi: 10.1016/0010-0285(77) 90012-3

Oliva, A., and Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cogn. Psychol.* 34, 72–107. doi: 10.1006/cogp.1997.0667

Otsuka, Y., Ichikawa, H., Kanazawa, S., Yamaguchi, M. K., and Spehar, B. (2014). Temporal dynamics of spatial frequency processing in infants. *J. Exp. Psychol. Hum. Percept. perform.* doi: 10.1037/a0035440 [Epub ahead of print].

Ozgen, E., Payne, H. E., Sowden, P. T., and Schyns, P. G. (2006). Retinotopic sensitisation to spatial scale: evidence for flexible spatial frequency processing in scene perception. *Vision Res.* 46, 1108–1119. doi: 10.1016/j.visres.2005. 07.015

Ozgen, E., Sowden, P., Schyns, P., and Daoutis, C. (2005). Top-down attentional modulation of spatial frequency processing in scene perception. *Vis. Cogn.* 12, 925–937. doi: 10.1080/13506280444000580

Palmer, S. (1993).“Modern theories of gestalt perception,”in *Understanding Vision*, ed. G. W. Humphreys (Oxford: Blackwell), 39–70.

Parker, D. M., Lishman, J. R., and Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1448–1466. doi: 10.1037/0096-1523.22.6.1448

Patai, E. Z., Buckley, A., and Nobre, A. C. (2013). Is attention based on spatial contextual memory preferentially guided by low spatial frequency signals? *PLoS ONE* 8:e65601. doi: 10.1371/journal.pone.0065601

Peyrin, C., Baciu, M., Segebarth, C., and Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *Neuroimage* 23, 698–707. doi: 10.1016/j.neuroimage.2004.06.020

Peyrin, C., Chauvin, A., Chokron, S., and Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain Cogn.* 53, 278–282. doi: 10.1016/S0278-2626(03)00126-X

Peyrin, C., Chokron, S., Guyader, N., Gout, O., Moret, J., and Marendaz, C. (2006). Neural correlates of spatial frequency processing: a neuropsychological approach. *Brain Res.* 1073–1074, 1–10. doi: 10.1016/j.brainres.2005.12.051

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., et al. (2010). The neural substrates and timing of top-down processes during coarseto-fine categorization of visual scenes: a combined fMRI and ERP study. *J. Cogn. Neurosci.* 22, 2768–2780. doi: 10.1162/jocn.2010.21424

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., and Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *Neuroimage* 28, 464–473. doi: 10.1016/j.neuroimage.2005.06.006

Poggio,G.F.(1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Invest. Ophthalmol.* 11, 368–377.

Powell, H. W., Guye, M., Parker, G. J., Symms, M. R., Boulby, P., Koepp, M. J., et al. (2004). Noninvasive in vivo demonstration of the connections of the human parahippocampal gyrus. *Neuroimage* 22, 740–747. doi: 10.1016/j.neuroimage.2004.01.011

Pratte, M. S., and Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. *J. Vis.* 14, 1–12. doi: 10.1167/14.3.22

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., and Tootell, R. B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol*. 9:e1000608. doi: 10.1371/journal.pbio.1000608

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Robertson, L. C., and Lamb, M. R. (1991). Neuropsychological contributions to theories of part/whole organization. *Cogn. Psychol.* 23, 299–330. doi: 10.1016/0010-0285(91)90012-D

Robertson, L. C., Lamb, M. R., and Knight, R. T. (1988). Effects of lesions of temporal-parietal junction on perceptual and attentional processing in humans. *J. Neurosci.* 8, 3757–3769.

Rotshtein, P., Schofield, A., Funes, M. J., and Humphreys, G. W. (2010). Effects of spatial frequency bands on perceptual decision: It is not the stimuli but the comparison. *J. Vis.* 10, 1–20. doi: 10.1167/10.10.25

Sasaki, Y., Hadjikhani, N., Fischl, B., Liu, A. K., Marret, S., Dale, A. M., et al. (2001). Local and global attention are mapped retinotopically in human occipital cortex. *Psychology* 98, 2077–2082. doi: 10.1073/pnas.98.4.2077

Saygin, A. P., and Sereno, M. I. (2008). Retinotopy and attention in human occipital, temporal, parietal, and frontal cortex. *Cereb. Cortex* 18, 2158–2168. doi: 10.1093/cercor/bhm242

Schettino, A., Loeys, T., Delplanque, S., and Pourtois, G. (2011). Brain dynamics of upstream perceptual processes leading to visual object recognition: a high density ERP topographic mapping study. *Neuroimage* 55, 1227–1241. doi: 10.1016/j.neuroimage.2011.01.009

Schulman, G., Sulivan, M., Gisch, K., and Sadoka, W. (1986). The role of spatial frequency channels in the perception of local and global structure. *Perception* 15, 259–273. doi: 10.1068/p150259

Schyns,P.G.,andOliva,A.(1994). From blobs to boundary edges: Evidence for time and spatial-scale-dependant scene recognition. *Am. Psychol. Soc.* 5, 195–200. doi: 10.1111/j.1467-9280.1994.tb00500.x

Schyns, P. G., and Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception* 26, 1027–1038. doi: 10.1068/p261027

Schyns, P. G., and Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* 69, 243–265. doi: 10.1016/S0010-0277(98)00069-9

Sergent, J. (1982). Theoretical and methodological consequences of variations in exposure duration in visual laterality studies. *Percept. Psychophys.* 31, 451–461. doi: 10.3758/BF03204855

Sergent, J. (1983). Role of the input in visual hemispheric asymmetries. *Psychol. Bull.* 93, 481–512. doi: 10.1037/0033-2909.93.3.481

Sergent,J.,andHellige,J.B.(1986). Role of input factors in visual-field asymmetries. *Brain Cogn.* 5, 174–199. doi: 10.1016/0278-2626(86)90054-0

Sheremata, S. L., Bettencourt, K. C., and Somers, D. C. (2010). Hemispheric asymmetry in visuotopic posterior parietal cortex emerges with visual short-term memory load. *J. Neurosci.* 30, 12581–12588. doi: 10.1523/JNEUROSCI.268910.2010

Silver, M. A., Ress, D., Heeger, D. J., and Michael, A. (2005). Topographic maps of visual spatial attention in human parietal cortex. *J. Neurophysiol.* 1358–1371. doi: 10.1152/jn.01316.2004

Silver, M. A., Ress, D., Heeger, D. J., Michael, A., and Neural, D. J. H. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *J. Neurophysiol*. 229–237. doi: 10.1152/jn.00677.2006

Singh,K.D.,Smith,A.T.,andGreenlee,M.W.(2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage* 12, 550–564. doi: 10.1006/nimg.2000.0642

Spiridon,M.,andKanwisher,N.(2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35, 1157–1165. doi: 10.1016/S0896-6273(02)00877-2

Szczepanski, S. M., and Kastner, S. (2013). Shifting attentional priorities: control of spatial attention through hemispheric competition. *J. Neurosci.* 33, 5411–5421. doi: 10.1523/JNEUROSCI.4089-12.2013

Szczepanski, S. M., Konen, C. S., and Kastner, S. (2010). Mechanisms of spatial attention control in frontal and parietal cortex. *J. Neurosci.* 30, 148–160. doi: 10.1523/JNEUROSCI.3862-09.2010

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

Tolhurst, D. J., Tadmor, Y., and Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic Physiol. Opt.* 12, 229–232. doi: 10.1111/j.1475-1313.1992. tb00296.x

Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., et al. (1998). The retinotopy of visual spatial attention. *Neuron* 21, 1409–1422. doi: 10.1016/S0896-6273(00)80659-5

Tootell, R. B., Silverman, M. S., Hamilton, S. L., Switkes, E., and De Valois, R. L. (1988). Functional anatomy of macaque striate cortex. V. Spatial frequency. *J. Neurosci.* 8, 1610–1624.

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

Van Essen, D. C., and Deyoe, E. A. (1995). "Concurrent processing in the primate visual cortex," in *The Cognitive Neurosciences*, ed. M. Gazzaniga (Cambridge: Bradford Book), 383–400.

Vass, L. K., and Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *J. Neurosci.* 33, 6133–42. doi: 10.1523/JNEUROSCI.3873-12.2013

Voyer, D. (1996). On the magnitude of laterality effects and sex differences in functional lateralities. *Laterality* 1, 51–83. doi: 10.1080/7137 54209

Wandell, B. A., Dumoulin, S. O., and Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron* 56, 366–383. doi: 10.1016/j.neuron.2007. 10.012

Walther, D. B., Caddigan, E., Fei-Fei, L., and Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.* 29, 10573–10581. doi: 10.1523/JNEUROSCI.0559-09.2009

Watanabe, T., Sasaki, Y., Miyauchi, S., Putz, B., Fujimaki, N., Nielsen, M., et al. (1998). Attention-regulated activity in human primary visual cortex. *J. Neurophysiol.* 79, 2218–2221.

Weissman, D. H., and Woldorff, M. G. (2005). Hemispheric asymmetries for different components of global/local attention occur in distinct temporo-parietal loci. *Cereb. Cortex* 15, 870–876. doi: 10.1093/cercor/bhh187

Wilkinson, D. T., Halligan, P. W., Marshall, J. C., Buchel, C., and Dolan, R. J. (2001). Switching between the forest and the trees:

brain systems involved in local/global changed-level judgments. *Neuroimage* 13, 56–67. doi: 10.1006/nimg.2000.0678

Xu, X., Anderson, T. J., and Casagrande, V. A. (2007). How do functional maps in primary visual cortex vary with eccentricity? *J. Comp. Neurol.* 501, 741–755. doi: 10.1002/cne.21277

Yamaguchi, S., Yamagata, S., and Kobayashi, S. (2000). Cerebral asymmetry of the "top-down" allocation of attention to global and local features. *J. Neurosci.* 20, RC72.

Zeidman,P.,Mullally,S. L.,Schwarzkopf,D. S.,and Maguire,E.A. (2012). Exploring the parahippocampal cortex response to high and low spatial frequency spaces. *Neuroreport* 23, 503–507. doi: 10.1097/WNR.0b013e328353766a

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

Received: 16 February 2014; accepted: 19 April 2014; published online: 07 May 2014. Citation: Kauffmann L, Ramanoël S and Peyrin C (2014) The neural bases of spatial frequency processing during scene perception. Front. Integr. Neurosci. ***8****:37. doi: 10.3389/fnint.2014.00037*

This article was submitted to the journal Frontiers in Integrative Neuroscience. Copyright © 2014 Kauffmann, Ramanoël and Peyrin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CCBY). The use, distribution or reproduction in other forums is permitted, provided the original author(s)or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Article sous presse dans la revue *PlosOne*

Age-Related Differences in Spatial Frequency Processing During Scene Categorization

Stephen Ramanoël[1,2,3¶], Louise Kauffmann[1,2¶], Emilie Cousin[1,2,4¶], Michel Dojat[3,5¶], and Carole Peyrin[1,2¶*]

[1] Univ. Grenoble Alpes, LPNC, F-38000 Grenoble, France
[2] CNRS, LPNC, F-38000 Grenoble, France
[3] Univ. Grenoble Alpes, GIN, F-38000 Grenoble, France
[4] UMS IRMaGe, F-38000 Grenoble, France
[5] INSERM U836, GIN, F-38000 Grenoble, France

Abstract

Visual analysis of real-life scenes starts with the parallel extraction of different visual elementary features at different spatial frequencies. The global shape of the scene is mainly contained in low spatial frequencies (LSF), and the edges and borders of objects are mainly contained in high spatial frequencies (HSF). The present fMRI study investigates the effect of age on the spatial frequency processing in scenes. Young and elderly participants performed a categorization task (indoor vs. outdoor) on LSF and HSF scenes. Behavioral results revealed performance degradation for elderly participants only when categorizing HSF scenes. At the cortical level, young participants exhibited retinotopic organization of spatial frequency processing, characterized by medial activation in the anterior part of the occipital lobe for LSF scenes (compared to HSF), and the lateral activation in the posterior part of the occipital lobe for HSF scenes (compared to LSF). Elderly participants showed activation only in the anterior part of the occipital lobe for LSF scenes (compared to HSF), but not significant activation for HSF (compared to LSF). Furthermore, a ROI analysis revealed that the parahippocampal place area, a scene-selective region, was less activated for HSF than LSF for elderly participants only. Comparison between groups revealed greater activation of the right inferior occipital gyrus in young participants than in elderly participants for HSF. Activation of temporo-pariet al regions was greater in elderly participants irrespective of spatial frequencies. The present findings indicate a specific low-contrasted HSF deficit for normal elderly people, in association with an occipito-temporal cortex dysfunction, and a functional reorganization of the categorization of filtered scenes.

Keywords: Normal aging; Functional cerebral reorganization; Low spatial frequency; High spatial frequency; Elderly people.

Introduction

Our visual world is hierarchically organized. Visual scenes of daily life, such as outdoor or indoor environments, contain many visual objects made of smaller objects comprising multiple visual details. There is considerable evidence suggesting that the spatial frequency content of scenes is important for organized perception. Convergent data from the functional neuroanatomy of magnocellular and parvocellular visual pathways [1], neurophysiological recordings in primates (for a review, see [2]), psychophysical studies in humans [3,4], and simulations [5,6] suggest that the first stage of visual perception consists of the parallel extraction of different visual elementary features at different spatial frequencies. Low spatial frequencies (LSF), conveyed by fast magnocellular pathways, provide coarse information about the scene (e.g., the global shape and structure), whereas high spatial frequencies (HSF), conveyed more slowly by the parvocellular pathways, provide more detailed information about the scene (e.g., the edges and borders of an object). Temporal precedence of LSF over HSF processing has generally been observed in behavioral studies using sinusoidal grating [3,4,7] and complex filtered scenes as stimuli [8,9,10].

Recent neuroimaging studies have aimed to identify the cerebral regions differentially involved in LSF and HSF processing in young adults (for a recent review, see [11]). Using either sinusoidal gratings or more complex scene stimuli, recent studies revealed that spatial frequency processing is retinotopically organized in the visual cortex [12,13,14]. More precisely, LSF scene categorization activated occipital areas in relation to the peripheral representation of the visual field, whereas HSF scene categorization activated occipital areas in relation to the foveal representation. However, the way spatial frequencies are processed in normal aging remains unclear.

Normal aging is characterized by a decline in many cognitive functions, such as selective attention [15,16,17], working memory [18], processing speed [19,20] and executive control [21]. Many visual functions, such as visual acuity [22,23,24,25] and contrast sensitivity [26,27,28,29,30] also decline with age. Studies on spatial contrast sensitivity present conflicting results, but the majority have shown that, with increasing age, contrast sensitivity decreases mainly for medium and high spatial frequencies [26,27,29,30] ; see however [31]. Owsley, Sekuler and Siemsen [30] showed that contrast sensitivity for stationary LSF gratings remained the same throughout adulthood, whereas sensitivity for stationary HSF gratings decreased with age (beginning at around 40 to 50 years), suggesting more pronounced impairment of the parvocellular pathway. Hardy et al. [32] measured sensitivity to chromatic contrasts (e.g., green and red contrast) for different spatial frequencies using sinusoidal gratings in young and older participants. Results showed that the threshold of sensitivity to chromatic contrast was higher in older than in younger participants, especially for HSF gratings. Given the parvocellular pathway's sensitivity to chrominance and HSF, these results have been interpreted in favor of a functional decline of the parvocellular pathway in normal aging. Elliot and Werner [33] directly investigated age-related changes in magno- and parvocellular pathways using two paradigms thought to separate these two pathways based on their contrast gain signature [34]. Results showed a functional deficit of the two pathways with age, and this was more pronounced in the parvocellular pathway. Loss of contrast sensitivity for HSF with increasing age could, therefore, be explained by a higher sensitivity of the parvocellular pathway to normal aging.

Similarly, behavioral experiments investigating the perception of global and local visual information in normal elderly people have produced conflicting results on

the processing of spatial frequencies. Studies using hierarchical stimuli (global forms composed of several local elements; [35]) show that young adults identify the global form more rapidly than local elements (global precedence effect). Based on the assumption that global information is preferentially conveyed by LSF, and that local information is conveyed by HSF [36,37,38], the global precedence effect has been interpreted as additional evidence of the temporal precedence of LSF over HSF processing. However, a number of studies have shown that with age, the advantage of global processing tends to be reversed in favor of local processing [39,40,41]. Lux et al. [39] for example reported that reaction times of young adult participants were faster when detecting global forms, while older participants had faster reaction times when detecting local elements. These results could be interpreted as reflecting a temporal precedence of HSF over LSF processing in normal elderly people. However, in other studies, global precedence was not reduced, and became even more pronounced with increased age [42,43].

Only a few studies have investigated the effects of age on spatial frequency processing using more complex and ecological stimuli than gratings and hierarchical forms. Viggiano, Righi and Galli [44] presented sequences of nine images of objects (animals or tools). Sequences started with an LSF filtered object and HSF information was added progressively. Participants had to identify the object in each image in these sequences. Results showed that in order to identify the objects correctly, older participants needed more HSF information than young participants. Musel et al. [8] presented sequences of six images of scenes (indoors or outdoors) in which the spatial frequency content differed from one image to the other, going either from LSF to HSF or from HSF to LSF. Results showed that young participants categorized low-to-high sequences more quickly than high-to-low sequences, consistent with the temporal advantage of LSF over HSF. The LSF advantage tended to be reversed in older participants (over the age of 60). These two studies suggest that HSF are more important to older than to younger participants for visual perception. However, because it was not possible to compare HSF and LSF directly, the authors were unable to reach any conclusion on the existence of a potential deficit in spatial frequency processing related to normal aging. Musel et al. [45] tested the categorization of LSF and HSF scenes in normal elderly participants in order to establish normative data to assess the processing of spatial frequencies in age-related macular degeneration. Results showed that the performance of healthy elderly participants was not affected by the spatial frequency content of scenes, but provided no answer concerning a possible deficit in spatial frequency processing related to normal aging because no comparison with young participants was made. Overall, these behavioral studies suggest that the visual mechanisms involved in spatial frequency processing may change with age.

The small number of neuroimaging studies which have investigated age-related differences on visual processing reveal cortical anatomical changes, characterized by atrophy of white and gray matter in the primary visual cortex with age [46,47], and changes in the retinotopy of visual areas [48,49,50]. The latter may influence the way spatial frequencies are processed in the occipital cortex. However to our knowledge, no studies have as yet been conducted on the effects of age on the neural bases of spatial frequency processing.

The present study was adapted from a previous fMRI study conducted in young adults only [13] to investigate elderly adults' ability to process spatial frequencies in natural environments compared to that of young adults. For this purpose, young and elderly participants had to categorize natural indoor and outdoor scenes filtered in LSF and HSF. Exemplars from the outdoor and indoor categories were chosen in order to

have similar dominant orientations in the amplitude spectrum and to avoid categorization based on this type of visual cue. From a pragmatic point of view, this categorization task can be performed whatever the type of filtering (low-pass, high-pass, or pass-band). Furthermore, this task is simple and quick to administer, but also easy to perform even for patients with age-related degeneration (Musel et al., 2011). It should be noted that luminance contrast in scenes decreases as spatial frequency increases, following a $1/f\alpha$ function [51]. Therefore, LSF scenes are characterized by high luminance contrast, while HSF scenes are characterized by low luminance contrast. FMRI was used to investigate any neural correlates of spatial frequency changes due to age-related differences in the whole brain. Furthermore, most studies agree that the parahippocampal cortex, especially the posterior part known as the parahippocampal place area (PPA), is a region of the human cortex involved in the processing of visual scenes. Early studies [52,53,54,55,56], showed that the PPA responds more strongly to images of real-world scenes (such as cityscapes and landscapes) than to other meaningful visual stimuli (such as faces and objects). We also investigated age-related differences in spatial frequency processing in the PPA. Studying the effects of normal aging on the structural and functional properties of cortex in human is essential to distinguish cortical changes related to healthy aging from those resulting from the pathophysiology of an underlying disease [49].

Method

Participants

Twenty-four right-handed participants divided into two age groups were included in the experiment (Table 1): 12 young participants (6 males; Mean age ± SD: 22 ± 3 years; Range 18-26 years) and 12 elderly participants (8 males; Mean age ± SD: 64 ± 3 years; Range 61-71 years) with normal or corrected-to-normal vision. Visual acuity was tested prior to the experiment, using the Monoyer chart designed for testing long distance vision. Participants requiring visual correction wore the MediGoggle Adult Research Set (Cambridge Research Systems Ltd, England; http://www.crsltd.com/), interchangeable prescriptive goggles suitable for use in MR environments. Participants had no neurological or ocular disorders (such as age-related macular degeneration, glaucoma or multiple sclerosis). All participants gave their informed written consent before participating in the study, which was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans, and approved by the local ethics committee (Comité de protection des personnes Sud-Est V, ID RCB: 2011-A01551-40).

Table 1: Age and visual acuity of Young (Y) and Elderly (E) participants

Participants	Age (years)	Visual Acuity (Log MAR)	
		left eye	right eye
Y1	22	0.0	0.0
Y2	25	0.0	0.0
Y3	24	-0.1	-0.1
Y4	18	0.0	0.0
Y5	18	-0.1	-0.1
Y6	25	-0.1	0.2
Y7	23	-0.1	0.0
Y8	24	0.0	0.0
Y9	18	0.0	0.0
Y10	23	0.0	0.0
Y11	25	0.0	0.0
Y12	19	0.1	0.2

Participants	Age (years)	Visual Acuity (Log MAR)	
		left eye	right eye
E1	60	0.0	0.0
E2	61	0.3	0.2
E3	71	0.5	0.0
E4	66	0.0	0.0
E5	62	0.0	0.0
E6	61	0.3	0.2
E7	64	0.0	0.0
E8	61	0.2	0.2
E9	67	0.3	0.0
E10	64	-0.1	-0.1
E11	68	0.0	0.0
E12	65	0.2	0.2

Stimuli and Procedure in the Spatial Frequency Experiment

Stimuli consisted of 20 black and white photographs (256-level grey-scales, 1042x768 pixels) of scenes classified into two distinct categories (10 indoor scenes and 10 outdoor scenes) with a visual angle of 24x18 degrees, ensuring to stimulate both the fovea and the peripheral visual field [13]. Outdoor scenes are views of houses or buildings with sky at the top and outdoor-relevant objects (e.g., car, tree). Indoor scenes are kitchens, offices and living rooms with indoor-relevant objects (e.g., table, sofa, chair). Scenes were displayed in their original version and in their mirrored version (left and right were reversed) in order to avoid any effect of the visual asymmetry of these large scene images. Exemplars from the two categories (outdoor and indoor) were chosen in order to have similar dominant orientations in the amplitude spectrum and to avoid categorization based on this type of visual cue.

To ensure that the chosen scenes have similar amplitude spectra, we first calculated the mean amplitude spectrum for the 10 indoor scenes (mean AS indoor) and the 10 outdoor scenes (mean AS outdoor). Then, for each scene, we calculated two 2D correlation coefficients, one between the scene's amplitude spectrum and the mean AS indoor and the other one between the scene's amplitude spectrum and the mean AS outdoor. The mean AS of the category corresponding to the scene of interest was calculated by excluding the scene's amplitude spectrum (i.e., for an indoor scene, the mean AS indoor was calculated based on the 9 remaining indoor scenes, whereas the mean AS outdoor was calculated based on the 10 outdoor scenes). The 2D correlation coefficient was calculated using the Matlab function "corr2d." A 2 × 2 variance analyses (ANOVA) with the category of the scene (indoor and outdoor) and the category of the mean AS (indoor and outdoor) as within-subject factors were conducted on the 2D correlation coefficients. Results show that the 2D correlation coefficients calculated between indoor scenes and the mean AS indoor did not significantly differ from those calculated between indoor scenes and the mean AS outdoor (0.76 ± 0.05 and 0.76 ± 0.05, respectively; $F_{1,18} < 1$). Similarly, the 2D correlation coefficients calculated between outdoor scenes and the mean AS outdoor did not significantly differ from those calculated between outdoor scenes and the mean AS indoor (0.78 ± 0.05 and 0.78 ± 0.04, respectively; $F_{1,18} < 1$). Outdoor and indoor categories were equivalent in terms of visual cluttering (Subband Entropy measures; see [57]). Mean subband entropy was equivalent for outdoors and indoors (2.91 ± 0.15 and 2.91 ± 0.15, respectively; $F_{1,18} < 1$).

Stimuli were elaborated using the image processing toolbox on MATLAB (Mathworks Inc., Sherborn, MA, USA). Each scene was filtered with three low-pass filters, with cutoff frequencies corresponding to 0.5, 1, and 2 cycles per degree (cpd; i.e. 12, 24, 49 cycles per image) and three high-pass filters with cutoff frequencies corresponding to 3, 6, and 12 cpd (i.e. 71, 144, 293 cycles per image), or left unfiltered (NF). Cut-off frequencies followed a logarithmic scale in order to obtain a better sampling of the amplitude spectrum of natural scenes (see [58] for a similar procedure). Furthermore, these values were chosen in order to include cut-off frequencies of 2 cpd for LSF and 6 cpd for HSF, as used in Schyns & Oliva's pioneer study on spatial frequency processing during scene perception [10]. The resulting images were then normalized to obtain a mean luminance equal to 128 on a grey-level scale ranging from 0 to 256. This resulted in 7 versions of each scene (1 NF, 3 LSF and 3 HSF, see Figure 1). Stimuli were displayed using E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, USA) and back-projected onto a translucent screen positioned at the rear of the magnet. Participants viewed this screen at a distance of about 222 cm via a mirror fixed on the head coil. We used a backward mask, built

with 1/f white noise, to prevent retinal persistence of the scene.

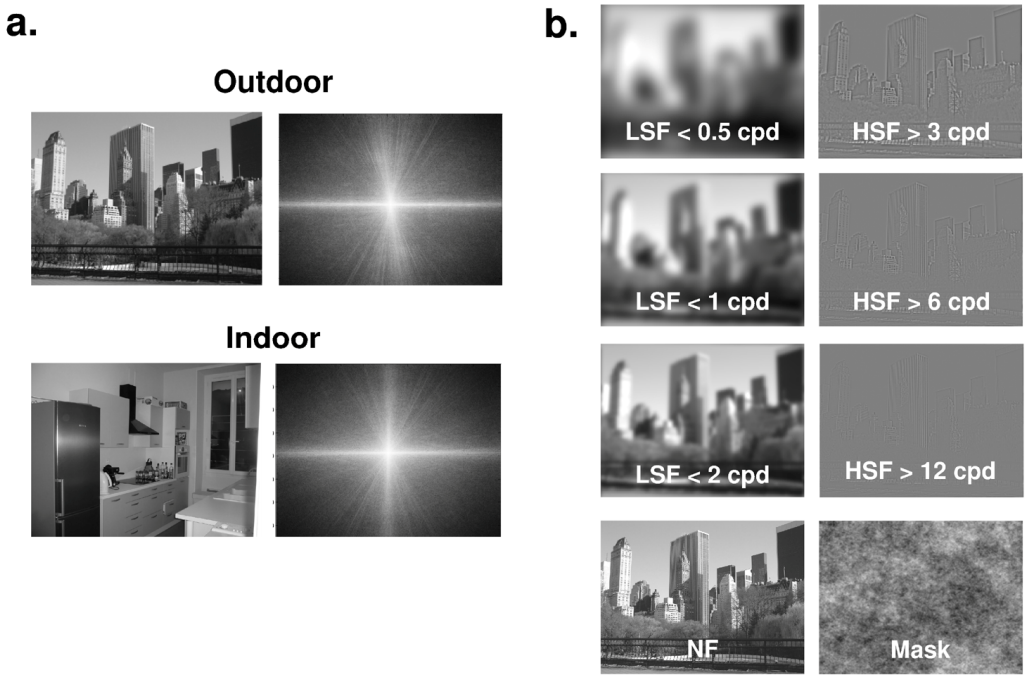


Figure 1: (a) Examples of scenes belonging to two different categories (outdoors and indoors) with the mean amplitude spectrum of each category. On the amplitude spectrum images, low spatial frequencies are close to the center and high spatial frequencies are on the periphery. Vertical orientations are represented on the x-axis and horizontal orientations on the y-axis. (b) Example of non-filtered scenes (NF), low-spatial frequency scenes (LSF) below 0.5, 1, and 2 cycles per degree (cpd), and high-spatial frequency scenes (HSF) above 3, 6, and 12 cpd. The mask used is also presented (bottom). It should be noted that the perception of spatial frequencies could be affected by picture reduction of scenes for illustrative purposes. In the experiment, picture size was about 10 times bigger than in this figure.

We used a block-design paradigm with the NF, LSF, and HSF scenes. The Spatial Frequency experiment consisted of four functional runs. Each functional scan lasted five minutes and was composed of nine 25-second task blocks (one block per spatial frequency cut-off and three blocks of NF scenes), including 10 scenes (5 indoors and 5 outdoors), interspersed with three 25-second blocks with a fixation dot in the center of the screen (Fixation condition) displayed against a grey background. Each scene was presented in all spatial frequency conditions. The order of images was randomized within blocks. It should be noted that the block-design paradigm did not allow us to analyze individual responses to trials in the way an event-related paradigm would have, and it did not allow us to investigate the neural correlates of semantic category effect and priming effect in our study. Each stimulus was displayed for 100 ms, followed by a mask for 30 ms and a fixation dot in the center of the screen. The interval between the onsets of two successive stimuli was 2.5 seconds. Participants had to give a categorical answer on the scenes (“indoors” or “outdoors”) by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. They were instructed to fixate on the center of the screen (fixation dot) during the entire run and to respond as quickly and as accurately as possible by pressing one of two response buttons. Half of the participants had to answer “indoors” with their forefinger and “outdoors” with their middle finger, while the other half had

to answer 'indoors" with their middle finger and "outdoors" with their forefinger. Response accuracy and reaction times (RT, in milliseconds) were recorded. If participants were not able to categorize the scene, they were allowed to not give a response. An error could be either a non-response or a false categorization. Analyses were conducted on mean non-response error rate (mNR), mean false categorization rate (mFC) and mean correct reaction times (mRT).

Stimuli and Procedure in the Localizer experiment of scene-selective areas

Following the main Spatial Frequency experiment, we performed a separate functional Localizer experiment (adapted from [59,60]) to localize the PPA. Participants viewed grayscale photographs of scenes, faces, and common objects in separate blocks of a block design paradigm. Scene pictures used in the localizer experiment were not shown during the spatial frequency experiment. Stimuli were black and white photographs (256 greyscales), all sized 700 x 700 pixels (or 16.4 x 16.4 degrees of visual angle). The localizer experiment consisted of one functional run. The functional run lasted 3 minutes and was composed of eight 15-s task blocks (four blocks of scenes, two blocks of faces, and two blocks of objects), including 15 different images of the same type, interspersed with four 15-s blocks with a fixation dot in the center of the screen displayed against a gray background. Participants performed a "one-back" repetition detection task. They were instructed to press a button whenever they saw two identical stimuli repeated. Only two repetitions per block were presented. Each stimulus was presented for 300 ms, with a 700 ms interstimulus interval with a fixation dot in the center of the screen. For each participant, the PPA was identified in both hemispheres by a [Scenes > Faces + Objects] contrast.

fMRI Acquisition

Experiments were performed using a whole-body 3T Philips scanner (Achieva 3.0T TX Philips - Philips Medical Systems, Best, NL) with a 32-channel head coil at the Grenoble MRI facility IRMaGe in France. For all functional Spatial Frequency scans, the manufacturer-provided gradient-echo/T2* weighted EPI method was used. Forty-four adjacent axial slices parallel to the bi-commissural plane were acquired in sequential mode. Slice thickness was 3 mm. The in-plane voxel size was 2.5×2.5 mm (220×220 mm field of view acquired with a 88×85 pixel data matrix; reconstructed with zero filling to 96×96 pixels). The main sequence parameters were: TR = 2.5 s, TE = 30 ms, flip angle = 80°. Finally, a T1-weighted high-resolution three-dimensional anatomical volume was acquired, by using a 3D Modified Driven Equilibrium Fourier Transform (MDEFT) sequence (field of view = 256×224×175 mm; resolution: 1.33×1.70s×1.37 mm; acquisition matrix: 192×132×128 pixels; reconstruction matrix: 288×288×128 pixels).

Data Analysis

Data analysis was performed using the general linear model [61] for block designs in SPM12b (Wellcome Department of Imaging Neuroscience, London, UK, www.fil.ion.ucl.ac.uk/spm/) implemented in MATLAB 7 (Mathworks Inc., Sherborn, MA, USA). Functional volumes were realigned to correct for head movements to the mean functional image using a rigid body transformation. The T1-weighted anatomical volume was then realigned (affine transformation) to match the mean functional image of each participant, and was then normalized (non rigid, non linear transformation) into the MNI space. A default 4th degree B-Spline interpolation was applied. The anatomical normalization parameters were subsequently used for the normalization of functional volumes. Finally, each functional volume was smoothed by an 8-mm FWHM (Full Width

at Half Maximum) Gaussian kernel. Times-series for each voxel were high-pass filtered (1/128 Hz cutoff) to remove low-frequency noise and signal drift. We used an efficient non-linear deformation algorithm [62] with a high level number of degrees of freedom to realign all individual brains, and by visual inspection, we did not notice any realignment problem in some parts of the brain between elderly and young anatomical scans.

For the Spatial Frequency experiment, eight conditions of interest (LSF-0.5cpd, LSF-1cpd, LSF-2cpd, HSF-3cpd, HSF-6cpd, HSF-12cpd, NF, and Fixation) were modeled as eight regressors constructed as box-car functions convolved with a canonical hemodynamic response function for each participant. Accuracy for each trial (either correct responses or error) and movement parameters derived from realignment corrections (three translations and three rotations) were also entered in the design matrix as additional factors of no interest to account for no-response related variance and head motion, respectively. On an individual level, we first identified the brain regions involved in the processing of each spatial frequency content relative to the fixation ([LSF > fixation], [HSF > fixation], and [NF > fixation]), irrespective of the spatial frequency cut-off. We also tested the effect of spatial frequency cut-off for each spatial frequency band relative to the fixation ([LSF-05cpd > fixation], [LSF-1cpd > fixation], [LSF-2cpd > fixation], [HSF-3cpd > fixation], [HSF-6cpd > fixation], [HSF-12cpd > fixation]). To allow population inference, three ANOVAs were performed based on individual analysis by means of a flexible-factorial design following the guidelines by Glascher and Gitelman [63]. The first ANOVA aimed to test the interaction between groups and the processing of spatial frequencies. We modeled groups (young and aged) and spatial frequency bands (LSF, HSF, and NF) as factors. This ANOVA allowed us to investigate the processing of spatial frequencies in each group separately, and to test differences between young and aged participants on each spatial frequency band. The second ANOVA aimed to assess the effect of LSF cut-off on within- and between-group differences. We modeled groups (young and aged) and LSF cut-offs (LSF-0.5cpd, LSF-1cpd, and LSF-2cpd) as factors. The third ANOVA aimed to assess the effect of HSF cut-off on within- and between-group differences. We modeled groups (young and aged) and HSF cut-offs (HSF-3cpd, HSF-6cpd, and HSF-12cpd) as factors. Areas of activation were considered significant if they exceeded a voxel and cluster threshold of $p < 0.05$ family-wise error (FWE) corrected for multiple comparisons, with a minimum cluster extent of 5 voxels ($T < 5.18$ for the first ANOVA and $T < 5.23$ for the second and third ANOVA). To facilitate comparisons with other studies, a transformation of MNI into Talairach and Tournoux [64] coordinates was performed using the MNI2TAL function (created by Matthew Brett, available at www.mrc-cbu.cam.ac.uk/Imaging).

For the Localizer experiment, the fMRI signal in the localizer run was analyzed using single-participant general linear model. For each participant, four conditions of interest (scenes, faces, objects, and fixation) were modeled as four regressors, constructed as box-car functions convolved with a canonical hemodynamic response function. Movement parameters derived from realignment corrections (three translations and three rotations) were also entered into the design matric as additional factors of no interest. Left and right PPA were defined independently for each participant as regions of interest (ROIs) using the [Scenes > Faces + Objects] contrast. Significant voxel clusters on individual t maps were identified using a false-discovery correction at $qFDR < .05$ to control for the overall false-positive rate. As the clusters were generally large and involved several regions of interest, small sphere ROIs (3 mm radius) were created at individual peaks of activation in each scene-

selective region and in each hemisphere. These clusters were selected as ROIs for the data analysis in the Spatial Frequency experiment. Parameter estimates (% signal change relative to the global mean intensity of signal) of block responses were extracted from the two sphere ROIs for each participant. The average parameter of activity was calculated for the NF, LSF, and HSF. These values were submitted to an ANOVA for mixed designs with Group (Young and Elderly) as between-subject factor, and Hemisphere (Left and Right) and Spatial Frequency (NF, LSF, and HSF) as within-subject factors. In order to examine whether activity in PPA was influenced by the spatial frequency cut-offs for the LSF and HSF scenes, we additionally performed two separate ANOVAs for mixed designs with Groups (young and aged participants) as between-subject factor, and Hemisphere (Left and Right) and Spatial frequency cut-off (either 0.5, 1, and 2 cpd for LSF scenes or 3, 6, and 12 cpd for HSF scenes) as within-subject factors.

Behavioral Results

Three 2x3x2 analyses of variance (ANOVA) with Group (young and elderly participants) as between-subjects factor, and Spatial frequency (NF, LSF, and HSF) and Category (indoors and outdoors) as within-subjects factors were conducted on mNR, mFC and mRT.

The ANOVA on mNR (Figure 2a) revealed that elderly participants responded less often than young participants (Mean ± SD: 17.5 ± 10.9% and 1.9 ± 3.4%, respectively; $F_{1,22} = 26.54$, $p < 0.001$). The expected Group x Spatial frequency interaction was significant ($F_{2,44} = 37.35$, $p < 0.001$). Planned comparison showed that elderly participants responded less often when categorizing HSF scenes (49.2 ± 26.9%) than NF scenes (0.3 ± 0.8%; $F_{1,22} = 80.79$, $p < 0.001$) and LSF scenes (2.8 ± 5.0%; $F_{1,22} = 85.01$, $p < 0.001$), and when categorizing LSF than NF scenes ($F_{1,22} = 7.52$, $p < 0.05$). For young participants, there was no effect of spatial frequencies (NF: 1.2 ± 3.6%; LSF: 1.3 ± 2.8%; HSF: 3.3 ± 3.90%; all $F_{1,22} < 1$). In addition, elderly participants responded less often than young participants only when categorizing HSF scenes ($F_{1,22} = 34.67$, $p < 0.001$; NF: $F_{1,22} < 1$; LSF; $F_{1,22} = 1.01$, $p = 0.33$). The Group x Spatial frequency x Category interaction was significant ($F_{2,44} = 5.76$, $p < 0.01$). Planned comparison revealed a significant Spatial frequency x Category interaction only for elderly participants ($F_{2,44} = 4.77$, $p < 0.05$; young participants: $F_{2,44} = 1.80$, $p = 0.18$) due to the fact that they only responded less often for categorizing indoor than outdoor scenes filtered in HSF (51.1 ± 27.2% and 47.4 ± 26.6%, respectively; $F_{1,22} = 11.03$, $p < 0.01$; NF: $F_{1,22} = 1.11$, $p = 0.305$; LSF: $F_{1,22} = 3.13$, $p = 0.09$).

The ANOVA on mFC (Figure 2b) did not show a main effect of Groups ($F_{1,22} < 1$), interaction between Groups and Spatial frequencies ($F_{2,44} < 1$) or interaction between Groups, Spatial frequencies and Categories ($F_{2,44} < 1$). There was only a main effect of Spatial frequency ($F_{2,44} = 18.23$, $p < 0.001$). All participants made more false categorizations when categorizing HSF scenes (10.7 ± 10.9%) than NF scenes (2.4 ± 5.0%; $F_{1,22} = 22.04$, $p < 0.001$) and LSF scenes (3.7 ± 6.1%; $F_{1,22} = 15.48$, $p < 0.001$), and when categorizing LSF than NF scenes ($F_{1,22} = 6.39$, $p < 0.05$).

Finally, the ANOVA on mRT (Figure 2c) revealed a main effect of Groups ($F_{1,22} = 7.05$, $p < 0.05$). Elderly participants were slower to categorize scenes than healthy participants (699 ± 92 and 614 ± 81 ms, respectively). The expected Group x Spatial frequency interaction was not significant ($F_{2,44} = 1.43$, $p = 0.25$), and the Group x Spatial frequency x Category interaction was not significant ($F_{2,44} < 1$). There was a main effect of Spatial frequency ($F_{2,44} = 77.07$, $p < 0.001$). All participants had longer RT when categorizing HSF scenes (741 ± 101 ms) than NF scenes (603 ± 79 ms; $F_{1,22} = 86.90$, $p < 0.001$) and LSF scenes (626 ± 81 ms; $F_{1,22} = 69.83$, $p < 0.001$), and when categorizing LSF than NF scenes ($F_{1,22} = 31.65$, $p < 0.001$).

The relationship between visual acuity and performance was statistically assessed for each spatial frequency condition by using Pearson correlation tests between each group's performance (mNR, mFC, and mRT) and visual acuity. Results for young participants showed no correlations between LSF and Visual acuity (mNR: $r = -0.19$, $p = 0.56$; mFC: $r = -0.05$, $p = 0.89$; mRT: $r = 0.29$, $p = 0.36$) or HSF and Visual acuity (mNR: $r = -0.26$, $p = 0.42$; mFC: $r = -0.01$, $p = 0.99$; mRT: $r = 0.17$, $p = 0.59$). For elderly participants, there was a significant correlation between LSF and Visual acuity on the mFC ($r = 0.60$, $p < 0.05$). The lower participants' visual acuity, the higher were their false categorization rates. However, descriptive analysis on single participants data showed that the mFC of the elderly participant with the worst visual acuity (E6 in Table 1) was the higher (26%), well above the group average (4 ± 7%), and suggest that the significant correlation could be due to this outlier. When removing this participant, results did not show significant correlation between LSF and Visual acuity ($r = 0.24$, $p = 0.48$). No correlation was observed between LSF and Visual acuity for the mNR ($r = -0.23$, $p = 0.47$) and the mRT ($r = -0.09$, $p = 0.78$) and between HSF and Visual acuity for either measure (mNR: $r = 0.35$, $p = 0.26$; mFC: $r = -0.07$, $p = 0.82$; mRT: $r = 0.03$, $p = 0.92$).

We also tested whether the spatial frequency cut-off for the LSF and HSF conditions could influence mNR. Separate ANOVAs were conducted for the LSF and HSF stimuli with Group (young and elderly participants) as between-subjects factor, and Spatial frequency cut-off (either 0.5, 1, and 2 cpd for LSF scenes or 3, 6, and 12 cpd for HSF scenes) and Category (indoors and outdoors) as within-subjects factors. For the LSF scenes, there was neither Group x Spatial frequency cut-off interaction ($F_{2,44} = 2.53$, $p = 0.09$), nor Group x Spatial frequency cut-off x Category interaction ($F_{2,44} = 1.31$, $p = 0.28$). There was only a main effect of Spatial frequency cut-off ($F_{2,44} = 3.77$, $p < 0.05$). However, planned comparisons did not show that participants responded less often as the cut-off for the LSF information decreased (0.5 cpd: 4.4 ± 8.5%; 1 cpd: 0.7 ± 2.0%; 2 cpd: 1.0 ± 2.5%; 0.5 vs. 1 cpd: $F_{1,22} = 3.72$, $p = 0.07$; 0.5 vs. 2 cpd: $F_{1,22} = 4.07$, $p = 0.07$; 1 vs. 2 cpd: $F_{1,22} < 1$). For the HSF scenes, there was a significant Group x Spatial frequency cut-off interaction ($F_{2,44} = 27.74$, $p < 0.001$). Planned comparisons revealed that elderly participants responded less often as the cut-off for the HSF information increased (3 cpd: 14.6 ± 20.5%; 6 cpd: 57.7 ± 35.4%; 12 cpd: 75.4 ± 35.3%; 3 vs. 6 cpd: $F_{1,22} = 62.70$, $p < 0.001$; 3 vs. 12 cpd: $F_{1,22} = 81.38$, $p < 0.001$; 6 vs. 12 cpd: $F_{1,22} = 15.70$, $p < 0.001$), but no effect of cut-off was observed for young participants (3 cpd: 2.1 ± 3.8%; 6 cpd: 1.3 ± 2.4%; 12 cpd: 6.5 ± 8.7%; all $F_{1,22} < 1$). In addition, elderly participants responded less often than young participants for all spatial frequency cut-off of HSF scenes (3 cpd: $F_{1,22} = 4.43$, $p < 0.05$; 6 cpd: $F_{1,22} = 31.05$, $p < 0.001$; 12 cpd: $F_{1,22} = 44.88$, $p < 0.001$). The Group x Spatial frequency cut-off x Category interaction was not significant ($F_{1,44} < 1$).

It should be noted that luminance contrast is higher for LSF than for HSF. A control behavioral study was conducted in the same participants to verify that the specific HSF deficit in elderly participants did not result from a lower contrast in HSF scenes (see Supplementary material). We have equalized the luminance contrast of filtered stimuli using the RMS (root mean square) contrast normalization. The RMS contrast is the most used normalization. It corresponds to the standard deviation of the luminance values and has been evidenced to be the most reliable indicator of the visibility of broadband filtered images [65]. In this control experiment, scenes were normalized to obtain a RMS contrast of 0.1 (i.e. 25.6 on a gray-level scale). This contrast normalization reduces contrast in LSF while enhancing HSF contrast. Results showed that elderly participants responded less often and made more false categorization when categorizing HSF than NF and LSF scenes

even when contrast between spatial frequencies was equalized, whereas no differences were observed between spatial frequencies in young participants. They also responded less often and made more false categorization than young participants only for categorizing HSF scenes. Results revealed also that categorization of HSF scenes by elderly participants was improved by the RMS contrast normalization. However, it should be noted that RMS contrast normalization induces severe modifications in the amplitude spectrum properties of scenes that may bias behavioral and neurobiological responses.

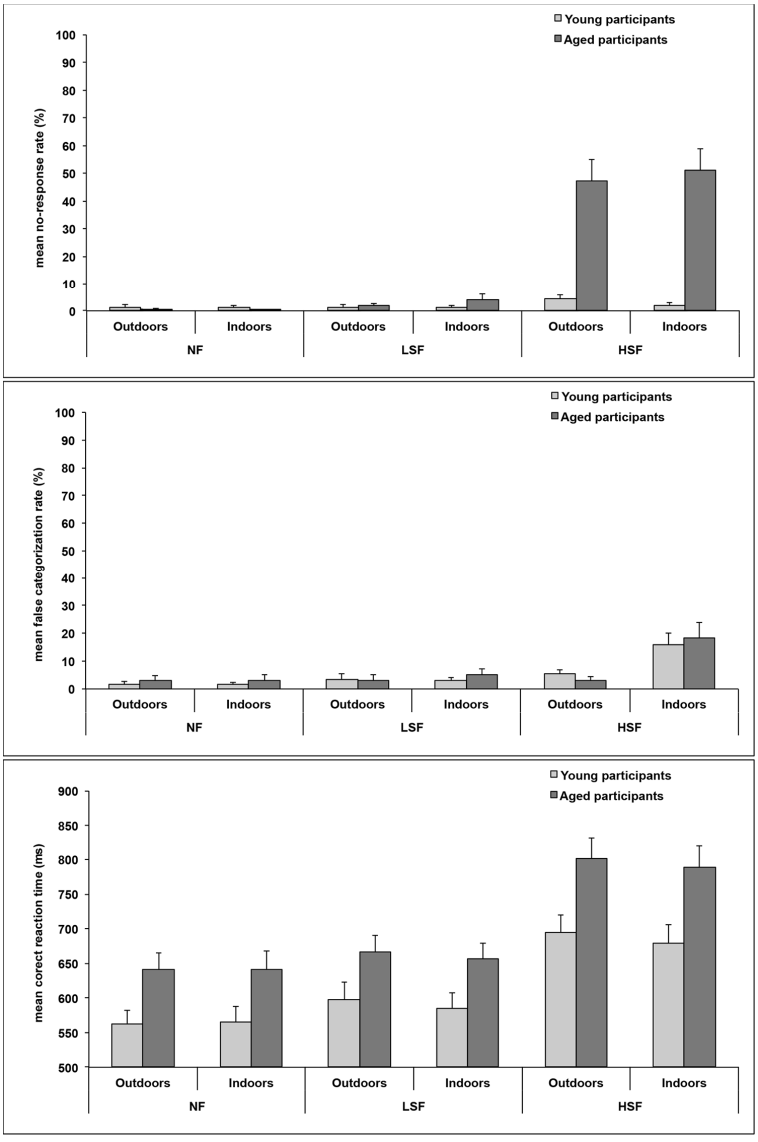


Figure 2: Mean no-response error rates (%), mean false categorization rates (%) and mean correct reaction times (in milliseconds) for the categorization of non-filtered scenes (NF), low-spatial frequency scenes (LSF), and high-spatial frequency scenes (HSF). Error bars correspond to the standard error.

Brain Activation Results

Whole Brain Analysis

The interaction between Groups (young and elderly participants) and Spatial frequencies (LSF, HSF and NF) was associated with isolated activation in the right inferior occipital gyrus (peak coordinate: 20x, -82y, -7z, F = 19.51; Figure 4). The nature of this interaction was explored firstly by calculating the contrasts that would allow us to compare the processing of different spatial frequencies for each group (within-group analysis), and then the contrasts that directly compared groups for each spatial frequency content (between-group analysis).

Within-group Analysis:

Results for the within-group analysis are shown in Table 2 and Figure 3. We began by contrasting the processing of natural scenes filtered in LSF to HSF ([LSF > HSF] contrast) by young participants, and observed extensive bilateral recruitment of the occipital cortex (in the cuneus), the posterior cingulate gyrus (involving the retrosplenial cortex in the right hemisphere), the left middle temporal gyrus, the left superior and inferior pariet al lobule areas, and the right postcentral gyrus. The opposite contrast ([HSF > LSF] contrast) showed that HSF scenes activated the occipital cortex bilaterally and the left inferior temporal gyrus. Critically, in the occipital cortex, results showed that LSF processing specifically activated the medial aspect of the occipital lobe, in the anterior half of the calcarine fissures (peak coordinates: 0x, -69y, 27z; Figure 3 and Table 2a). The reverse [HSF > LSF] contrast elicited significant, rather more posterior bilateral activation in the cuneus (right hemisphere: 20x, -82y, -7z; left hemisphere: -22x, -89y, 4z; Figure 3 and Table 2a). It should be noted that in the right hemisphere, the activation overlapped with the one associated with the interaction between groups and spatial frequencies. Contrasts relative to NF scenes revealed no significant activation.

The [LSF > HSF] contrast for elderly participants showed that LSF processing specifically activated the posterior cingulate gyri bilaterally (in the retrosplenial cortex, extending to the anterior part of the cuneus) and the left superior temporal gyrus. Greater activation in the left motor area, the right cerebellum, and the anterior cingulate gyrus was also observed for the categorization of LSF than for HSF scenes. The opposite contrast [HSF > LSF] contrast revealed no significant activation considering the statistical threshold used in the present study ($p < 0.05$ FWE corrected for multiple comparisons. It should be noted that we observed an extensive recruitment of the right middle frontal gyrus (peak coordinates: 43x, 38y, 22z) for the processing of HSF at a more lenient threshold ($p < 0.001$ uncorrected). Once again contrasts relative to NF scenes revealed no significant activation. However, similarly to the [LSF > HSF] contrast, the [NF > HSF] contrast revealed a bilateral activation of the posterior cingulate gyri (BA 31, peak coordinates: 3x, -57y, 27z) extending to the anterior part of the cuneus at a more lenient threshold of $p < 0.001$ uncorrected. This result is not surprising given that in NF scenes, the whole frequency spectrum was present (LSF and HSF information). Therefore, the [NF > HSF] contrast in which activation from HSF scenes was subtracted from NF scenes indirectly highlight activation related to LSF processing.

Table 2. Cerebral regions specifically activated during the categorization of scenes for (a) young and (b) elderly participants. The statistical significance threshold for individual voxels was set at $P < 0.05$ FWE corrected for multiple comparisons ($T > 5.18$). Only the contrasts revealing significant activation are reported. For each cluster, the region showing the maximum T value is listed first, followed by the other regions belonging to the cluster [between brackets]. The Talairach coordinates (x, y, z) of the peak and the spatial extent (k = number of voxels in the cluster) are indicated.

Contrast	Area	Side	BA	k	x	y	z	T
(a) Young participants								
[LSF > HSF]	Cuneus	L/R	17/18	11	0	-73	30	5.49
	Posterior cingulate gyrus	L	31	27	-17	-37	34	6.45
		R	31	13	5	-49	26	5.87
	Middle temporal gyrus	L	39	13	-44	-68	36	6.00
	Superior pariet al lobule	L	7	17	-24	-33	62	6.15
	Inferior pariet al lobule/Supramarginal gyrus	L	40	6	-59	-28	23	5.86
	Postcentral gyrus	R	1/2/3	65	60	-1	13	6.50
[HSF > LSF]	Inferior occipital gyrus	R	18	20	20	-82	-7	6.51
	Middle occipital gyrus	R	19	7	33	-81	17	5.59
	Cuneus	L	17	6	-14	-87	-2	5.50
	Inferior temporal gyrus	L	37	10	-44	-58	-3	5.76
	Middle frontal gyrus	R	9	84	43	38	22	7.57
	Insula	R	-	27	38	22	-4	5.91
[HSF-6 > HFS-3]	Middle frontal gyrus	R	9	17	45	35	19	6.13
	Inferior frontal gyrus/Insula	R	47	16	33	22	-2	5.68
[HSF-12 > HFS-3]	Inferior frontal gyrus/Insula	R	47	123	35	25	1	6.92
	Middle frontal gyrus	R	9	27	45	35	19	6.59
	Anterior cingulate gyrus	L/R	32	19	-7	24	37	5.96
(b) Elderly participants								
[LSF > HSF]	Posterior cingulate gyrus	R	31	42	10	-59	16	5.87
	[Cuneus]	R	19		8	-71	28	5.57
	Posterior cingulate gyrus	L	31	6	-17	-62	16	5.50
	Superior temporal gyrus	L	22	5	-47	-18	19	5.56
	Postcentral gyrus	L	1/2/3	498	-42	-28	59	8.52
	[Precentral gyrus]		4		-39	-11	58	8.15
	Medial frontal gyrus/SMA	L	6	30	-22	-4	66	5.95
		R	6	18	18	-4	57	6.71
	Anterior cingulate gyrus	L/R	32	13	5	3	43	6.15
	Cerebellum	R	-	125	15	-51	-16	9.93

Abbreviations: R = right hemisphere; L = left hemisphere; BA = Brodmann area.

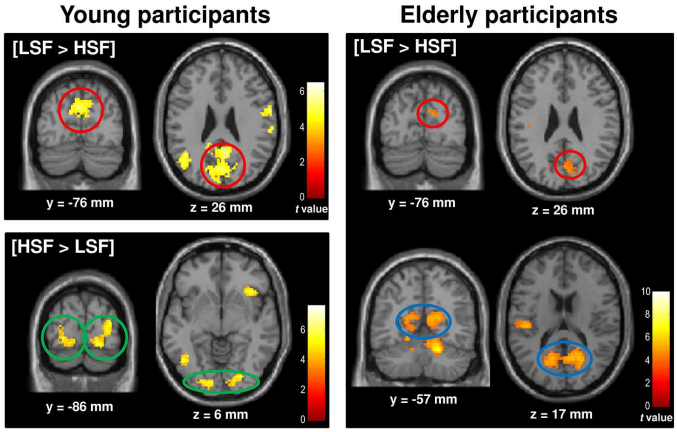


Figure 3: Cerebral regions activated by contrasting natural scenes filtered in LSF and HSF to fixation periods, and to each other in young and elderly participants. A retinotopic organization for spatial frequency processing was observed in young participants: medial activation of the anterior part of the cuneus for LSF (red circles), extending to the retrosplenial cortex (blue circles), and more lateral activation of the posterior part of the cuneus for HFS (green circles). In elderly participants, LSF scene categorization elicited bilateral activation of the retrosplenial (blue circles) cortex, extending to the anterior part of the cuneus (red circles). For illustrative purposes, statistical maps were generated with P < .0001 uncorrected.

Between-group Analysis:

Results for the between-group analysis are shown in Table 3 and Figure 4. The between group analysis revealed greater activation of the right inferior occipital gyrus (peak coordinates: 20x, -82y, -7z) in young participants ([Young > Elderly] contrast) for the categorization of HSF scenes. This activation overlapped with the one associated with the interaction between groups and spatial frequencies. No greater activation was observed for the categorization of LSF and NF scenes. In contrast, greater activation in the left temporal areas (the middle temporal gyrus) and pariet al areas bilaterally (the inferior pariet al lobules), as well as in the right superior frontal gyrus was observed in elderly participants ([Elderly > Young] contrast) for the categorization of both LSF and HSF scenes. A similar cerebral network was activated during the categorization of HSF, with the exception of the middle temporal gyrus which showed greater activation only in the left hemisphere, the primary motor cortex where activation was greater only in the right hemisphere and the putamen and cerebellum where no significant activation appeared. No greater activation was observed for the categorization of NF scenes.

Table 3. Brain activity comparison between young and elderly participants for spatial frequency processing (LSF and HSF). The statistical significance threshold for individual voxels was set at P < 0.05 FWE corrected for multiple comparisons (T > 5.23). The Talairach coordinates (x, y, z) of the peak and the spatial extent (k = number of voxels in the cluster) are indicated.

	LSF x, y, z	HSF x, y, z
[Young > Elderly]		
Right inferior occipital gyrus (BA 18)	-	20, -82, -7 (12)
[Elderly > Young]		
Left middle temporal gyrus (BA 37)	-49, -60, 5 (38)	-51, -58, 5 (31)
Left inferior pariet al lobule (BA 40)	-51, -29, 36 (6)	-54, -35, 26 (52)
Right inferior pariet al lobule/supramarginal gyrus (BA 40)	55, -29, 36 (7)	57, -32, 28 (106)
Right middle frontal gyrus (BA 9)	33, 41, 33 (8)	33, 43, 27 (27)

Abbreviations: BA = Brodmann area.

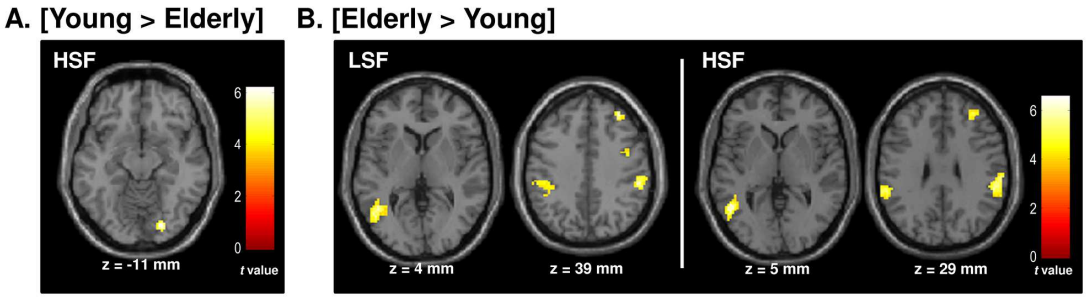


Figure 4: Cerebral regions differentially activated by (a) young and (b) elderly participants during the categorization of low-spatial frequency (LSF) and high-spatial frequency (HSF) scenes. For illustrative purposes, statistical maps were generated with P < .0001 uncorrected.

Effect of Spatial Frequency Cut-off:

The interaction between Groups (young and elderly participants) and LSF cut-offs (0.5, 1, and 2 cpd) was not significant. The interaction between Groups (young and elderly participants) and HSF cut-offs (3, 6, and 12 cpd) was associated with isolated activation in the anterior cingulate gyrus (peak coordinate: 0x, 28y, 25z, F = 18.80). We calculated the contrasts which enabled us to investigate the effect of the spatial frequency cut-off for each group (within-group analysis), and then the contrasts that compared groups directly for each spatial frequency cut-off (between-group analysis) in order to test the main effect of the interaction. The within-group analysis showed significant results only for young participants (Table 1). Activation in the cingulate gyrus was significantly greater for 12 cpd than for 3 cpd scenes ([HSF-12cpd > HSF-3cpd] contrast; 0x, 28y, 25z, F = 18.80). Furthermore, activation in the right inferior frontal gyrus (BA 47) and right middle frontal gyrus (BA 9) was significantly greater for both [HSF-12cpd > HSF-3cpd] and [HSF-6cpd > HSF-3cpd] contrasts. The between-group analysis revealed no significant results.

ROI analysis

The PPA ROIs were defined in each individual, based on the independent Localizer experiment. This served as the structural constraint for the analysis of the data in the Spatial Frequency experiment. No recordings were obtained in one young participant due to technical problems. The PPA was localized for all other participants (11 young and 12 elderly participants) in the two hemispheres (see Figure 5a for an illustration on representative young and elderly participants) based on the ([Scenes > Faces + Objects] contrast). Parameter estimates (% signal change relative to the global mean intensity of signal) of block responses were extracted from the two sphere ROIs for each participant. The average parameter of activity was calculated for each experimental condition. These values were submitted to a first ANOVA for mixed designs with Groups (young and aged participants) as between-subject factor and Hemisphere (Left and Right) and Spatial frequency (NF, LSF, and HSF) as within-subject factors. The three spatial frequency cut-offs used for LSF and HSF scenes were grouped together for this ANOVA. There was no main effect of Groups ($F_{1,21} < 1$) but there was a main effect of Spatial frequencies ($F_{2,42} = 15.83$, $p < 0.05$) and a significant Spatial Frequency × Group interaction ($F_{2,42} = 5.86$, $p < 0.05$; Figure 5b). Planned comparisons showed that for young participants, there was no effect of Spatial frequencies (NF vs. HSF: $F_{1,21} = 1.86$, $p = 0.19$; LSF vs. HSF: $F_{1,21} < 1$; NF vs. LSF: $F_{1,21} < 1$), whereas for elderly participants, HSF scenes elicited less activation than did NF and LSF scenes (NF vs. HSF: $F_{1,21} = 27.52$, $p < 0.05$; LSF vs. HSF: $F_{1,21} = 26.06$, $p < 0.05$; NF vs. LSF:

$F_{1,21} < 1$). Hemispheres did not interact with either Groups ($F_{1,21} = 1.42$, $p = 0.25$), Spatial frequencies ($F_{2,42} = 1.85$, $p = 0.17$), or their interaction ($F_{2,42} = 2.77$, $p = 0.07$).

Then, we tested whether the spatial frequency cut-off for the LSF and HSF scenes could influence the activation observed within PPA. Separate ANOVAs for mixed designs were conducted for the LSF and HSF stimuli with Groups (young and aged participants) as between-subject factor, and Hemisphere (Left and Right) and Spatial frequency cut-off (either 0.5, 1, and 2 cpd for LSF scenes or 3, 6, and 12 cpd for HSF scenes) as within-subject factors. For LSF, there was not effect of spatial frequency cut-off neither for young participants (0.5 vs. 1 cpd: $F_{1,21} = 2.33$, $p = 0.14$; 0.5 vs. 2 cpd: $F_{1,21} < 1$; 1 vs. 2 cpd: $F_{1,21} = 1.68$, $p = 0.21$) nor for elderly participants (all $F_{1,21} < 1$). For HSF, there was not effect of spatial frequency cut-off for young participants (all $F_{1,21} < 1$), whereas HSF-6cpd and HSF-12cpd scenes elicited less activation than HSF-3cpd scenes for elderly participants (3 vs. 6 cpd: $F_{1,21} = 5.74$, $p < 0.05$; 3 vs. 12 cpd: $F_{1,21} = 13.98$, $p < 0.05$; 6 vs. 12 cpd: $F_{1,21} = 2.62$, $p = 0.12$).

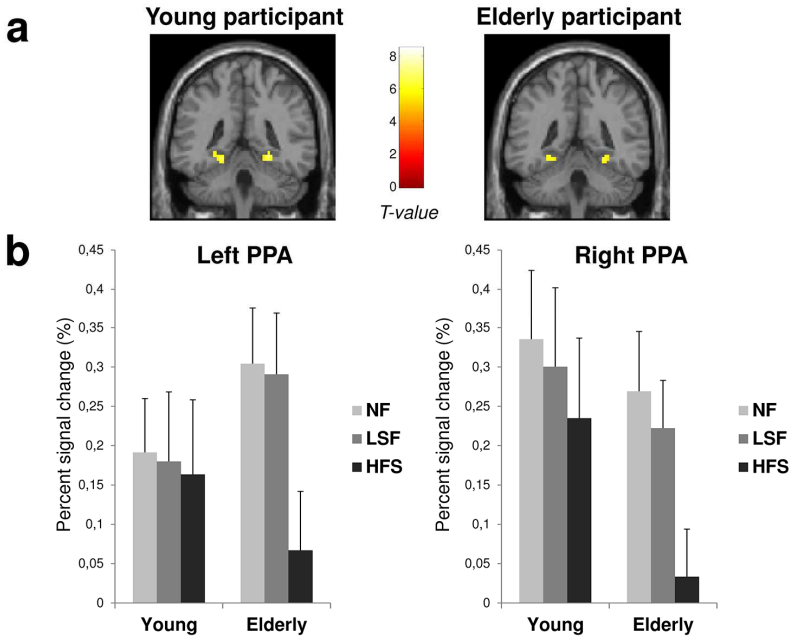


Figure 5: (a) Left and right parahippocampal place area (PPA) activated during the perception of scenes compared to faces and objects ([Scenes > Faces + Objects] contrast) for a young participant and an elderly participant. The ROIs were defined independently for each participant by contrasting scenes to other stimuli: [Scenes > Faces + Objects]. (b) Signal changes were then extracted from the scene-selective ROIs for each participant and each experimental condition (NF, LSF, and HSF). Graphics represent the mean percentage of signal change of the 11 young and 12 elderly participants for each ROI. Error bars correspond to the standard errors.

Discussion

The present fMRI study investigated for the first time the effects of age-related differences on spatial frequency processing during the categorization of visual scenes. Overall, our results suggest that differences exist in spatial frequency processing in young and elderly participants, and are characterized by a deficit in low-contrasted HSF processing in older participants compared to younger participants.

Behavioral HSF deficit in normal elderly people

Results showed that on a behavioral level, elderly participants gave more non-responses when categorizing HSF than NF and LSF scenes, irrespective of scene category, whereas no difference was observed between spatial frequencies in young participants. Elderly participants also made more non-responses than young participants during the categorization of HSF scenes. Complementary analyses revealed that as the spatial frequency cut-off of HSF scenes increased, elderly participants made more non-responses whereas no effect of cut-off was observed in young participants. These results suggest the presence of a selective impairment in elderly participants when processing HSF information. Musel et al. [45] previously investigated the categorization of filtered scenes in healthy elderly participants in order to establish normative data to assess the processing of spatial frequencies in age-related macular degeneration. They did not observe any selective HSF processing impairment due to age-related differences. However, their paradigm was constructed in order to avoid any bias in spatial frequency processing for healthy elderly controls. Presentation time, for example, was longer (300 ms vs. 100 ms in the present study) favoring HSF processing [10], and the spatial frequency cut-off of HSF scenes was very low (1 cpd vs. 3, 6, and 12 cpd in the present study) including thus middle spatial frequencies. Our complementary analyses conducted on spatial frequency cut-offs revealed that elderly participants are in fact sensitive to the amount of HSF information in a scene, and that performances on HSF processing are improved by a decrease in the cut-off value.

Concerning rapidity of categorization, elderly participants categorized scenes less rapidly than young participants. More specifically, their performance was poorer than that of younger participants when categorizing scenes filtered in HSF, in support to the impairment of HSF processing observed in response accuracy in elderly participants. Furthermore, both young and elderly participants categorized LSF scenes more rapidly than HSF scenes, in favor of a temporal processing precedence of LSF over HSF. For young participants, this result is consistent with previous behavioral studies which used either hierarchal stimuli and showed a global precedence effect [35] or filtered scenes as stimuli and showed temporal precedence of LSF over HSF [8,9,10]. However, for elderly participants, this result is not consistent with previous studies conducted in normal elderly people which used hierarchical forms as stimuli [39,41]. These studies showed either a local precedence effect or a reduced global precedence effect in older participants, and suggest the existence of a selective impairment in the processing of LSF, which convey global information, associated with normal aging. However, it should be noted that the relationship between spatial frequencies and local and global information in hierarchical stimuli is far from univocal [66]. Global information can, for example, be conveyed by not only LSF but also by HSF. When attempting to understand spatial frequency processing, any inferences based on hierarchical stimuli should therefore be considered with caution. Our results were also not consistent with previous behavioral studies which used filtered scenes as stimuli. Using sequences of filtered scenes simulating either a LSF to HSF processing or a reverse HSF to LSF processing as stimuli, Musel et al. [8] showed that elderly

participants categorized outdoor scenes faster when LSF were presented before HSF, but they categorized indoor scenes faster when HSF was presented before LSF information. These results suggest a either LSF or HSF temporal precedence in normal elderly people depending on the perceptual properties of categories, which is not consistent with the LSF temporal precedence observed irrespective of categories in the present study.

The discrepancies between our results and previous findings on hierarchical stimuli and scenes could be explained by the difficulty of the visual task. Critically, in all the studies mentioned above, error rates were very low even for elderly participants. This suggests that the tasks used were not suited to the investigation of visual deficits, but rather to the investigation of the temporal aspects of the visual processing. Interestingly, results showed that elderly participants were always slower than young participants. Musel et al. [8] previously hypothesized that elderly participants privileged high accuracy to the detriment of speed using additional processes that lengthened reaction times. In fact, they would process further semantic information to categorize scenes. Such additional processes would have changed the temporal precedence of spatial frequency processing depending on the categories. For example, they would preferentially use global spatial invariants in LSF (such as the ground, the sky, and the direction of natural light) for outdoor categorization, leading to a temporal precedence of LSF over HSF processing, and local elements in HSF (e.g., table, sofa, chair) for indoor categorization, leading to a temporal precedence of HSF over LSF processing. The preferred HSF information for categorizing indoor scenes in elderly people may explained why, in the present study, the HSF deficit in elderly participants was more pronounced for indoor than outdoor scenes. Indeed, elderly participants responded less often for categorizing indoors than outdoors only when scenes were filtered in HSF. This result suggests that the HSF deficit increased when the categorization task is preferentially performed on the basis of HSF information.

To summarize, our behavioral results point to a selective deficit in the categorization of HSF scenes in elderly people, which is consistent with the loss of contrast sensitivity to HSF previously observed in studies using sinusoidal gratings [26,27,29,30]. It should be noted that the results from the control behavioral experiment, in which the luminance contrast between NF, LSF and HSF scenes was equalized, revealed that performance of elderly participants was than that of younger participants when categorizing scenes in HSF. However, their performance was improved by the enhanced contrast in HSF scenes, suggesting that the visual deficit of scene perception in elderly participants was also partially driven by the low luminance contrast in HSF scenes. Further research is needed to clarify the extent to which luminance contrast influences spatial frequency processing during scene perception in elderly people. For example, this could be done by systematically manipulating different levels of luminance contrast as a function of different spatial frequencies.

Neurobiological correlates of the HSF deficit

The neural correlates of spatial frequency processing were investigated under fMRI. We compared activation elicited by LSF and HSF scenes in young and elderly participants separately (within-group analysis), and then we compared groups for each type of spatial frequency content (between-group analysis). The within-group analysis revealed a retinotopic organization of spatial frequency processing in young participants. The categorization of LSF scenes (relative to HSF) activated the medial aspect of the occipital lobe in the anterior half of calcarine fissure linked to the peripheral visual fields, in accordance with the retinotopic properties of early visual areas. The categorization of HSF scenes

(compared to LSF) elicited activation in the posterior part of the occipital lobes, which are linked to the fovea. Similar results were observed previously in fMRI studies conducted on young adults using either sinusoidal gratings [12,14] or filtered scenes [13]. Retinotopic processing of spatial frequencies could be explained by the distribution and neurophysiological properties of photoreceptor and ganglion cells in the human retina [67,68]. The density of cones and midget ganglion cells, which are used to process HSF information, is greatest in the fovea. Because the fovea is represented in the posterior parts of the visual areas, HSF information may be predominantly processed in these areas. In contrast, the density of rods and parasol ganglion cells, which are used to process LSF information, increases with foveal eccentricity. Because the peripheral retina is represented in progressively more anterior parts of the visual areas, LSF information may be predominantly processed in these areas.

However, retinotopic processing of spatial frequency was not always observed in elderly participants. LSF scene categorization elicited greater activation in the anterior part of the cuneus than HSF scene categorization. However, HSF scene categorization did not elicit stronger significant activation than LSF scenes. The absence of significant activation was consistent with the observed behavioral deficit in HSF processing in older participants compared to younger participants. This may result from anatomical and functional changes in the primary visual cortex associated with normal aging. Salat et al. [47], for example, observed a global thinning of the cerebral cortex, including the primary visual areas, with age. Brewer and Barton [48,49] showed that the size of the receptive field corresponding to central vision was larger in elderly than in young adults. To be precise, the size of the receptive fields corresponding to an eccentricity of 3° in elderly adults (over the age of 57) is equivalent to one of 8° in young adults (under the age of 36). This result may suggest that absence of the receptive field tuned to HSF processing in elderly people. The authors also showed that the surface of projection of the fovea was smaller in the elderly than in young adults [49], suggesting a decrease in the cortical surface dedicated to HSF processing. Further anatomical studies using, for instance, voxel-based morphometry (VBM) should be conducted to determine whether the HSF deficit in normal elderly participants could be linked to changes in gray matter in the visual areas related to central vision.

Direct comparison between groups (between-group analysis) revealed greater activation of the right inferior occipital gyrus in young participants (peak coordinates: 20x, -82y, -7z) than in elderly participants for the categorization of HSF scenes only (greater activation was not observed for NF and LSF scenes). This activation overlaps with the retinotopic processing of HSF information in the right hemisphere in young participants (peak coordinates: 20x, -82y, -7z), and mirrors the absence of HSF specific significant activation in older participants (within-group analysis). Overall, results from within- and between-analyses suggest that an occipital cortex dysfunction underlies HSF scene categorization **in elderly people**. However, a high level of activation was noted in a temporo-pariet al cortical network in elderly participants for the categorization of LSF and HSF (but not NF) scenes. We cannot rule out that this enhanced activity was linked to increased attentional resources in elderly participants.

Firstly, we observed greater activation in the inferior pariet al lobules, with a greater extent for HSF than LSF scenes. These regions have previously been described as mediating and sustaining the allocation of attention to low and high spatial frequencies in scenes [69,70] and may be engaged by elderly participants to emphasize spatial frequency information, in particular HSF. We also observed stronger activation in the right middle frontal gyrus. It should be noted that, for elderly

participants, only the right middle frontal gyrus was more activated by the processing of HSF than LSF at a more lenient statistical threshold. This region has previously been described as being involved in sustained attention [71] and in people with low vision when more attentional resources are required for words recognition [72]. Both activations therefore suggest that for elderly people the categorization of filtered scenes requires greater attentional resources.

We also observed stronger activation in the left posterior middle temporal cortex in elderly participants during the categorization of both LSF and HSF scenes. This region is known to be involved in the semantic categorization of visual stimuli [70,73,74,75], and could be linked to the retrieval of semantic concepts related to the visual stimulus [74]. Alternatively, this region is also known to be involved in word retrieval [76] and word generation [77]. Its activation could be related to internal dialogue which elderly participants used to generate the semantic category label (indoor or outdoor). High-level visual areas associated with semantic categorization were, therefore, more highly activated in older than in young participants. The elderly participants may engage such additional semantic processes to compensate for their deficit in low-level vision and diminished occipital cortex activation. Interestingly, despite their behavioral deficit when categorizing HSF scenes, persistent activation in the left posterior middle temporal cortex was present in this experimental condition. Stronger activation of left high-level visual areas (left middle temporal gyrus) paralleled with weaker activation of right low-level visual areas (right inferior occipital gyrus). It seems that elderly participants made more demands on areas in the left than in the right hemisphere in order to categorize HSF scenes. This is consistent with the involvement of verbal processes for which the left hemisphere is dominant. However, it should be noted that a greater activation of the left temporal cortex, but also of the left primary motor cortex (BA 4) and right cerebellum directly linked to the motor response given by participants, was observed for LSF than HSF scenes for elderly participants (see the within-group analysis), and is consistent with the higher non-response rate observed for the categorization of HSF scenes.

Finally, we also investigated age-related differences in spatial frequency processing in the PPA. The PPA was evidenced to be a scene-selective cortical region [53,55]. This region has been linked to high-order functions during scene perception, such as navigation, spatial layout processing and scene recognition [54,56,78,79,80], but also contextual association [52,81]. Recent studies revealed that this region is also sensitive to spatial frequency information in scenes [59,60,82]. In a recent fMRI study in which the experimental paradigm was the same as that used in the present study [59], results showed that NF and LSF scenes elicited greater activation than HSF scenes in the PPA for young participants. The present study showed that activation of the PPA did not differ between spatial frequencies for young participants maybe due to the smaller sample size (12 instead of 16 young participants). Importantly, results showed that the age of participants interacted with the processing of spatial frequencies in the PPA. Activation of the PPA significantly decreased during HSF scene categorization (compared to LSF) for elderly participant, suggesting that the specific age-related HSF deficit could also impact the processing of spatial frequencies in PPA.

Conclusion

The present findings indicate a specific low-contrasted HSF deficit in elderly people, in association with occipito-temporal cortex dysfunction. Our study provides important additional information for the development of models of scene perception and has an impact on public health. Understanding of the cerebral interpretation of scenes in elderly people is therefore of primary importance for the

improvement of their quality of life. Our study provides also further support to the understanding of visual impairments in pathological aging. Indeed, the number of visual impairments related to age is in constant progression in developed countries. Age-related macular degeneration (AMD) is the first cause of central vision loss in the elderly population and it mainly affects people over the age of 50 [83,84,85]. The investigation of spatial frequency processing in natural environments is even more important since elderly people with AMD exhibit a specific deficit in the processing of HSF in photographs of natural scenes [45].

Supplementary material

A control behavioral study was conducted in the same participants (12 young and 12 elderly participants), in which the luminance contrast of filtered stimuli was equalized using a RMS (root mean square) contrast normalization. Scenes were normalized to obtain an RMS contrast of 0.1 (i.e. 25.6 on a gray-level scale). We chose a value situated between LSF and HSF contrast values in natural conditions (i.e. 0.21 and 0.04, respectively in the LUM condition), in order to avoid affecting one spatial frequency condition more than another. This contrast normalization reduces contrast in LSF while enhancing HSF contrast. Participants performed the same categorization task except that scenes with equalized RMS contrast replaced the original scenes.

Results showed that the global performance was improved by the RMS contrast normalization (mNR: 1.9 ± 3.3%, mFC: 4.1 ± 5.9%, mRT: 632 ± 83 ms) in comparison with the original experiment (mNR: 9.7 ± 7.2%, mFC: 5.6 ± 7.3%, mRT: 657 ± 87 ms). However, results still showed that elderly participant performance was poorer than that of younger participants when categorizing scenes filtered in HSF only. Three 2x2x2 analyses of variance (ANOVA) with Groups (young and elderly participants) as between-subjects factor, and Spatial frequencies (NF, LSF, and HSF) and Categories (indoors and outdoors) as within-subjects factors were conducted on mNR, mFC and mRT. The ANOVA on mNR revealed that elderly participants responded less often than young participants (3.1 ± 4.9% and 0. 1 ± 0.1%, respectively; $F_{1,22}$ = 4.30, $p < 0.05$). The expected Group x Spatial frequency interaction was significant ($F_{2,44}$ = 4.42, $p < 0.05$). Planned comparison showed that elderly participants responded less often when categorizing HSF scenes (6.1 ± 9.2%) than NF scenes (0.1 ± 0.2%; $F_{1,22}$ = 7.49, $p < 0.05$) and LSF scenes (2.2 ± 3.8%; $F_{1,22}$ = 7.32, $p < 0.05$), while there was no difference between LSF and NF scenes ($F_{1,22}$ = 2.17, $p = 0.15$). For young participants, there was no effect of spatial frequencies (NF: 0.1 ± 0.2%; LSF: 0.1 ± 0.2%; HSF: 0.1 ± 0.1%; all $F_{1,22} < 1$). In addition, elderly participants responded less often than young participants only when categorizing HSF scenes ($F_{1,22}$ = 5.35, $p < 0.05$; NF: $F_{1,22} < 1$; LSF; $F_{1,22}$ = 1.12, $p = 0.30$). The ANOVA on mFC did not show a main effect of Groups ($F_{1,22} < 1$). The Group x Spatial frequency interaction was significant ($F_{2,44}$ = 6.84, $p < 0.01$). Planned comparison showed that elderly participants made more false categorizations when categorizing HSF scenes (7.0 ± 8.0%) than NF scenes (0.3 ± 0.7%; $F_{1,22}$ = 7.43, $p < 0.05$) and LSF scenes (4.9 ± 6.9%; $F_{1,22}$ = 14.86, $p < 0.001$), and when categorizing LSF than NF scenes ($F_{1,22}$ = 5.52, $p < 0.05$). For young participants, there was no effect of spatial frequencies (NF: 3.3 ± 4.3%; LSF: 3.4 ± 5.0%; HSF: 2.8 ± 4.1%; all $F_{1,22} < 1$). Finally, the ANOVA on mRT revealed a main effect of Groups ($F_{1,22}$ = 5.95, $p < 0.05$). Elderly participants were slower to categorize scenes than healthy participants (671 ± 82 and 593 ± 82 ms, respectively). The Group x Spatial frequency interaction was not significant ($F_{2,44}$ = 2.18, $p = 0.12$). There was a main effect of Spatial frequency ($F_{2,44}$ = 23.10, $p < 0.001$). All participants had longer RT when categorizing HSF scenes (661 ± 91 ms) than NF scenes (601 ± 76 ms; $F_{1,22}$ = 37.30, $p < 0.001$) and LSF scenes (633 ± 82 ms; $F_{1,22}$ = 10.69, $p <$

0.005), and when categorizing LSF than NF scenes ($F_{1,22} = 15.81$, $p < 0.001$).

Acknowledgments

This work was supported by the RECOR "Agence Nationale pour la Recherche" Grant (ANR-12-JHS2-0002-01 RECOR). Grenoble MRI facility IRMaGE was partly funded by the French program "Investissement d'Avenir" run by the "Agence Nationale pour la Recherche": Grant "Infrastructure d'Avenir en Biologie Santé" (ANR-11-INBS-0006). Louise Kauffmann was supported by Région Rhône-Alpes (Cible Grants). The authors warmly thank the "Délégation à la Recherche Clinique et à l'Innovation" of the University Hospital of Grenoble and the "Cellule de Neuroimagerie Fonctionnelle" of the Laboratory of Psychology and NeuroCognition for sponsoring. We thank Catherine Dal Molin for the English revision of the manuscript.

References

1. Van Essen DCV, Deyoe EA (1995) Concurrent processing in the primate visual cortex. In M. S. Gazzaniga (Éd.).The Cognitive Neurosciences. Cambridge: Bradford Book. pp 383-400.
2. Bullier J (2001) Integrated model of visual processing. Brain Research, Brain Research Reviews 36(2-3): 96‑107.
3. Ginsburg AP (1986) Spatial filtering and visual form perception. In Boff K, Kauman L, Thomas J, eds. Editors. Handbook of perception and human performance. New York: Wiley-Interscience 1-41.
4. Hughes HC, Nozawa G, Kitterle F (1996) Global precedence, spatial frequency channels, and the statistics of natural images. Journal of Cognitive Neuroscience 8(3): 197‑230.
5. Guyader N, Chauvin A, Peyrin C, Hérault J, Marendaz C (2004) Image phase or amplitude? Rapid scene categorization is an amplitude-based process. Comptes Rendus Biologie 327(4): 313‑318.
6. Torralba A, Oliva A (2003) Statistics of natural image categories. Network. 14(3): 391‑412.
7. Breitmeyer B, Julesz B (1975) The role of on and off transients in determining the psychophysical spatial frequency response. Vision Research 15(3): 411‑415.
8. 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.
9. Parker DM, Roly J, 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.
10. Schyns PG, Oliva A (1994) From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. Psychological Science 5(4): 195‑200.
11. Kauffmann L, Ramanoël S, Peyrin C (2014) The neural bases of spatial frequency processing during scene perception. Frontiers in Integrative Neuroscience 8: 37.
12. Henriksson L, Nurminen L, Hyvärinen A, Vanni S (2008) Spatial frequency tuning in human retinotopic visual areas. Journal of Vision 8(10): 5.1‑13.
13. Musel B, Bordier C, Dojat M, Pichat C, Chokron S, et al. (2013) Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. Journal of Cognitive Neuroscience 25(8): 1315‑1331.
14. Sasaki Y, Hadjikhani N, Fischl B, Liu AK, Marret S, et al. (2001) Local and global attention are mapped retinotopically in human occipital cortex. Proceedings of the National Academy of Sciences of the United States of America 98(4): 2077‑2082.
15. De Fockert JW, Ramchurn A, van Velzen J, Bergström Z, Bunce D (2009) Behavioral and ERP evidence of greater distractor processing in old age. Brain Research 1282: 67‑73.
16. Geerligs L, Saliasi E, Maurits NM, Renken RJ, Lorist MM (2014) Brain mechanisms underlying the effects of aging on different aspects of selective attention. NeuroImage 91: 52‑62.

17. Haring AE, Zhuravleva TY, Alperin BR, Rentz DM, Holcomb PJ, et al. (2013) Age-related differences in enhancement and suppression of neural activity underlying selective attention in matched young and old adults. Brain Research 1499: 69‑79.
18. Park DC, Reuter-Lorenz P (2009) The adaptive brain: aging and neurocognitive scaffolding. Annual Review of Psychology 60(1): 173‑196.
19. Park DC, Lautenschlager G, Hedden T, Davidson NS, Smith AD, et al. (2002) Models of visuospatial and verbal memory across the adult life span. Psychology and Aging 17(2): 299‑320.
20. Park DC, Schwarz N (1999) Cognitive Aging: A Primer (1 edition.). Philadelphia, PA: Psychology Press.
21. Marchand WR, Lee JN, Suchy Y, Garn C, Johnson S, et al. (2011) Age-related changes of the functional architecture of the cortico-basal ganglia circuitry during motor task execution. NeuroImage 55(1): 194‑203.
22. Brown B, Lovie-Kitchin J (1993) Repeated visual acuity measurement: establishing the patient's own criterion for change. Optometry and Vision Science 70(1): 45‑53.
23. Elliott DB, Yang KC, Whitaker D (1995) Visual acuity changes throughout adulthood in normal, healthy eyes: seeing beyond 6/6. Optometry and Vision Science 72(3): 186‑191.
24. Gittings NS, Fozard JL (1986) Age related changes in visual acuity. Experimental Gerontology 21(4-5): 423‑433.
25. Rubin GS, West SK, Muñoz B, Bandeen-Roche K, Zeger S, et al. (1997) A comprehensive assessment of visual impairment in a population of older Americans. The SEE study. Salisbury Eye Evaluation Project. Investigative Ophthalmology & Visual Science 38(3): 557‑568.
26. Elliott DB (1987) Contrast sensitivity decline with ageing: a neural or optical phenomenon? Ophthalmic & Physiological Optics 7(4): 415‑419.
27. Elliott D, Whitaker D, MacVeigh D (1990) Neural contribution to spatiotemporal contrast sensitivity decline in healthy ageing eyes. Vision Research 30(4): 541‑547.
28. Higgins KE, Jaffe MJ, Caruso RC, deMonasterio FM (1988) Spatial contrast sensitivity: effects of age, test-retest, and psychophysical method. Journal of the Optical Society of America A 5(12): 2173-80.
29. Owsley C (2011) Aging and vision. Vision Research 51(13): 1610‑1622.
30. Owsley C, Sekuler R, Siemsen D (1983) Contrast sensitivity throughout adulthood. Vision Research 23(7): 689‑699.
31. Sekuler R, Hutman LP, Owsley CJ (1980) Human aging and spatial vision. Science (New York, N.Y.). 209(4462): 1255‑1256.
32. Hardy JL, Delahunt PB, Okajima K, Werner JS (2005) Senescence of spatial chromatic contrast sensitivity. I. Detection under conditions controlling for optical factors. Journal of the Optical Society of America A 22(1): 49‑59.
33. Elliott SL, Werner JS (2010) Age-related changes in contrast gain related to the M and P pathways. Journal of Vision 10(4):1-15.
34. Pokorny J, Smith VC (1997) Psychophysical signatures associated with magnocellular and parvocellular pathway contrast gain. Journal of the Optical Society of America A 14(9): 2477‑2486.
35. Navon D (1977) Forest before trees: The precedence of global features in visual perception. Cognitive Psychology 9(3): 353‑383.
36. Badcock JC, Whitworth FA, Badcock DR, Lovegrove WJ (1990) Low-frequency filtering and the processing of local-global stimuli. Perception 19(5): 617‑629.
37. Lamb MR, Yund EW (1993) The role of spatial frequency in the processing of hierarchically organized stimuli. Perception & Psychophysics. 54(6): 773‑784.
38. Shulman GL, Sullivan MA, Gish K, Sakoda WJ (1986) The role of spatial-frequency channels in the perception of local and global structure. Perception 15(3): 259‑273.

39. Lux S, Marshall JC, Thimm M, Fink GR (2008) Differential processing of hierarchical visual stimuli in young and older healthy adults: implications for pathology. Cortex 44(1): 21‑28.
40. Oken BS, Kishiyama SS, Kaye JA, Jones DE (1999) Age-related differences in global-local processing: stability of laterality differences but disproportionate impairment in global processing. Journal of Geriatric Psychiatry and Neurology 12(2): 76‑81.
41. Staudinger MR, Fink GR, Mackay CE, Lux S (2011) Gestalt perception and the decline of global precedence in older subjects. Cortex 47(7): 854‑862.
42. Bruyer R, Scailquin JC (2000) The fate of global precedence with age. Experimental Aging Research 26(4): 285‑314.
43. Roux F, Ceccaldi M (2001) Does aging affect the allocation of visual attention in global and local information processing? Brain and Cognition 46(3): 383‑396.
44. Viggiano MP, Righi S, Galli G (2006) Category-specific visual recognition as affected by aging and expertise. Archives of Gerontology and Geriatrics 42(3): 329 ‑338.
45. Musel B, Hera R, Chokron S, Alleysson D, Chiquet C, et al. (2011) Residual abilities in age-related macular degeneration to process spatial frequencies during natural scene categorization. Visual Neuroscience 28(6): 529‑541.
46. Lemaître H, Crivello F, Grassiot B, Alpérovitch A, Tzourio C et al. (2005) Age- and sex-related effects on the neuroanatomy of healthy elderly. NeuroImage 26(3): 900‑911.
47. Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RSR, et al. (2004) Thinning of the cerebral cortex in aging. Cerebral Cortex 14(7): 721‑730.
48. Brewer AA, Barton B (2012) Effects of healthy aging on human primary visual cortex. Health 04(09): 695‑702.
49. Brewer AA, Barton B (2014) Visual cortex in aging and Alzheimer's disease: changes in visual field maps and population receptive fields. Frontiers in Psychology 5: 74.
50. Crossland MD, Morland AB, Feely MP, von dem Hagen E, Rubin GS (2008) The effect of age and fixation instability on retinotopic mapping of primary visual cortex. Investigative Ophthalmology & Visual Science 49(8): 3734‑3739.
51. Field DJ (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.
52. Aminoff E, Gronau N, Bar M (2007) The parahippocampal cortex mediates spatial and nonspatial associations. Cerebral Cortex 17(7): 1493-1503.
53. Downing PE, Chan AWY, Peelen MV, Dodds CM, Kanwisher N (2006) Domain specificity in visual cortex. Cerebral Cortex 16(10):1453-1461.
54. Epstein RA, Harris A, Stanley D, Kanwisher N (1999) The parahippocampal place area: recognition, navigation or encoding? Neuron 23(1):115-125.
55. Epstein RA, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392(6676):598-601.
56. Epstein RA, Ward EJ (2010) How reliable are visual context effects in the parahippocampal place area? Cerebral Cortex 20(2)294-303.
57. Rosenholtz R, Li Y, Nakano L (2007) Measuring visual clutter. Journal of Vision 7(2): 17.1‑22.
58. Willenbockel V, Fiset D, Chauvin A, Arguin M, Tanaka JW et al. (2010) Does face inversion change spatial frequency tuning? Journal of Experimental Psychology. Human Perception and Performance 36:122-35.
59. Kauffmann L, Ramanoël S, Guyader N, Chauvin A, Peyrin C (2015) Spatial frequency processing in scene-selective cortical regions. NeuroImage 112:86-95.
60. Musel B, Kauffmann L, Ramanoël S, Giavarini C, Guyader N, et al. (2014) Coarse-to-fine categorization of visual scenes in scene-selective cortex. Journal of Cognitive Neuroscience 26(10):2287-2297.
61. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, et al. (1995) Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping 2(4): 189‑210.

62. Ashburner J (2007) A fast diffeomorphic image registration algorithm. NeuroImage 38(1):95-113.
63. Glascher J, Gitelman D (2008) Contrast weights in flexible factorial design with multiple groups of subjects. Unpublished tutorial. Retrieved from http://jiscmail.ac.uk/cgi-bin/webadmin?A2=ind0803&L=SPM&P=R16629 (assessed 26.09.2012).
64. Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.
65. Bex PJ, Makous W (2002) Spatial frequency, phase and the contrast of natural images. Journal of the Optical Society of America A 19(6):1096-1106.
66. Palmer S (1993) Modern theories of gestalt perception, in Understanding Vision ed. Humphreys G. W., editor. (Oxford: Blackwell;) 39-70.
67. Curcio CA, Allen KA (1990) Topography of ganglion cells in human retina. The Journal of Comparative Neurology 300(1): 5‑25.
68. Curcio CA, Sloan KR, Kalina RE, Hendrickson AE (1990) Human photoreceptor topography. The Journal of Comparative Neurology 292(4): 497‑523.
69. Peyrin C, Baciu M, Segebarth C, Marendaz C (2004) Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. NeuroImage 23(2): 698‑707.
70. Peyrin C, Michel CM, Schwartz S, Thut G, Seghier M, et al. (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.
71. Cabeza R, Nyberg L (2002) Neural bases of learning and memory: functional neuroimaging evidence. Current Opinion in Neurology 13(4): 415‑421.
72. Szlyk JP, Little DM (2009) An FMRI study of word-level recognition and processing in patients with age-related macular degeneration. Invest Ophthalmol Vis Sci 50: 4487–4495.
73. Gerlach C, Law I, Gade A, Paulson OB (2000) Categorization and category effects in normal object recognition: a PET study. Neuropsychologia 38(13): 1693‑1703.
74. Mummery CJ, Patterson K, Hodges JR, Price CJ (1998) Functional neuroanatomy of the semantic system: divisible by what? Journal of Cognitive Neuroscience 10(6): 766‑777.
75. Thioux M, Pesenti M, Costes N, De Volder A, Seron X (2005) Task-independent semantic activation for numbers and animals. Cognitive Brain Research 24(2): 284‑290.
76. Abrahams S, Goldstein LH, Simmons A, Brammer MJ, Williams SCR, et al. (2003) Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses. Human Brain Mapping 20(1): 29‑40.
77. Friedman L, Kenny JT, Wise AL, Wu D, Stuve TA, et al. (1998). Brain activation during silent word generation evaluated with functional MRI. Brain and Language. 64(2): 231‑256.
78. Epstein RA (2005) The cortical basis of visual scene processing. Visual Cognition 12:954-978.
79. Epstein RA, Graham KS, Downing PE (2003) Viewpoint-specific scene representation in parahippocampal cortex. Neuron 37:865-876.
80. Epstein RA, Higgins JS, Jablonsky K, Feiler AM (2007) Visual scene processing in familiar and unfamiliar environment. Journal of Neurophysiology 97:3670-3683.
81. Bar M, Aminoff E, Schacter DL (2008) Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. The Journal of Neuroscience 28(34):8539-8544.
82. Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RBH (2011) The “parahippocampal place area” responds preferentially to high spatial frequencies in humans and monkeys. PLoS Biology 9(4):e1000608.
83. Friedman DS, O’Colmain BJ, Muñoz B, Tomany SC, McCarty C, et al. (2004)

Prevalence of age-related macular degeneration in the United States. Archive of Ophthalmology 122(4): 564‑572.

84. Klein R, Chou CF, Klein BEK, Zhang X, Meuer SM, et al. (2011) Prevalence of age-related macular degeneration in the US population. Archives of Ophthalmology 129(1): 75‑80.
85. Vingerling JR, Dielemans I, Hofman A, Grobbee DE, Hijmering M, et al. (1995) The prevalence of age-related maculopathy in the Rotterdam Study. Ophthalmology 102(2): 205‑210.

Louise KAUFFMANN

Doctorante en Psychologie Cognitive

Née le 28/03/1989 (26 ans)
Nationalité Française

Coordonnées personnelles :
72 rue Abbé Grégoire
38000 GRENOBLE

Tel: 06 80 59 04 03
louise.kauffmann@gmail.com

Coordonnées professionnelles :
Laboratoire de Psychologie et Neurocognition (LPNC)
Université Pierre Mendès-France
BP47, 38040 Grenoble Cedex 9 France
Tel: 04 76 82 54 00
louise.kauffmann@upmf-grenoble.fr

Page web: http://web.upmf-grenoble.fr/lpnc/membre_louise_kauffmann

CURSUS UNIVERSITAIRE

En préparation : Thèse de Doctorat en Psychologie Cognitive.
Sujet de recherche : Mécanismes et bases cérébrales du traitement des fréquences spatiales lors de la catégorisation de scènes visuelles.
Lieu : Laboratoire de Psychologie et Neurocognition (LPNC), Grenoble.
Direction : Carole Peyrin (CR1, LPNC) et Alan Chauvin (MCF, LPNC).

Sept-Déc 2014 : Séjour de Recherche « Indoc » à l'étranger.
Sujet de recherche : Stratégies d'analyse de scènes et traitement des fréquences spatiales dans la myopie.
Lieu : Laboratoire de Psychophysique et de Perception Visuelle, Université de Montréal, Canada.
Direction : Guillaume Giraudet (Professeur) et Jocelyn Faubert (Professeur).

2011-2012 : Master 2 Recherche Psychologie Cognitive et Sociale, UPMF, Grenoble.
Sujet de recherche : Perception de scènes narturelles : l'analyse 'Coarse-to-Fine' et ses bases cérébrales.
Lieu : Laboratoire de Psychologie et Neurocognition, Grenoble.
Direction : Carole Peyrin (CR1, LPNC) et Alan Chauvin (MCF, LPNC).

2010-2011 : Master 1 Psychologie Cognitive, UPMF, Grenoble.
Sujet de recherche : Informations diagnostiques pour la reconnaissance de scènes : application de la méthode " Bubbles " à l'espace de Fourier.
Lieu : Laboratoire de Psychologie et Neurocognition, Grenoble.
Direction : Alan Chauvin (MCF, LPNC).

2007-2010 : Licence de Psychologie
Lieu : Université Pierre Mendès-France, Grenoble.

2006-2007 : Baccalauréat Section Littéraire
Lieu : Lycée Aristide Bergès, Seyssinet.

ACTIVITE SCIENTIFIQUE

Publications

Ramanoël, S., **Kauffmann, L.**, Cousin, E., Dojat, M., Peyrin, C. (Sous presse). Age-related differences in spatial frequency processing during scene categorization. *PlosOne*.

Kauffmann, L., Bourgin, J., Guyader, N., Peyrin, C. (Sous presse). The neural bases of the semantic interference of spatial frequency-based information in scenes. *Journal of Cognitive Neuroscience.*

Kauffmann, L. Chauvin, A., Pichat, C., Peyrin, C. (2015). Effective connectivity in the neural network undelying coare-to-fine categorization of scenes. A dynamic causal modelling study. *Brain and Cognition*. *99*, 46-56.

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., Peyrin, C. (2015). Spatial frequency Processing in scene-selective cortical regions. *NeuroImage*. *112,* 86-95.

Kauffmann, L., Chauvin, A., Guyader, N., Ramanoël, S., 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*, 1-14.

Musel, B., **Kauffmann, L**., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., Peyrin, C. (2014) Coarse-to-fine processing of visual scenes within scene-selective cortex. *Journal of cognitive neuroscience.26(10),* 2287-2297.

Communications affichées

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing within scene-selective cortical regions. Organization of Human Brain Mapping (HBM), USA.

Ramanoël, S., **Kauffmann, L.**, Cousin, E., Dojat, M., & Peyrin, C. (2015). Age-related differences in spatial frequency processing during scene categorization. Organization of Human Brain Mapping (HBM), USA.

Kauffmann. L., Chauvin, A., Guyader, N., Ramanoël, S., Peyrin, C. (2014). Does RMS contrast normalization impair coarse-to-fine processing of natural scenes? *Visual Science Society.* St Pete Beach, Florida.

Kauffmann, L., Musel, B., Chauvin, A., Guyader, N., Pichat, C., Ramanoël, S., Le Bas, J.-F., Peyrin, C. (2013). Dynamic scenes as an experimental tool to investigate the coarse-to-fine processing within scene-selective cortical areas. Colloque de la Société des Neurosciences, Lyon, France.

Ramanoël, S., **Kauffmann, L.**, Guyader, N., Chauvin, A., Pichat, C., Dojat, M., Peyrin, C. (2013). RMS contrast normalization induces changes in the retinotopic processing of spatial frequencies in scenes. Colloque de la Société des Neurosciences, Lyon, France.

Kauffmann, L., Musel, B., Chauvin, A., Guyader, N., Pichat, C., Ramanoel, S., Le Bas, J. F., Peyrin, C. (2013), Dynamic scenes as an experimental tool to investigate the neural substrates of coarse-to-fine processing in scene perception. Cognitive Neuroscience Society, San Fransisco, USA.

Peyrin, C., Musel, B., **Kauffmann, L**., Guyader, N., Pichat, C., Ramanoel, S., Chauvin, A. (2013), Coarse-to-Fine categorization of scenes within the parahippocampal place area. Cognitive Neuroscience Society, San Fransisco, USA.

Kauffmann, L., Musel, B., Guyader, N., Peyrin, C., & Chauvin, A. (2012). Dynamic scenes as an experimental tool to provide behavioral evidence for a Coarse-to-Fine processing sequence during scene categorization. European Conference on Visual Processing, Alghero, Sardinia, Italy.

Bourgin, J., Maire, C., **Kauffmann, L.**, Chauvin, A., Guyader, N., & Peyrin, C. (2012). Comparisons of static and dynamic paradigms to investigate the Coarse-to-Fine categorization of scenes. European Conference on Visual Processing, Alghero, Sardinia, Italy.

Communications orales

Kauffmann, L. Chauvin, A., Guyader, N., Ramanoël, S., Peyrin, C. (2014). Quand un contrôle devient un biais : effet de l'égalisation du contraste sur le traitement des fréquences spatiales. *5ème rencontres du Pôle Grenoble Cognition*. Grenoble, France.

Expertises

Revues internationales : NeuroImage

ACTIVITE D'ENSEIGNEMENT

2013-2015 Monitorat

Travaux dirigés :

Psychologie cognitive (L2)	30h	Méthode expérimentale (VI, VD, hypothèses) et analyse d'expériences en psychologie cognitive.
Psychologie cognitive (L1)	30h	Introduction à la méthode expérimentale (VI, VD, hypothèse, résultats…).
Analyse et Interprétation de données expérimentales (L2)	48h	Approfondissement de la méthode expérimentale et de l'analyse de données (t de Student, ANOVA).
Perception, Action et Catégorisation (L3)	6h	Analyse d'expériences sur la formation de concepts au cours du développement.

Cours magistraux :

Cognition Adulte (L2)	6h	Introduction à la cognition visuelle, reconnaissance visuelle, modèles neurocognitifs de la perception visuelle.

Mars 2014 Organisation d'une formation (3 jours)

« Apports de la Psychologie à l'activité d'enseignement » dans le cadre des formations proposées aux autres doctorants par le service Doctoral pour la Formation, l'Initiation et l'insertion professionnelles de l'Université de Grenoble.

Co-encadrement d'étudiants de Master

Dorante Miller (2013-2014), co-encadré avec Alan Chauvin dans le cadre de son mémoire de Master 1.

COMPETENCES

Protocoles expérimentaux	Maîtrise du logiciel E-prime.
Analyses statistiques	Maîtrise de Statistica, SPSS. Bases de programmation (R, Matlab).
Neuro-imagerie	Acquisition de données IRMf Traitement de données IRMf (SPM) Analyses en régions d'intérêt Connectivité (Dynamic Causal Modeling, Psycho-physiological interactions)
Langues	Français (langue maternelle) Anglais
Formations scientifiques	Connectivité cérébrale fonctionnelle (GIN, Grenoble) Journées inter-régionales de formation en IRMf (GIN, Grenoble)

RESUME

L'analyse visuelle de scènes débute par l'extraction en parallèle de l'information à différentes fréquences spatiales, en suivant un mode de traitement par défaut « coarse-to-fine ». L'analyse rapide de l'information grossière (« coarse ») en basses fréquences spatiales fournirait un aperçu global de la scène, qui serait ensuite affiné par l'analyse des détails de la scène (« fine ») en hautes fréquences spatiales. L'objectif de cette thèse a été de préciser les bases cérébrales du traitement des fréquences spatiales lors de la catégorisation de scènes. A travers deux études comportementales, nous avons tout d'abord montré qu'une analyse « coarse-to-fine » est plus avantageuse pour la catégorisation rapide de scènes, et ce, indépendamment de la valeur de contraste de luminance associée aux différentes fréquences spatiales (Expériences 1 et 2). Des études en IRMf nous ont par la suite permis de mettre en évidence l'implication d'un large réseau cérébral lors de l'analyse « coarse-to-fine » des scènes, incluant les aires visuelles primaires et occipito-temporales, mais également le cortex frontal inférieur (Expérience 3). Une analyse de la connectivité a révélé que lors de cette analyse, le cortex frontal inférieur exercerait une influence « top-down » sur le cortex visuel primaire et les gyri fusiforme et parahippocampique au sein du cortex occipito-temporal. Ces résultats soulignent le rôle du cortex visuel primaire comme région intégrative, codant à la fois les afférences rétino-thalamiques et les influences « top-down » de régions supérieures. Nous avons également observé que le gyrus frontal inférieur et le gyrus fusiforme participaient activement à l'intégration de l'information sémantique contenue dans les basses et hautes fréquences spatiales d'une scène (Expérience 4). Enfin, nous avons spécifiquement étudié le traitement des fréquences spatiales au sein de régions occipito-temporales sélectives aux scènes : la « parahippocampal place area » (PPA), le cortex retrosplenial et l'« occipital place area ». Nous avons montré que ces trois régions participent de façon distincte au traitement des fréquences spatiales dans les scènes (Expérience 5) et qu'une stratégie d'analyse « coarse-to-fine » serait privilégiée par la PPA (Expérience 6). Les résultats de ces travaux nous permettent de conforter et de préciser les modèles actuels de la catégorisation visuelle de scènes basés sur un traitement fréquentiel de l'information visuelle.

Mots clés : Scènes naturelles, Fréquences spatiales, Coarse-to-fine, IRMf, Cortex visuel, Cortex frontal inférieur, Cortex inférotemporal, Régions sélectives aux scènes.

ABSTRACT

Visual analysis begins with the parallel extraction of different attributes at different spatial frequencies following a predominantly coarse-to-fine default processing sequence. Rapid processing of low spatial frequency information would permit a coarse parsing of the visual input, prior to the detailed analysis of fine information in high spatial frequencies. Our aim was to further address the neural bases of spatial frequency processing during scene categorization. We first demonstrated in two behavioral studies that a coarse-to-fine processing is indeed an advantageous strategy for rapid scene categorization, and is independent of the luminance contrast values associated with the different spatial frequencies (Experiments 1 and 2). In two fMRI studies, we showed first the involvement of a large cerebral network during coarse-to-fine processing of scenes, including early visual and occipito-temporal areas, but also the inferior frontal cortex. Effective connectivity analysis revealed that the inferior frontal gyrus exerts top-down influence on the early visual cortex as well as on the parahippocampal and fusiform gyri in the occipito-temporal cortex (Experiment 3). These results highlight the role of the primary visual cortex in integrating top-down influences from frontal areas to retino-thalamic incoming signals. We also evidenced that the inferior frontal and fusiform gyri actively participate to the integration of the semantic information contained in low and high spatial frequency (Experiment 4). Finally, we specifically investigated the spatial frequency processing of scenes within scene-selective areas of the occipito-temporal cortex: the parahippocampal place area (PPA), the retrosplenial cortex, and the occipital place area. We demonstrated that these regions participate differently in the spatial frequency processing of scenes (Experiment 5) and that a coarse-to-fine processing is favored within the PPA (Experiment 6). Overall, results allow us to refine current model of visual scene categorization based on a spatial frequency analysis.

Keywords: Natural scenes, Spatial frequencies, Coarse-to-fine, fMRI, Visual cortex, Inferior frontal cortex, Inferotemporal cortex, Scene-selective regions.