



HAL
open science

Etude de la contribution de la vision centrale et de la vision périphérique lors de la catégorisation de scènes

Audrey Trouilloud

► **To cite this version:**

Audrey Trouilloud. Etude de la contribution de la vision centrale et de la vision périphérique lors de la catégorisation de scènes. Psychologie. Université Grenoble Alpes [2020-..], 2021. Français. NNT : 2021GRALS042 . tel-03774838

HAL Id: tel-03774838

<https://theses.hal.science/tel-03774838v1>

Submitted on 12 Sep 2022

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : PCN - **Sciences cognitives, psychologie et neurocognition**

Arrêté ministériel : 25 mai 2016

Présentée par

Audrey TROUILLOUD

Thèse dirigée par **Carole PEYRIN**, Université Grenoble Alpes

préparée au sein du **Laboratoire de Psychologie et Neuro Cognition**
dans **l'École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

Etude de la contribution de la vision centrale et de la vision périphérique lors de la catégorisation de scènes

Thèse soutenue publiquement le **07 décembre 2021**
devant le jury composé de :

Dr. Quentin LENOBLE

Maître de conférence, Université de Lille, France.
Rapporteur

Dr. Benoit COTTEREAU

Chercheur CNRS, Université de Toulouse, France.
Rapporteur

Pr. Anne GUERIN-DUGUE

Professeur, Université Grenoble-Alpes, France.
Examinatrice

Dr. Nathalie GUYADER

Maître de conférence, Université Grenoble-Alpes, France.
Examinatrice

Pr. Olivier KOENIG

Professeur, Université de Lyon, France.
Examinateur, Président du jury

Dr. Carole PEYRIN

Directrice de recherche CNRS, Université Grenoble-Alpes.
Directrice de thèse



Résumé

Le système visuel humain est particulièrement efficace pour saisir rapidement l'essentiel de l'environnement visuel, et ce en dépit de sa complexité. Une première représentation globale et rudimentaire d'une scène visuelle serait possible grâce au traitement rapide des basses fréquences spatiales (BFS) du signal visuel. Cette première représentation servirait également à générer très rapidement des prédictions sur l'environnement visuel afin d'anticiper et guider une analyse plus minutieuse de la scène qui repose sur le traitement des hautes fréquences spatiales (HFS). Étant donné que les BFS sont majoritairement extraites par la rétine périphérique, nous avons défendu l'hypothèse que les prédictions sont majoritairement générées sur la base des informations disponibles en vision périphérique. Ceci nous a amené dans ce travail de thèse à confronter les traitements réalisés respectivement en vision centrale et vision périphérique lors de la catégorisation de scènes. Nous avons pour cela mené plusieurs expériences comportementales auprès de jeunes adultes avec une vision normale. Nous avons tout d'abord testé la catégorisation d'images hybrides superposant deux scènes différentes, une en BFS et l'autre en HFS. Nous avons montré que la catégorisation de la scène en HFS était meilleure lorsque la scène en BFS qui se superposait était sémantiquement congruente qu'incongruente suggérant un traitement automatique de l'information en BFS, même lorsqu'elle n'est pas explicitement utile à la catégorisation, pour guider le traitement des HFS. Cet effet était plus fort lorsque les images hybrides étaient présentées en vision périphérique qu'en vision centrale. Nous avons également testé la catégorisation d'images révélant différentes parties d'une même scène à travers des anneaux circulaires de différentes excentricités rétinienne. Nous avons observé que les performances de catégorisation étaient meilleures pour les anneaux les plus périphériques suggérant que malgré leur faible résolution spatiale, les informations disponibles en vision périphérique sont plus utiles et plus efficaces que celles disponibles en vision centrale pour catégoriser très rapidement une scène. Nous avons ensuite montré que la catégorisation de la partie centrale d'une scène était meilleure lorsqu'on présentait simultanément un anneau périphérique sémantiquement congruent qu'incongruent. Ainsi, les informations disponibles en vision périphérique guideraient automatiquement les traitements réalisés en vision centrale. Nous avons enfin étudié les conséquences d'une perte de la vision périphérique sur la catégorisation de scènes en menant une expérience comportementale auprès de patients atteints d'un glaucome primitif à angle ouvert (GPAO). Nous avons testé la catégorisation d'images hybrides dans la vision centrale des patients et nous avons observé que l'influence des BFS sur la catégorisation de scènes en HFS diminuait avec l'avancée dans la maladie et la perte progressive de stimulation en vision périphérique. L'ensemble de ces travaux de thèse démontre le rôle primordial de la vision périphérique dans la catégorisation de scènes et supportent le rôle prédictif de la vision périphérique pour guider l'analyse de l'information présente au point de fixation visuelle.

Abstract

The human visual system is particularly efficient at quickly understanding most of the visual environment despite its complexity. A first global and rudimentary representation of a visual scene would be possible with the rapid processing of the low spatial frequencies (LSF) of the visual signal. This first representation would also be used to very quickly generate predictions about the visual environment in order to anticipate and guide a more detailed analysis of the scene based on high spatial frequency (HSF) processing. Since LSF are mostly extracted in the peripheral retina, we hypothesized that predictions are mostly generated based on information available in peripheral vision. This led us in this doctoral thesis to compare visual processing in central and peripheral vision during scene categorization. To this end, we conducted several behavioral experiments in young adults with normal vision. First, we tested the categorization of hybrid images superimposing two different scenes, one in LSF and the other in HSF. We showed that the categorization of the HSF scene was better when the LSF scene was semantically congruent than incongruent, suggesting that LSF information, even when irrelevant for categorization, was automatically processed to guide HSF processing. This effect was stronger when the hybrid images were presented in peripheral vision than in central vision. We also tested the categorization of images revealing different parts of the same scene through circular rings of different retinal eccentricities. We observed that categorization performance was better for the most peripheral rings, suggesting that despite their low spatial resolution, the information available in peripheral vision is more useful than information available in central vision to categorize a scene very quickly. We then showed that the categorization of the central part of a scene was better when a semantically congruent peripheral ring was presented simultaneously than when an incongruent one was presented. Thus, the information available in peripheral vision would automatically guide the processing performed in central vision. Finally, we investigated the consequences of peripheral vision loss on scene categorization by conducting a behavioral experiment in patients with primary open-angle glaucoma (POAG). We tested hybrid image categorization in the patients' central vision and observed that the influence of LSF on HSF scene categorization decreased with the disease progression and progressive loss of stimulation in peripheral vision. All of this thesis work demonstrates the primordial role of peripheral vision in scene categorization and supports the predictive role of peripheral vision to guide the analysis of information present at the point of visual fixation.

Remerciements

Je remercie chaleureusement Quentin Lenoble, Benoit Cottreau, Anne Guerin-Dugué, Nathalie Guyader et Olivier Koenig d'avoir accepté d'évaluer ce travail de thèse en faisant partie de mon jury. Je remercie également Marcela Perrone et Denis Pelisson d'avoir suivi l'avancée de mes travaux de thèse lors du comité de suivi individuel.

Ce travail de thèse n'aurait pas été possible sans l'implication, de près ou de loin, de nombreuses personnes que je souhaite remercier ici.

Je remercie d'abord ma Directrice de thèse, Carole Peyrin. Celle sans qui rien n'aurait été possible. Plus que des remerciements, je souhaite lui exprimer ma profonde gratitude pour m'avoir accompagnée durant mes études, et ce depuis ma première année de Master. Au-delà des compétences qu'elle a su me transmettre, à ses côtés j'ai appris la rigueur scientifique et la persévérance. Son soutien sans faille, sa disponibilité et ses nombreux conseils durant la rédaction de ma thèse m'ont permis de la concrétiser. Enfin, je souhaite remercier la petite Aurore qui a permis à sa maman de m'accompagner « jusqu'au bout ». Les intéressées se reconnaîtront !

Je souhaite remercier Muriel Boucart pour son accueil à Lille et sa collaboration dans un projet qui a impliqué des patients atteints d'un glaucome. Mais aussi l'équipe du service ophtalmologique du CHU de Lille, notamment le Professeur Jean-François Rouland pour son implication dans ce projet et son accueil au sein du service. Merci aussi à Elvia Ferry qui a réalisé les nombreuses passations expérimentales auprès des patients.

Merci également à tous les étudiants de psychologie de l'Université de Grenoble Alpes et aux patients du service ophtalmologique du CHU de Lille qui ont participé aux expériences de ce travail de thèse.

Je remercie infiniment Alexia Roux-Sibilon et Louise Kauffmann qui m'ont précédée dans l'équipe. Alexia, merci pour le temps que tu m'as consacré, tu as été d'une grande aide et d'un grand soutien. Je suis reconnaissante d'avoir pu bénéficier de tes conseils et de tes compétences. Louise, merci pour ton aide, pour ta disponibilité et ta bienveillance. Mais surtout, pour ta patience vis-à-vis de mes nombreuses erreurs sur Matlab.

Merci à Pauline Rossel pour ton implication dans l'article et pour ton aide précieuse dans le traitement des données en modèles mixtes, tes compétences en statistiques m'ont beaucoup apporté. C'est un véritable plaisir de travailler avec toi.

Je remercie également Cynthia Faurite et Eva Aprile qui ont contribué à l'acquisition des données expérimentales et qui ont passé de longues heures enfermées dans un box expérimental.

Je tiens à remercier chaleureusement le Laboratoire de Psychologie et NeuroCognition et tous ses membres pour m'avoir permis de réaliser cette thèse dans des conditions plus que parfaites. Merci à tous pour votre bonne humeur quotidienne et votre sympathie. Merci à Martial, Jean-Charles, David, Céline et Aurélie de m'avoir accordé votre confiance pour assurer les différents enseignements, vous m'avez donné le goût d'enseigner et c'est grâce à vous que j'ai souhaité poursuivre une année en tant qu'ATER.

Evidemment, merci aux doctorants du laboratoire qui sont aujourd'hui bien plus que de simples collègues. Notamment, merci à mes copains du bureau E118 : Méline, Merrick, Rémi, Thomas et Martin, vous m'avez tant fait rire, vous êtes de véritables rayons de soleil. A mes copains du bureau E121 : Cynthia, Adeline, Léa, Elie et Sarah, vous avez été d'un soutien incroyable. Merci à tous les autres, Olivier, Samuel, Maëlle, Sonja, Laura, Lucrèce, Céline, Elise, Lucie, Brice et Ali d'avoir égayé ces trois années de thèse par votre bonne humeur et votre humour.

J'adresse toute ma reconnaissance au Docteur Brinzeu, neurochirurgien à l'hôpital Pierre Weithermer de Lyon. Sans lui, cette thèse n'aurait tout simplement pas pu aboutir.

Un grand merci à ma famille de cœur, spécialement à Mélanie, Alicia, les 'psychopotes' Clara, Manon et Malena pour leur amitié et leur soutien. Nous avons partagé tous les moments importants de nos vies, de l'adolescence aux années étudiantes et des années étudiantes à l'âge adulte. A travers ces âges et dans toutes les situations, j'ai toujours pu compter sur vous. Merci aussi à ma famille de cœur de Grenoble, Emilie, Marine, Yanis et toute la team, pour votre présence au quotidien, vos attentions et vos encouragements.

Je remercie avec émotion ma famille pour son soutien inconditionnel. Merci Maman de m'avoir apporté cette force au quotidien, d'être toujours disponible pour m'aider dans n'importe quelle situation, à n'importe quelle heure. Merci Estelle, ma sœur jumelle, de me supporter tous les jours et de me faire prendre confiance en moi ; tu n'as jamais douté de moi. Merci aussi pour tes bonnes intuitions, je me souviens encore du jour où tu m'as dit « tu feras de la recherche plus tard », alors que j'étais seulement en première année de licence et que je n'y avais jamais pensé. Apparemment, tu ne t'es pas trompée. Merci Papou et Mamou, pour vos encouragements et votre bienveillance incroyable. Merci Philippe et Sam pour votre soutien et votre humour, notamment sur mon avenir professionnel (vous savez de quoi je parle). Philippe, merci d'avoir corrigé les fautes d'orthographe de mon manuscrit.

Enfin, je remercie par-dessus tout mon David, pour avoir cru en moi et m'avoir tirée vers le haut. Tu m'as écouté m'entraîner pour ma soutenance de Master 2 et tu m'écoutes aujourd'hui m'entraîner pour ma soutenance de thèse. Je suis reconnaissante de pouvoir bénéficier d'un tel soutien et d'une telle bienveillance. Ton optimisme quotidien m'a permis de relativiser de nombreuses fois et grâce à toi, je suis devenue une meilleure version de moi-même. Merci aussi d'avoir respecté ma bulle dans laquelle je m'étais enfermée les dernières semaines de rédactions.

Table des matières

RESUME	I
ABSTRACT	II
REMERCIEMENTS	III
LISTE DES FIGURES	IX
LISTE DES TABLEAUX	XII
ABREVIATIONS	XIII
PREAMBULE	1
CHAPITRE 1 - CADRE THEORIQUE DE LA RECONNAISSANCE VISUELLE	3
Encart 1 : Traitement des fréquences spatiales d'images numériques	4
1. Traitement de l'information visuelle de la rétine au cerveau	6
1.1. De la rétine au cortex visuel primaire	6
1.2. Le cortex visuel primaire et le phénomène de magnification corticale	8
1.3. Les voies visuelles corticales	10
2. Traitement des fréquences spatiales et modèle cognitif <i>coarse-to-fine</i> de la perception de scènes visuelles	12
2.1. Démonstration psychophysique de canaux fréquentiels et d'une précedence temporelle du traitement des basses fréquences spatiales	12
2.2. Modèle cognitif <i>coarse-to-fine</i> de la perception de scènes visuelles	17
2.3. Limites du modèle <i>coarse-to-fine</i>	21
3. Traitements ascendants et descendants de l'information visuelle	24
3.1. Modèle hiérarchique du cortex visuel	24
3.2. Limites d'une conception purement hiérarchique du traitement cortical	25
3.3. Modèle intégratif du traitement visuel	26
3.4. Théorie de la hiérarchie inversée	28
4. Traitements visuels prédictifs et rôle des fréquences spatiales	29
4.1. Théorie du codage prédictif appliquée à la perception visuelle	29
4.2. Modèle proactif de la reconnaissance visuelle d'objets	30
4.3. Bases neurales de l'analyse <i>coarse-to-fine</i> de scènes visuelles	32
4.4. Spécialisation hémisphérique du traitement des fréquences spatiales et du traitement visuel prédictif	34

4.5. Intégration des fréquences spatiales lors de la perception de scènes visuelles	36
4.6. Modèle fréquentiel prédictif de la perception de scènes	40
5. Problématique et objectifs du travail de thèse	41

CHAPITRE 2 - ETUDE DE L'INTEGRATION DES FREQUENCES SPATIALES EN VISION CENTRALE VERSUS PERIPHERIQUE LORS DE LA CATEGORISATION DE SCENES

45

1. Expérience 1 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides physiquement dissimilaires	47
1.1. Méthode	47
1.2. Hypothèses	51
1.3. Analyse des résultats	53
1.4. Discussion	60
2. Expérience 2 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides physiquement similaires	65
2.1. Méthode	65
2.2. Analyse des résultats	67
2.3. Discussion	73
3. Expérience 3 - Test de l'influence de la similarité physique (scènes hybrides physiquement similaire et dissimilaires)	77
3.1. Méthode	79
3.2. Analyse des résultats	80
4. Discussion générale	91

CHAPITRE 3 - ETUDE DE LA CONTRIBUTION RESPECTIVE DE LA VISION CENTRALE ET DE LA VISION PERIPHERIQUE LORS DE LA CATEGORISATION DE SCENES

98

I. Expériences 4 et 5 - Importance de la vision périphérique par rapport à la vision centrale	102
1. Expérience 4 - Catégorisation de larges scènes et de parties de scènes de différentes excentricités rétiniennes	104
1.1. Méthode	104
1.2. Hypothèses	108
1.3. Analyse des résultats	109
1.4. Discussion	115
2. Expérience 5 - Catégorisation de larges scènes dynamiques en expansion ou en contraction	117
2.1. Méthode	117
2.2. Hypothèses	119
2.3. Analyse des résultats	120
3. Discussion générale	124
4. Conclusion et limites	127

II. Expériences 6, 7 et 8 - Importance de la vision périphérique par rapport à la vision centrale : Contrôle de biais méthodologiques 130

1. Expérience 6 – Catégorisation de larges scènes et de parties de scènes de différentes excentricités rétiniennes 131

- 1.1. Méthode 131
- 1.2. Hypothèses 136
- 1.3. Analyse des résultats 136
- 1.4. Discussion 142

2. Expérience 7 - Catégorisation de larges scènes dynamiques en expansion ou en contraction 144

- 2.1. Méthode 144
- 2.2. Hypothèses 146
- 2.3. Analyse des résultats 147
- 2.4. Discussion 150

3. Expérience 8 – Contrôle d’un effet de masquage latéral rétroactif 151

- 3.1. Méthode 151
- 3.2. Hypothèses 154
- 3.3. Analyse des résultats 154

4. Discussion générale 158

CHAPITRE 4 - INFLUENCE DES INFORMATIONS SEMANTIQUES ET PHYSIQUES EN VISION PERIPHERIQUE SUR LA CATEGORISATION DES SCENES EN VISION CENTRALE 163

1. Expérience 9 - Influences mutuelles des informations provenant de la vision périphérique et de la vision centrale lors de la catégorisation de scènes 170

- 1.1. Méthode 170
- 1.2. Hypothèses 172
- 1.3. Analyse des résultats 173
- 1.4. Discussion 175

2. Expérience 10 - Nature de l’interférence de la vision périphérique sur la catégorisation de scènes en vision centrale 176

- 2.1. Méthode 176
- 2.2. Hypothèses 179
- 2.3. Analyse des résultats 179

3. Discussion générale 181

CHAPITRE 5 - ETUDE DU ROLE DE LA VISION PERIPHERIQUE A TRAVERS UN MODELE PATHOLOGIQUE : LE GLAUCOME PRIMITIF A ANGLE OUVERT 186

1. Contexte théorique 186

- 1.1. Définition du glaucome primitif à angle ouvert (GPAO) 186

Encart 2 : Evaluation de la perte du champ visuel à l’aide d’une périmétrie de Humphrey 188

1.2. Troubles visuels cognitifs dans le glaucome	189
1.3. Objectifs et hypothèses	191
2. Expérience 11 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides dans la vision centrale de patients atteints d'un glaucome	195
2.1. Méthode	195
2.2. Analyse des résultats	200
3. Discussion générale	205
CHAPITRE 6 – DISCUSSION GENERALE	211
1. Implications théoriques des résultats	211
1.1. Apport des résultats aux modèles existants	211
1.2. Extension du modèle FP-Péri	213
2. Perspectives de recherche	218
2.1. Limites et perspectives théoriques	218
2.2. Perspectives cliniques	221
ANNEXE 1	223
ANNEXE 2	237
BIBLIOGRAPHIE	264

Liste des figures

Figure A	Exemples d'une photographie filtrée en basses fréquences spatiales et en hautes fréquences spatiales	4
Figure B	Illustration de la procédure de filtrage d'une scène grâce à la transformée de Fourier	5
Figure C	Évaluation de la perte du champ visuel à l'aide d'une périmétrie statique automatisée Humphrey SITA-Standard 24-2	188
Figure D	Exemples de champs visuels de patients atteints d'un GPAO obtenus sur l'œil testé à l'aide d'une périmétrie statique automatisée (Humphrey SITA-Standard 24-2)	188
Figure 1.1	Organisation rétinotopique des aires visuelles et illustration de la magnification corticale dans V1	10
Figure 1.2	Exemples de réseaux sinusoidaux et carrés en basses fréquences spatiales et en hautes fréquences spatiales et fonction de sensibilité au contraste	13
Figure 1.3	Fonctions de sensibilité au contraste qui représente les différents canaux fréquentiels sélectifs à une bande de fréquences spatiales	14
Figure 1.4	Fonction de sensibilité au contraste chez un observateur normal en fonction de la fréquence spatiale et de l'excentricité rétinienne	15
Figure 1.5	Exemple de formes hiérarchiques utilisées par Navon (1977)	16
Figure 1.6	Exemple d'images hybrides utilisées par Schyns et Oliva (1994)	18
Figure 1.7	Exemple de séquences dynamiques utilisées par Musel et al. (2012)	19
Figure 1.8	Latences de réponse visuelle de neurones de différentes aires corticales (Bullier, 2001)	26
Figure 1.9	Schématisation du modèle intégratif du traitement visuel de Bullier (2001)	27
Figure 1.10	Modèle de hiérarchie inversée de Hochstein et Ahissar (2002)	29
Figure 1.11	Modèle proactif de la reconnaissance visuelle d'objets de Moshe Bar (2006, 2007)	31
Figure 1.12	Exemple de stimuli hybrides utilisés dans Kauffmann, Bourgin et al. (2015) et Kauffmann et al. (2017)	37
Figure 1.13	Modèle Fréquentiel Prédicatif adapté de la thèse d'Alexia Roux-Sibilon (2020)	41
Figure 2.1	Exemples de stimuli utilisés dans l'Expérience 1	49
Figure 2.2	Schéma de la procédure expérimentale de la Session 2 et de la Session 3 de l'Expérience 1	51
Figure 2.3	Résultats de la Session 1 de l'Expérience 1 (catégorisation de scènes filtrées en BFS et en HFS en vision centrale)	54
Figure 2.4	Résultats des Sessions 2 et 3 de l'Expérience 1 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : Pourcentages d'erreurs moyens	56
Figure 2.5	Résultats des Sessions 2 et 3 de l'Expérience 1 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : Temps de réponses correctes moyens en ms	57

Figure 2.6	Résultats de la Session 3 de l'Expérience 1 (catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche) : Pourcentage d'erreurs moyens	59
Figure 2.7	Résultats de la Session 3 de l'Expérience 1 (catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche) : Temps de réponses correctes moyens en ms	60
Figure 2.8	Exemple de stimuli utilisés dans l'Expérience 2	66
Figure 2.9	Résultats de la Session 1 de l'Expérience 2 (catégorisation de scènes filtrées en BFS et en HFS en vision centrale)	68
Figure 2.10	Résultats des Sessions 2 et 3 de l'Expérience 2 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : Pourcentages d'erreurs moyens	70
Figure 2.11	Résultats des Sessions 2 et 3 de l'Expérience 2 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : Temps de réponses correctes moyens en ms	71
Figure 2.12	Résultats de la Session 3 de l'Expérience 2 (catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche)	73
Figure 2.13	Exemple d'images hybrides utilisées dans l'Expérience 3	80
Figure 2.14	Résultats de la Session 1 de l'Expérience 3 (catégorisation de scènes filtrées en BFS et en HFS en vision centrale et en vision périphérique)	82
Figure 2.15	Résultats de la Session 1 de l'Expérience 3 (catégorisation de scènes filtrées en BFS et en HFS dans le champ visuel droit et dans le champ visuel gauche)	83
Figure 2.16	Résultats des Sessions 2 de l'Expérience 3 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : Pourcentage d'erreurs moyens	86
Figure 2.17	Résultats des Sessions 2 de l'Expérience 3 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : Temps de réponses correctes moyens en ms	89
Figure 3.1	Exemple de stimuli utilisés dans Larson et Loschky (2009)	100
Figure 3.2	Représentation schématique des séquences <i>central-to-peripheral</i> et <i>peripheral-to-central</i>	101
Figure 3.3	Exemple de stimuli utilisés dans la Session 1 (Disque entier) et dans la Session 2 (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4) de l'Expérience 4	105
Figure 3.4	Schéma de la procédure expérimentale de l'Expérience 4	108
Figure 3.5	Résultats de la Session 1 (catégorisation du Disque entier) et de la Session 2 (catégorisation du Disque central, de l'Anneau 1, de l'Anneau 2, de l'Anneau 3 et de l'Anneau 4) de l'Expérience 4	110
Figure 3.6	Schéma de la procédure expérimentale de l'Expérience 5	119
Figure 3.7	Résultats de l'Expérience 5	123
Figure 3.8	Exemples de stimuli utilisés dans les différentes conditions expérimentales de l'étude de Geuzebroek et van den Berg (2018)	126
Figure 3.9	Illustration de l'effet de masquage latéral rétroactif ou masquage par métacontraste (<i>Metacontract masking</i>) sur un extrait d'une séquence <i>central-to-peripheral</i>	129

Figure 3.10	Exemple de stimuli utilisés dans la Session 1 (Disque entier) et dans la Session 2 (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4) de l'Expérience 6	134
Figure 3.11	Schéma de la procédure expérimentale de l'Expérience 6	135
Figure 3.12	Résultats de la Session 1 (catégorisation du Disque entier) et de la Session 2 (catégorisation du Disque central, de l'Anneau 1, de l'Anneau 2, de l'Anneau 3 et de l'Anneau 4) de l'Expérience 6	141
Figure 3.13	Schéma de la procédure expérimentale de l'Expérience 7	146
Figure 3.14	Résultats de l'Expérience 7	149
Figure 3.15	Schéma de la procédure expérimentale de l'Expérience 8	153
Figure 3.16	Résultats de l'Expérience 8	157
Figure 4.1	Exemples de stimuli utilisés dans l'étude de Roux-Sibilon et al. (2019)	164
Figure 4.2	Exemples de paires de scènes avec leur spectre d'amplitude utilisés dans Peyrin et al. (2021) et schéma de la procédure expérimentale	165
Figure 4.3	Modèle Fréquentiel Prédicatif incluant la vision périphérique issu de la thèse d'Alexia Roux-Sibilon (2020)	167
Figure 4.4	Exemple de stimuli utilisés dans Lukavský (2019)	168
Figure 4.5	Exemple de stimuli utilisés dans l'Expérience 9	171
Figure 4.6	Résultats de l'Expérience 9	175
Figure 4.7	Exemple de stimuli utilisés dans l'Expérience 10	178
Figure 4.8	Résultats de l'Expérience 10	181
Figure 4.9	Modèle FP-Peri dans le cadre du glaucome	185
Figure 5.1	Schématisation de la périmétrie statique automatisée des 30 degrés centraux d'un œil gauche	186
Figure 5.2	Périmétrie automatisée de de Humphrey SITA-Standard 24-2 de deux patients et exemples de stimuli faiblement contrastés utilisés dans l'étude de Roux-Sibilon et al. (2018)	190
Figure 5.3	Illustration du paradigme des <i>Steady-Pulsed pedestal</i> utilisé par McKendrick, Badcock et Morgan (2004)	192
Figure 5.4	Exemples de stimuli utilisés dans l'étude de Roux-Sibilon et al. (2019)	194
Figure 5.5	Exemples de stimuli utilisés dans l'Expérience 11	199
Figure 5.6	Résultats de la Session 1 de l'Expérience 11 (catégorisation de scènes filtrées en BFS et en HFS en vision centrale)	201
Figure 5.7	Résultats de la Session 2 de l'Expérience 11 (catégorisation d'images hybrides en vision centrale)	204
Figure 5.8	Résultats de la Session 2 de l'Expérience 11 (catégorisation d'images hybrides en vision centrale) : Relation entre l'indice MD des patients et l'effet d'interférence sémantique en pourcentages d'erreurs moyens	205
Figure 6.1	Extension du modèle FP-Péri	214
Figure 6.2	Exemple de paradigme expérimental de choix saccadique	215
Figure 6.3	Exemple de stimuli utilisé dans l'étude de Cajar et al. (2020)	217
Figure 6.4	Exemple de stimuli de visages utilisés pour tester le phénomène de crowding dans l'étude de Louie et al. (2007)	219

Liste des tableaux

Tableau 1.1	Ensemble des expériences réalisées au cours de cette thèse et aperçu des principaux stimuli visuels utilisés	44
Tableau 2.1	Ensemble des expériences réalisées au cours de cette thèse et aperçu des principaux stimuli visuels utilisés	46
Tableau 2.2	Résultats de la Session 1 de l'Expérience 1	54
Tableau 2.3	Résultats des Sessions 2 et 3 de l'Expérience 1	55
Tableau 2.4	Résultats de la Session 3 de l'Expérience 1	58
Tableau 2.5	Résultats de la Session 1 de l'Expérience 2	67
Tableau 2.6	Résultats des Sessions 2 et 3 de l'Expérience 2	68
Tableau 2.7	Résultats de la Session 3 de l'Expérience 2	72
Tableau 2.8	Résultats de la Session 1 de l'Expérience 3	81
Tableau 2.9	Résultats de la Session 1 de l'Expérience 3	82
Tableau 2.10	Résultats de la Session 2 de l'Expérience 3	84
Tableau 2.11	Résultats de la Session 2 de l'Expérience 3	90
Tableau 3.1	Résultats de la Session 1 de l'Expérience 4	109
Tableau 3.2	Résultats de la Session 2 de l'Expérience 4	111
Tableau 3.3	Résultats de l'Expérience 5	120
Tableau 3.4	Résultats de la Session 1 de l'Expérience 6	137
Tableau 3.5	Résultats de la Session 2 de l'Expérience 6	138
Tableau 3.6	Résultats de l'Expérience 7	147
Tableau 3.7	Résultats de l'Expérience 8	155
Tableau 4.1	Résultats de l'Expérience 9	174
Tableau 4.2	Résultats de l'Expérience 10	180
Tableau 5.1	Classification HPA (Hodapp, Parrish & Anderson, 1993) de l'importance de l'atteinte périmétrique du glaucome	187
Tableau 5.2	Données individuelles cliniques des 52 participants de l'Expérience 11	196
Tableau 5.3	Résultats de la Session 1 de l'Expérience 11	200
Tableau 5.4	Résultats de la Session 2 de l'Expérience 11	203

Abréviations

%ERm	Pourcentages d'erreurs moyens
ANOVA	Analyse de la variance
BFS	Basses Fréquences spatiales
Cpd	Cycle par degré
Cpi	Cycle par image
CVD	Champ visual droit
CVG	Champ visual gauche
DCM	Dynamic causal modelling
DMLA	Dégénérescence maculaire liée à l'âge
EEG	Electroencéphalographique
F	F de Fisher
FEF	Frontal Eye Fields
FP	Fréquentiel Prédicatif
FSC	Fonction de sensibilité au contraste
GPAO	Glaucome primitif à angle ouvert
HD	Hémisphère droit
HFS	Hautes Fréquences spatiales
HG	Hémisphère gauche
IRMf	Imagerie par Résonance Magnétique fonctionnelle
K	Koniocellulaire
M	Millisecondes
M	Magnocellulaire
MD	Mean Deviation
MEG	Magnétoencéphalographie
MST	Medial Superior Temporal
MT	Medial Temporal
P	Parvocellulaire
PPA	Parahippocampal place area
PRL	Preferred retinal locus
RMS	Root-Mean-Square
SOA	Stimulus Onset Asynchrony
t	T de student
TRm	Temps de réponses correctes moyens
V1	Cortex visuel primaire
VBM	Voxel-based morphometry

Préambule

Le système visuel humain est particulièrement efficace pour extraire rapidement l'essentiel d'une scène. Nous sommes capables de juger très facilement si nous sommes sur une plage de sable fin face à un océan qui s'étend loin à l'horizon ou sur un sentier boueux entouré d'arbres au milieu d'une forêt et ce, en dépit des variations de lumière au cours de la journée, des changements de contraste et de couleurs et de la variabilité des éléments qu'on trouve dans ces deux types d'environnement. Plus étonnant, nous sommes capables de catégoriser instantanément les larges scènes que représentent notre environnement visuel sans même avoir conscience que la majorité de l'information visuelle est floue.

Pour bien comprendre ce paradoxe, il faut tout d'abord préciser que la résolution spatiale de l'information visuelle n'est pas uniforme à travers le champ visuel humain. Le champ visuel correspond à la partie visible de l'environnement lorsque les yeux fixent un point. La vision centrale est la partie du champ visuel autour de cette fixation. Elle est considérée comme la partie la plus importante du champ visuel, certainement en raison de sa très bonne résolution spatiale, ou acuité visuelle. L'acuité visuelle est la capacité à discerner les détails. Elle est la plus élevée dans une petite zone de vision centrale, la fovéa, qui s'étend approximativement à 1° d'excentricité (Curcio & Allen, 1990 ; Curcio et al., 1990 ; Larson & Loschky, 2009 ; Polyak, 1941 ; Yamada, 1969). La vision centrale nous permet de lire, de reconnaître des visages et des objets, de conduire et bien d'autres actions quotidiennes. La partie restante du champ visuel se projette sur la parafovéa, qui s'étend jusqu'à environ 4-5° d'excentricité rétinienne, et sur la rétine périphérique, qui couvre tout le champ visuel restant (Larson & Loschky, 2009 ; Loschky et al., 2017). Bien que notre expérience visuelle subjective semble riche et détaillée, en partie grâce aux mouvements constants de nos yeux, la résolution spatiale diminue à mesure qu'on s'éloigne de la fixation et finalement, seules des informations de faible résolution spatiale sont disponibles dans la plupart du champ visuel.

Pourtant, de nombreuses recherches menées en psychologie et neurosciences cognitives suggèrent le rôle crucial des informations périphériques pour catégoriser les scènes visuelles. Ce travail de thèse a pour objectif d'approfondir nos connaissances sur le rôle de la vision périphérique lors de la catégorisation visuelle de scènes en considérant les différences de traitement entre la vision centrale et la vision périphérique. Cette distinction est plus ou moins arbitraire, puisqu'il n'existe pas de consensus scientifique sur la définition du champ visuel périphérique. Chez l'homme, le champ de vision binoculaire avec mouvements des yeux représente environ 200° d'angle visuel horizontalement et 125° verticalement (Loschky et al., 2017). Bien qu'il soit difficile de délimiter précisément la vision centrale et la vision périphérique, on peut toutefois distinguer la vision centrale qui s'étend sur un rayon d'environ 5° autour de la fixation et la vision périphérique qui correspondrait à la surface restante du champ visuel. Pour notre problématique de recherche, nous parlerons de vision centrale et de vision périphérique en opposant grossièrement la portion du champ visuel autour de la fixation versus le reste du champ visuel.

Dans le Chapitre 1 de ce manuscrit de thèse, nous présenterons différents modèles de la reconnaissance visuelle de scènes. La plupart de ces modèles reposent actuellement sur le traitement du contenu en fréquences spatiales du signal visuel, une caractéristique visuelle intrinsèquement liée à la résolution spatiale du champ visuel. Nous nous focaliserons ensuite sur les travaux actuels qui nous amènent à nous interroger sur le rôle mésestimé de la vision périphérique dans les modèles de reconnaissance visuelle.

Les chapitres 2 à 5 seront consacrés aux études que nous avons réalisées au cours de ce travail de thèse. Nous avons mené des expériences comportementales auprès de participants sains (Chapitres 2, 3 et 4), ainsi qu'auprès de patients atteints d'un glaucome (Chapitre 5) en utilisant des tâches de catégorisation de scènes visuelles. Dans le Chapitre 2, nous avons étudié comment les différentes fréquences spatiales d'une scène s'influencent mutuellement d'une part, en vision centrale et d'autre part, en vision périphérique. Dans le Chapitre 3, nous avons étudié distinctement la catégorisation de scènes en vision centrale et en vision périphérique afin de tester de quelle façon ces différentes régions du champ visuel contribuent respectivement à la reconnaissance rapide de scènes. Dans le Chapitre 4, nous avons étudié comment les informations visuelles disponibles en vision périphérique peuvent influencer la catégorisation de scènes en vision centrale. Enfin, dans le Chapitre 5, nous avons étudié la catégorisation de scènes dans la vision centrale de patients atteints d'un glaucome. Cette pathologie ophtalmique, qui affecte principalement la rétine périphérique, nous a permis d'étudier directement l'impact de la perte de la vision périphérique sur les traitements réalisés en vision centrale. Notre objectif ici était de tester si la perte de vision périphérique entraîne un déficit de la catégorisation de scènes en vision centrale.

Le Chapitre 6 de ce manuscrit sera consacré à la Discussion générale. Nous résumerons les principaux résultats des expériences que nous avons réalisées et nous tenterons de les articuler dans un modèle de catégorisation de scène intégrant les traitements inhérents à la vision périphérique.

Chapitre 1 - Cadre théorique de la reconnaissance visuelle

Plus que tout autre sens, la vision implique une grande diversité de traitements comme la perception de formes, de couleurs, de mouvements ou encore de profondeurs. Dans ce travail de thèse, nous nous focaliserons sur la perception de larges scènes de la vie quotidienne, statique et en deux dimensions, car elles permettent de décrire l'essentiel notre environnement visuel. Ces scènes sont généralement composées de formes globales, qui sont elles-mêmes composées de petites formes locales. Par exemple, lorsque nous regardons une scène de ville, nous visualisons d'abord les grandes maisons et les immeubles sous forme de carrés et rectangles juxtaposés les uns à côté des autres, puis à l'intérieur de ces grandes formes, nous distinguons les portes et les fenêtres. Ces informations visuelles correspondent à des variations de luminance, ou alternance entre des bandes claires et sombres, qui diffèrent par leur amplitude, leur orientation et leur position dans l'espace. Ces variations sinusoïdales de luminance sont appelées fréquences spatiales, les basses fréquences spatiales (BFS) correspondant à des variations espacées de niveaux de gris et les hautes fréquences spatiales (HFS) correspondant à des variations plus abruptes. Pour une définition détaillée des fréquences spatiales dans le domaine du traitement d'images numériques, nous renvoyons le lecteur à l'**Encart 1**. Les basses et les hautes fréquences spatiales fournissent alors des informations différentes sur la scène visuelle. Les BFS fournissent une information sur les formes globales et la structure de la scène, alors que les HFS fournissent les détails de la scène visuelle comme les contours précis des objets. L'être humain semble tout particulièrement adapté à percevoir les fréquences spatiales puisque le système visuel humain est doté de neurones sensibles aux variations de luminance en basses et en hautes fréquences spatiales. Dans ce chapitre théorique de thèse, nous verrons tout d'abord les mécanismes de traitement des fréquences spatiales dont dispose le système visuel. Nous présenterons ensuite différents modèles de la perception visuelle de scène, puis les récents travaux qui ont montré le rôle fondamental de la vision périphérique pour la perception de scènes. Nous finirons ce chapitre théorique sur l'objectif et la problématique de cette thèse.

Encart 1 : Traitement des fréquences spatiales d'images numériques

Une image numérique est avant tout un signal qui peut être décrit de deux façons différentes. Dans le domaine spatial, une image est considérée comme une matrice contenant des pixels qui expriment la valeur de la luminance en chaque point. Elle peut également être considérée comme une matrice contenant trois canaux de couleurs : Rouge, Vert, Bleu (RGB). Néanmoins, nous ne détaillerons pas la chrominance dans ce manuscrit en raison de son intérêt moindre.

Dans le domaine fréquentiel, une image est considérée comme une superposition de réseaux sinusoïdaux de différentes fréquences spatiales. Chaque réseau sinusoïdal se caractérise par sa fréquence spatiale et son amplitude. La fréquence spatiale est caractérisée quant à elle par le nombre d'ondulations de luminance par unité de distance. Elle est exprimée en cycle par degré d'angle visuel (cpd) ou en cycles par image (cpi) et l'amplitude représente la différence de luminosité entre les ondulations. Le cpd est une mesure qui varie car elle dépend de la distance entre l'observateur et l'image alors que le cpi est une mesure indépendante de l'observateur. Plus il y a de cpd et de cpi et plus la fréquence spatiale est élevée, et inversement (Figure A).

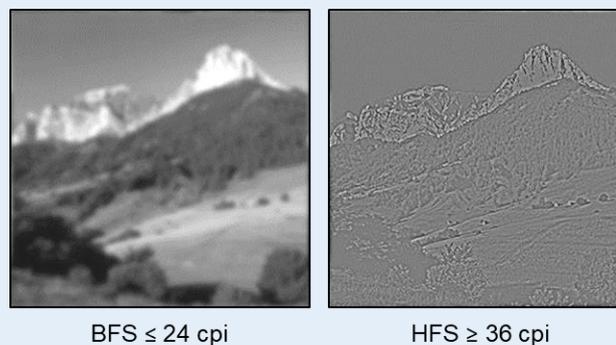


Figure A : Exemples d'une photographie filtrée en basses fréquences spatiales (BFS ; à gauche) et en hautes fréquences spatiales (HFS ; à droite). Pour la photographie filtrée en BFS, les fréquences spatiales inférieures à 24 cycles par image (cpi) ont été conservées et pour celle filtrée en HFS, les fréquences spatiales supérieures à 36 cpi ont été conservées.

Lorsque l'on s'intéresse à la perception de scènes visuelles, il peut être intéressant de manipuler les images en filtrant certaines informations. Ceci est possible grâce à la transformée de Fourier, une suite d'opérations mathématiques linéaires qui permet de décomposer un signal visuel en une somme de réseaux sinusoïdaux qui peuvent être caractérisées par leur fréquence, leur amplitude, leur phase et leur orientation. Grâce à la transformée de Fourier, une image dans le domaine fréquentiel peut alors être caractérisée par son spectre d'amplitude et son spectre de phase (Figure B). Le spectre d'amplitude fournit la distribution du contraste de luminance (amplitude) en fonction des orientations et des fréquences spatiales. Le spectre de phase fournit quant à lui la position relative de la luminance des fréquences spatiales dans l'espace de l'image. Sur le spectre d'amplitude, les BFS de l'image sont représentées au centre, là où l'amplitude est la plus importante, et les HFS de l'image sont représentées en périphérie. Pour les scènes visuelles naturelles, l'amplitude du signal décroît de manière inverse à la fréquence spatiale (i.e. le contraste est plus important dans les BFS que dans les HFS ; Field, 1987) et suit une forme $1/f^2$ où f correspond à la fréquence spatiale.

Dans le domaine fréquentiel, on peut donc contrôler les fréquences spatiales d'une image en filtrant certaines parties du spectre d'amplitude. Cette procédure est réalisable grâce à la transformée de Fourier inverse, un outil indispensable pour étudier le rôle des fréquences spatiales dans la perception visuelle. Par exemple, on peut appliquer au spectre d'amplitude un filtre passe-bas qui va laisser passer uniquement le signal en BFS du spectre en éliminant le signal en HFS supérieur à une

fréquence donnée (la fréquence de coupure). A l'inverse, un filtre passe-haut va laisser passer uniquement le signal en HFS du spectre en éliminant le signal en BFS inférieur à la fréquence de coupure choisie. Enfin, un filtre passe-bande va quant à lui permettre de conserver uniquement le signal concentré dans une bande de fréquence de coupure et d'atténuer le signal dont la fréquence spatiale est supérieure ou inférieure. A noter que la limite entre les BFS et les HFS est relative et le choix de la fréquence de coupure dépend des auteurs (e.g., 3 cpd pour DeValois & DeValois, 1988), mais fréquemment, on parle de BFS pour des cpd inférieurs à 2 et de HFS pour des cpd supérieurs à 6. Lors de ces manipulations, le spectre de phase n'est pas altéré afin de conserver la position relative des fréquences spatiales dans l'espace et la configuration spatiale de l'image. A l'aide de la transformée de Fourier inverse, il est également possible d'étudier la contribution du spectre de phase et d'amplitude d'une image. Par exemple, afin de conserver uniquement une information contenue dans le spectre de phase, on peut appliquer un spectre d'amplitude de bruit blanc à l'image. A l'inverse, afin de conserver uniquement une information contenue dans le spectre d'amplitude, on peut appliquer un spectre de phase de bruit blanc à l'image. La manipulation des fréquences spatiales dans une image permet d'étudier et de comprendre leurs rôles en perception visuelle.

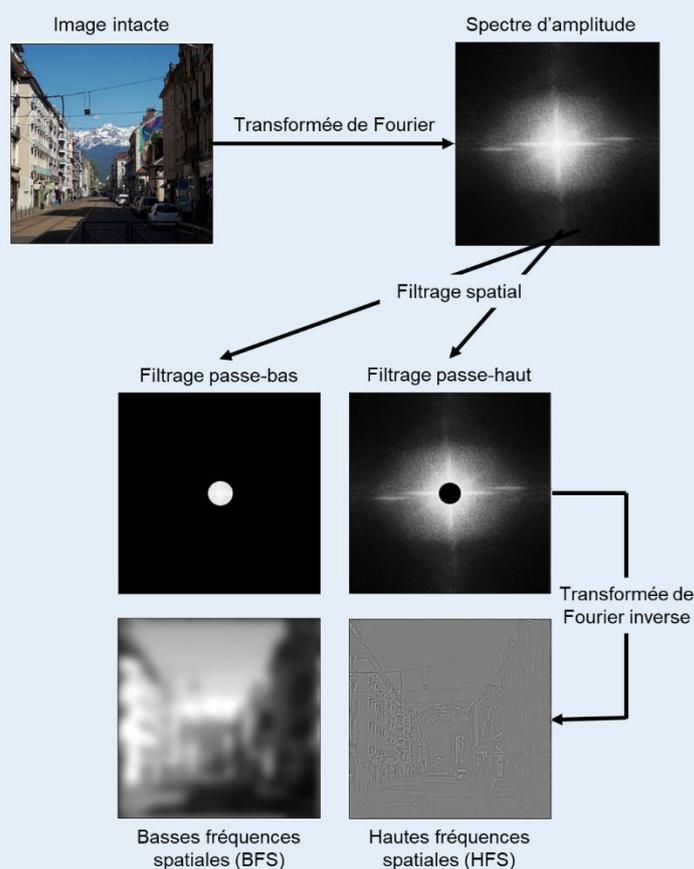


Figure B : Illustration de la procédure de filtrage d'une scène grâce à la transformée de Fourier.

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

1.1. De la rétine au cortex visuel primaire

La reconnaissance visuelle d'une scène ne se fait pas au niveau de l'œil. L'œil ne fait que capter des rayons lumineux qui proviennent de l'environnement. Cette lumière peut provenir directement d'une source lumineuse, comme le soleil, mais aussi de tous les objets qui nous entourent car ils reflètent la lumière. La rétine, qui se trouve au fond de l'œil, va ensuite transformer ces rayons lumineux en un message nerveux qui sera envoyé au cerveau et qui va donner un sens à cette lumière.

Dans cette partie du manuscrit, nous allons nous intéresser au traitement de l'information visuelle de la rétine au cerveau. La rétine, qui tapisse le fond de l'œil, se divise en deux parties : (1) la partie dite centrale de la rétine qui comprend la macula et la fovéa et qui s'étend jusqu'à 6 millimètres de diamètre et (2) la partie dite périphérique de la rétine qui comprend tout le reste et qui s'étend sur environ 17 millimètres. D'un point de vue fonctionnel, la rétine se compose de deux types de photorécepteurs situés dans sa couche la plus profonde : les cônes et les bâtonnets. Les photorécepteurs sont sensibles aux stimulations lumineuses (les photons). Les cônes, dont le nombre est d'environ 5 millions, sont des cellules sensibles aux informations chromatiques, mais sont très peu sensibles aux informations de luminance. Ils ont donc besoin d'une forte intensité lumineuse pour fonctionner correctement. On parle alors de vision photopique. Les bâtonnets, dont le nombre est d'environ 20 fois supérieur aux cônes, sont au contraire très sensibles aux informations de luminance mais très peu sensibles aux informations chromatiques. Ils permettent donc uniquement une vision en nuances de gris et ne fonctionnent que lorsque l'éclairage est faible. On parle ici de vision scotopique. La répartition des cônes et des bâtonnets n'est pas homogène sur la rétine. La densité des cônes est la plus élevée au niveau de la fovéa, tandis que la densité des bâtonnets augmente avec l'excentricité rétinienne (Curcio et al., 1990).

L'information lumineuse transformée en signal électrique par les photorécepteurs est ensuite véhiculée, par le biais des cellules bipolaires, jusqu'à la couche la plus superficielle de la rétine composée des cellules bipolaires. On peut distinguer trois types de cellules ganglionnaires présentant chacune des propriétés différentes : les cellules ganglionnaires magnocellulaires (M), les cellules ganglionnaires parvocellulaires (P) et les cellules koniocellulaires (K) que nous n'aborderons pas dans ce travail de thèse en raison de leur intérêt moindre. Les cellules ganglionnaires M sont également appelées 'parasol' en raison de leurs grands corps cellulaires et d'une arborisation dendritique étendue. Les axones de ces cellules sont larges et myélinisés permettant une transmission rapide de l'information visuelle. Elles sont sensibles aux contrastes de luminance et aux mouvements, mais peu sensibles aux informations chromatiques. Ces cellules possèdent un grand champ récepteur, c'est-à-dire que les réponses de nombreux photorécepteurs, positionnés sur une surface rétinienne relativement large, convergent vers ces cellules. Elles véhiculent donc une information de faible résolution spatiale (Kaplan, 2004 ; Kaplan & Shapley, 1986). On parle alors d'une sensibilité de ces cellules aux basses fréquences spatiales ou BFS (Curcio et al., 1990 ; Enroth-Cugell & Robson, 1966, 1984). A contrario, les cellules ganglionnaires P sont également appelées 'naines' en

raison de leurs petits corps cellulaires et d'une arborisation dendritique restreinte. Les axones de ces cellules étant fins et non myélinisés, leur transmission de l'information visuelle est plus lente. Elles sont sensibles aux informations chromatiques mais peu sensibles aux informations de luminance. Ces cellules possèdent un champ récepteur beaucoup plus petit que les cellules M, ce qui leur permet de véhiculer une information de très haute résolution spatiale (Kaplan, 2004 ; Kaplan & Shapley, 1986). On parle cette fois-ci d'une sensibilité de ces cellules aux hautes fréquences spatiales ou HFS (Curcio et al., 1990 ; Enroth-Cugell & Robson, 1966 ; Enroth-Cugell & Robson, 1984).

En sortie de la couche ganglionnaire, les axones des cellules ganglionnaires se rassemblent pour former le nerf optique. Ce nerf sensitif permet la conduction de l'information à destination des corps genouillés latéraux droit et gauche, deux noyaux situés dans la partie dorsale du thalamus. Avant de rejoindre les corps genouillés latéraux, les fibres du nerf optique 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 corps genouillé latéral gauche. 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 corps genouillé latéral droit. Ainsi, l'information issue d'un hémichamp visuel est envoyée au corps genouillé latéral controlatéral. Une fois arrivées aux corps genouillés latéraux, les caractéristiques de l'information transmise par les cellules ganglionnaires M et P sont préservées car elles se projettent respectivement vers les couches M et P des corps genouillés latéraux. A noter toutefois que si la majorité des fibres du nerf optique (90%) se projettent sur les différentes couches des corps genouillés latéraux, la minorité restante des fibres du nerf optique (10%) se projettent sur le colliculus supérieur, une structure sous-corticale impliquée essentiellement dans le contrôle des mouvements oculaires.

A la sortie des corps genouillés latéraux, l'information visuelle est véhiculée jusqu'au cortex visuel primaire (V1, ou cortex strié) le long de deux voies distinctes : la voie magnocellulaire et la voie parvocellulaire. Les voies magnocellulaire et parvocellulaire sont formées à partir des axones des neurones des couches M et P des corps genouillés latéraux, respectivement. Elles permettent d'acheminer, entre autres, les fréquences spatiales jusqu'à V1 (Bullier, 2001 ; Van Essen & DeYoe, 1995). Du fait des propriétés anatomiques et fonctionnelles des axones des cellules M et P, la voie magnocellulaire véhicule l'information visuelle plus rapidement que la voie parvocellulaire. Comme nous le verrons en détail dans la suite de ce manuscrit, des enregistrements neurophysiologiques réalisés sur le primate non humain (Bullier, 2001) montrent effectivement que les informations véhiculées par la voie magnocellulaire atteindraient le cortex visuel primaire 25 ms avant les informations véhiculées par la voie parvocellulaire. Des travaux neurophysiologiques menés auprès de participants humains en TMS (stimulation magnétique transcrânienne) confirment que la latence de la réponse de la voie magnocellulaire dans le cortex visuel primaire est de 25 à 45 ms plus précoce que la latence de la réponse de la voie parvocellulaire (Laycock et al., 2007). Il y aurait donc un avantage temporel du signal magnocellulaire qui aurait des implications sur la vitesse de traitement des informations en fréquences spatiales. Une information de basse résolution spatiale, donc en BFS, serait véhiculée plus rapidement jusqu'au cortex visuel primaire par la

voie magnocellulaire que l'information de très haute résolution spatiale, donc en HFS, par la voie parvocellulaire.

Le système visuel humain contient donc dès la rétine des cellules spécialisées dans le traitement des fréquences spatiales. Par ailleurs, une caractéristique du traitement rétino-cortical qui est particulièrement importante pour ce travail de thèse concerne la distinction des propriétés fonctionnelles du système visuel entre vision centrale et vision périphérique. Concrètement, le système visuel ne traite pas les mêmes caractéristiques visuelles suivant l'excentricité rétinienne du signal visuel. En effet, la répartition des photorécepteurs et des cellules ganglionnaires n'est pas homogène sur la rétine. Nous avons vu que la densité des cônes est la plus élevée au niveau de la fovéa, tandis que la densité des bâtonnets augmente avec l'excentricité rétinienne. De même, au niveau de la fovéa, on trouve majoritairement des cellules ganglionnaires P à petits champs récepteurs, tandis qu'au niveau de la rétine périphérique, on trouve majoritairement des cellules ganglionnaires M à grands champs récepteurs. Ces propriétés anatomiques expliqueraient (en partie, puisque nous verrons plus bas que le traitement cortical y contribue également) la baisse de l'acuité visuelle à mesure que l'on s'éloigne du centre de la rétine (Johnson, 2010). Du fait de cette répartition non homogène des différents types de photorécepteurs et des cellules ganglionnaires sur la rétine et de leurs spécificités, nous pouvons faire l'hypothèse que la rétine centrale (fovéa) serait particulièrement bien équipée pour encoder les HFS, tandis que la rétine périphérique encoderait plutôt les BFS. Nous avons ensuite vu que l'information visuelle en BFS est véhiculée plus rapidement via la voie magnocellulaire à travers le cortex que l'information en HFS via la voie parvocellulaire. Ainsi, les BFS issues de la rétine périphérique seraient envoyées plus rapidement au cortex visuel que les HFS issues de la rétine centrale. Ces propriétés sont fondamentales pour comprendre la problématique générale de la thèse que nous détaillerons dans la partie 5 de ce chapitre.

1.2. Le cortex visuel primaire et le phénomène de magnification corticale

Au niveau du cortex visuel primaire, la distinction entre les informations en BFS et les informations en HFS est conservée car le cortex possède lui aussi des cellules spécialisées dans le traitement de différentes fréquences spatiales (Poggio, 1972 ; DeValois, Albrecht & Thorell, 1982 ; DeValois, Yund & Hepler, 1982). De plus, comme les corps genouillés latéraux, le cortex visuel primaire conserve une représentation topographique du champ visuel depuis sa projection sur la rétine (pour des études en neurophysiologie sur le primate non humain : Daniel & Whitteridge, 1961 ; pour des études en IRM sur l'humain : Engel et al., 1994 ; Engel, Glover & Wandell, 1997 ; Sereno, McDonald & Allman, 1994 ; Tootell et al., 1982 ; Tootell et al., 1988 ; Wandell, Dumoulin & Brewer, 2007 ; Wandell & Winawer, 2011). Des points voisins de la rétine se projettent sur des régions voisines du cortex visuel primaire. On parle alors d'une organisation rétinotopique (Figure 1.1). Plus précisément, les informations en provenance du champ visuel droit se projettent sur le cortex visuel primaire de l'hémisphère gauche, et les informations en provenance du champ visuel gauche se projettent sur le cortex visuel primaire de l'hémisphère droit. On parle alors d'un principe de contralatéralité de la projection des

informations du champ visuel sur le cortex. Par ailleurs, pour chaque hémichamp visuel, le quart de champ visuel supérieur est représenté dans la partie inférieure de la scissure calcarine et le quart de champ visuel inférieur est représenté dans la partie supérieure. Le champ visuel fovéal (ou central) se projette sur la surface ventrolatérale de la partie postérieure du lobe occipital. Avec l'augmentation de l'excentricité rétinienne, le champ visuel se projette progressivement sur la surface médiane de la partie plus antérieure du lobe occipital. Cette projection est toutefois loin d'être uniforme. Le signal visuel provenant de la fovéa subit une magnification au niveau du cortex visuel. Cela signifie qu'un grand nombre de cellules du cortex visuel primaire est consacré au traitement des informations visuelles en provenance de la partie centrale de la rétine tandis qu'un plus petit nombre de cellules est dédié à la vision plus périphérique. Plus précisément, la partie fovéale de la rétine, qui représente seulement 1% de la surface totale de la rétine, est représentée sur 50% du cortex visuel primaire. Ce phénomène de magnification corticale de la fovéa contribue également à la baisse d'acuité visuelle en vision périphérique. Il ne doit pas être sous-estimé dans ce travail de thèse où nous étudions comment la vision centrale et la vision périphérique contribuent respectivement à la reconnaissance visuelle de scènes.

Par ailleurs, au niveau du cortex visuel primaire, plusieurs études en neurophysiologie menées chez le chat et le primate non-humain (Everson et al., 1998 ; Issa, Trepel & Stryker, 2000 ; Foster et al., 1985), ainsi que plusieurs études en IRM conduites chez l'humain (Sasaki et al., 2001 ; Henriksson et al., 2008 ; Musel et al., 2013 ; Xu, Anderson & Casagrande, 2007) mettent en évidence un traitement rétinotopique des fréquences spatiales. Par exemple, Musel et al. (2013) ont montré dans une étude en IRMf que la catégorisation de grandes scènes filtrées en HFS activait plus fortement les aires du cortex occipital correspondant à la projection de la vision centrale (partie latérale postérieure du cortex visuel), tandis que des grandes scènes filtrées en BFS activaient plus fortement les aires du cortex occipital correspondant à la projection du champ visuel périphérique (partie médiale antérieure du cortex visuel). Ce type de résultat supporte notre hypothèse d'une rétine centrale plutôt sélective aux HFS et d'une rétine périphérique plutôt sélective aux BFS.

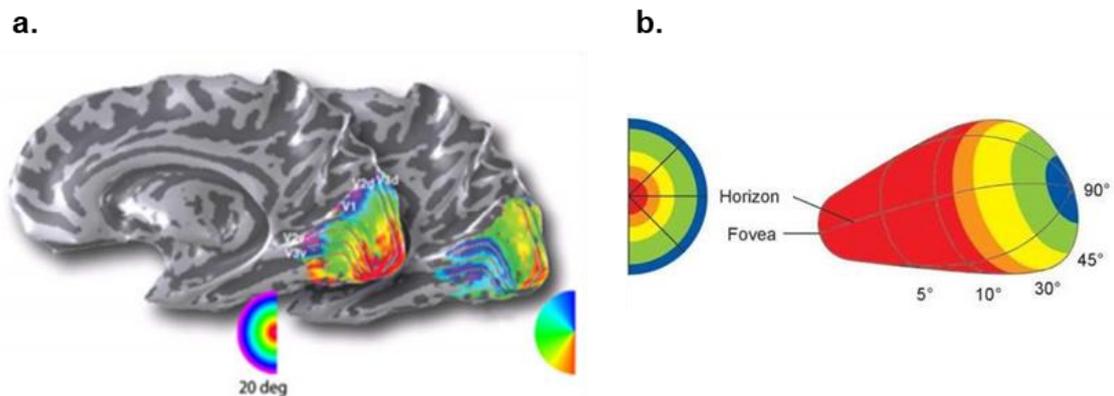


Figure 1.1 : (a) Organisation rétinotopique des aires visuelles (V1, V2 et V3). Les couleurs indiquent la correspondance entre la position dans le champ visuel (i.e. l'excentricité rétinienne par rapport à la fovéa) et la projection sur le cortex visuel. Figure adaptée de Wandell, Dumoulin & Brewer (2007). (b) Illustration de la magnification corticale dans V1. La partie fovéale de la rétine (en rouge), qui représente seulement 1% de la surface totale de la rétine (à gauche), est représentée sur 50% du cortex visuel primaire (à droite). Figure adaptée de Hubel (1994).

Il est également intéressant de souligner que Musel et al. (2013) ont également mis en évidence des différences interhémisphériques lors du traitement des fréquences spatiales contenues dans des scènes. En comparant les activations entre les deux hémisphères cérébraux pour chaque type de scènes filtrées (BFS et HFS) séparément, les auteurs ont observé une activation plus importante du gyrus occipital moyen dans l'hémisphère droit que dans l'hémisphère gauche lors de la catégorisation de scènes en BFS, et une activation plus importante du gyrus temporal moyen dans l'hémisphère gauche que dans l'hémisphère droit lors de la catégorisation des scènes en HFS. Cette spécialisation hémisphérique du traitement des fréquences spatiales est également à prendre en considération lorsqu'on cherche à étudier le rôle de la vision périphérique dans la reconnaissance visuelle, puisqu'il pose la question de différences de traitement de l'information visuelle entre l'information périphérique droite et gauche. Nous reviendrons sur la spécialisation hémisphérique des fréquences spatiales dans la partie [4.4](#) de ce chapitre théorique.

1.3. Les voies visuelles corticales

A partir du cortex visuel primaire (V1), l'information visuelle est projetée vers les aires visuelles extra-striées, dont les principales sont les aires V2, V3, V4 et V5 (aussi appelée aussi MT). Ces aires sont spécialisées dans l'extraction de caractéristiques élémentaires spécifiques du signal visuel (e.g., les contours, les couleurs, l'orientation, les mouvements ; Wandell et al., 2007). En sortie de V1, l'information visuelle est transmise à l'aire V2, à partir de laquelle on observe une séparation des traitements en deux voies corticales possédant des propriétés fonctionnelles distinctes : la voie ventrale, qui porte une information concernant la forme et l'identité du signal visuel (la voie du "quoi") et la voie dorsale, qui porte une information sur

la structure spatiale et les mouvements du signal visuel (la voie du “où” ; de Haan & Cowey, 2011 ; Goodale & Milner, 1992 ; Haxby et al., 1991 ; Ungerleider & Haxby, 1994).

La voie ventrale (voie du “quoi”) véhicule l’information jusqu’aux cortex inférotemporal (IT) en traversant les aires V1, V2, V3 et V4. L’organisation rétinotopique reste présente dans V1 et V2, mais disparaît progressivement à partir de V4. Par ailleurs, les champs récepteurs des neurones s’agrandissent à mesure qu’on avance dans la hiérarchie corticale de cette voie. Par exemple, les champs récepteurs des neurones du cortex inférotemporal sont 100 fois plus grands que ceux de V4, qui sont eux-mêmes 30 fois plus grands que ceux de V1 (Gross, Bender & Rocha-Miranda, 1969). Les neurones de l’aire V4 sont plutôt sélectifs aux informations de chrominance et les neurones du cortex inférotemporal sont plutôt sélectifs aux formes complexes et répondent à la combinaison de formes et de couleurs (Tanaka, Saito, Fukada & Moriya, 1991 ; Tanaka, 1993, 1996). Les aires de cette voie sont donc principalement impliquées dans le traitement des informations de chrominance et des formes. Une lésion de la voie ventrale est connue pour impliquer des déficits pour la reconnaissance (Rubens & Benson, 1971), telle que celle des objets (on parle d’agnosie visuelle d’objets), des couleurs (achromatopsie) ou des visages (prosopagnosie). Par ailleurs, la voie ventrale contient des projections parvocellulaires et, en plus en faible proportion, des projections magnocellulaires (Merigan & Maunsell, 1993). Elle répond donc surtout aux informations en HFS, mais aussi aux BFS.

La voie dorsale (voie du “où”) véhicule quant à elle l’information jusqu’aux cortex pariétal postérieur, en traversant les aires V1, V2, V3 et V5. La taille des champs récepteurs augmente dans la voie dorsale de la même façon que pour la voie ventrale (Raiguel et al., 1997). Les neurones de l’aire V3 présentent une forte sensibilité au contraste et les neurones de l’aire V5 sont sélectifs aux mouvements (Maunsell & Van Essen, 1983a). Les aires de cette voie sont donc principalement impliquées dans le traitement des relations spatiales entre les objets et du mouvement (Goodale & Milner, 1992). Une lésion de cette voie impliquerait principalement des déficits de ces fonctions appelés agnosies spatiales (Benson & Greenberg, 1969). De plus, la voie dorsale contenant des projections majoritairement magnocellulaires, elle répond principalement aux BFS et sa latence de réponse est plus courte que celle de la voie ventrale.

Les aires du système visuel seraient donc scindées en deux voies avec des fonctions bien distinctes. En ce sens, Livingstone et Hubel (1988) ont suggéré que les voies dorsale et ventrale seraient respectivement les extensions des voies magnocellulaire et parvocellulaire. Cette proposition a toutefois largement été controversée pour deux raisons. Premièrement, dans la réalité clinique, les lésions apparentées à la voie dorsale ou à la voie ventrale n’entraînent pas nécessairement des déficits caractéristiques des fonctions respectives de ces deux voies (Merigan et al., 1991 ; Schiller et al., 1990). Deuxièmement, des connexions ont été retrouvées entre ces deux voies, suggérant qu’elles ne fonctionnent pas de manière totalement indépendante (Merigan & Maunsell, 1993). 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 les 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). Ces connexions expliqueraient pourquoi, en pratique clinique, on observe des déficits bien plus complexes que ceux propres à chaque voie.

En dépit de la complexité des connexions corticales, il ressort de ces travaux que l'information corticale magnocellulaire est véhiculée plus rapidement que l'information parvocellulaire. Cette propriété fonctionnelle propre au système visuel amène à l'hypothèse fondamentale d'une précédece temporelle du traitement des caractéristiques visuelles véhiculées par la voie magnocellulaire et donc, des BFS. L'hypothèse d'une précédece temporelle du traitement des BFS est au cœur même des modèles actuels de la perception visuelle de scènes. Dans la partie qui suit, nous allons voir les études psychophysiques réalisées chez l'humain qui viennent supporter l'hypothèse d'une précédece temporelle des BFS et nous verrons ensuite une proposition de modèle de catégorisation de scènes qui en découle : le modèle *coarse-to-fine*.

2. Traitement des fréquences spatiales et modèle cognitif *coarse-to-fine* de la perception de scènes visuelles

2.1. Démonstration psychophysique de canaux fréquentiels et d'une précédece temporelle du traitement des basses fréquences spatiales

Les travaux que nous avons précédemment présentés sur la sélectivité des cellules du système visuel aux fréquences spatiales concernent pour la majorité des enregistrements neurophysiologiques sur l'animal. Dans les années 60, Fergus Campbell et ses collaborateurs (Campbell & Robson, 1968 ; Blakemore & Campbell, 1969) apportèrent des preuves supplémentaires de cette sélectivité en l'étudiant de façon comportementale chez l'humain et ce, grâce à des méthodes originales issues de la psychophysique. Leur travail prend sa source dans toute une série de travaux qui avait pour objectif d'étudier la capacité de l'être humain à discerner les contrastes, c'est-à-dire les différences de luminance entre des surfaces adjacentes. On parle alors d'évaluation de la sensibilité au contraste. Cette évaluation est fondamentale pour diagnostiquer des déficiences visuelles. Une méthode psychophysique d'évaluation de la sensibilité au contraste d'un observateur consiste à mesurer sa capacité à percevoir le contraste minimal d'un réseau. Un réseau ressemble à une alternance progressive de bandes sombres et claires dont la largeur définit la fréquence spatiale et la différence de luminance définit le contraste. Un réseau est dit sinusoïdal lorsqu'il alterne progressivement entre les bandes sombres et claires, ou carré lorsque le signal alterne régulièrement et instantanément entre les bandes sombres et claires (Figure [1.2a](#)). Dans la pratique, on fait varier le contraste d'un réseau d'une fréquence spatiale donnée jusqu'à atteindre la valeur de contraste minimum (seuil de visibilité). On répète cette opération pour des réseaux de différentes fréquences spatiales. On établit alors une fonction de sensibilité au contraste ou FSC (Figure [1.2b](#)) où les fréquences spatiales sont reportées sur l'axe des abscisses et la sensibilité au contraste (valeur inverse du seuil de contraste) sur l'axe des ordonnées. Pour un observateur normal face à des réseaux sinusoïdaux statiques présentés en vision centrale, les travaux de DeValois et DeValois (1988) ont montré que la sensibilité au contraste est la meilleure pour les fréquences spatiales

comprises entre 2 et 6 cpd. En d'autres termes, la capacité de l'être humain à distinguer des changements minimes entre des bandes sombres et claires est la meilleure pour ces valeurs de fréquences spatiales. On observe une diminution rapide de la sensibilité au contraste pour les fréquences spatiales plus hautes et une diminution plus lente pour les fréquences spatiales plus basses.

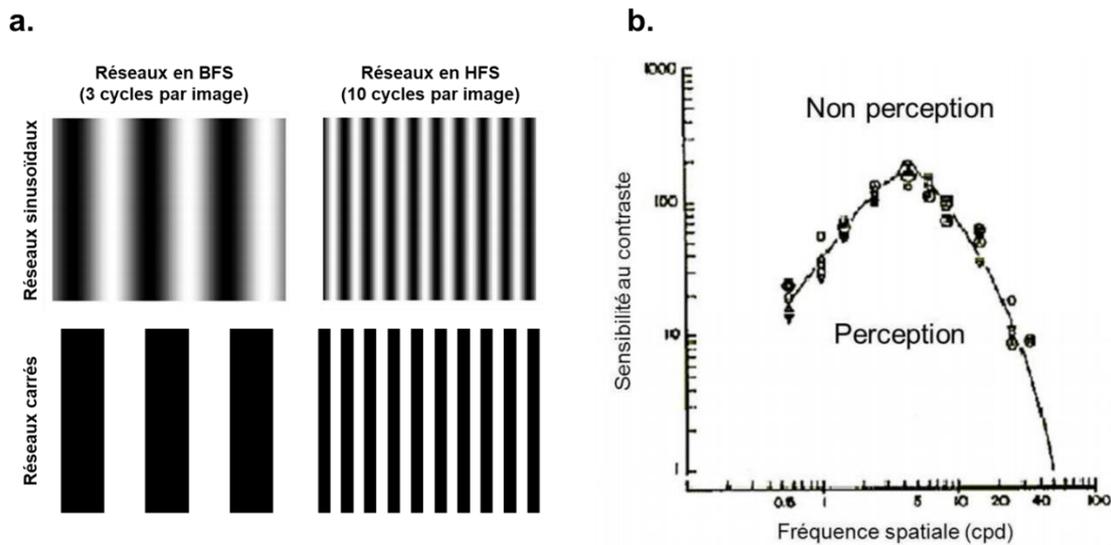


Figure 1.2 : (a) Exemples de réseaux sinusoidaux et carrés en basses fréquences spatiales (BFS, 3 cycles par image) et en hautes fréquences spatiales (HFS, 10 cycles par image). (b) Fonction de sensibilité au contraste (FSC) chez un observateur normal face à un réseau sinusoidal statique présenté à différentes fréquences spatiales en vision centrale. Les fréquences spatiales en cycle par degré (cpd) sont représentées en abscisse et la sensibilité au contraste est représentée en ordonnée. 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. Figure adaptée de De Valois et De Valois (1988).

Selon Campbell et Robson (1968), cette fonction ne refléterait pas un seul mécanisme perceptif, mais la combinaison de plusieurs mécanismes indépendants qu'ils appellent des "canaux fréquentiels", chaque canal étant sensible à une bande limitée de fréquences spatiales. Cette hypothèse a été testée par Blakemore et Campbell (1969) à l'aide d'une méthode psychophysique d'adaptation sélective. Cette méthode consiste à présenter de façon prolongée un réseau d'une fréquence spatiale donnée afin de diminuer la sensibilité du système visuel à cette fréquence spatiale (et d'augmenter le seuil de sensibilité au contraste). Si la présentation ultérieure d'un réseau d'une autre fréquence spatiale n'est pas affectée par cette adaptation, on parle alors d'adaptation sélective. Après avoir adapté des observateurs à un réseau sinusoidal d'une fréquence spatiale donnée, les auteurs ont observé une élévation du seuil de sensibilité au contraste uniquement pour des réseaux sinusoidaux dont la fréquence spatiale était proche de celle du réseau d'adaptation. Ils ont ainsi pu identifier des bandes de fréquences spatiales adaptées où la baisse de sensibilité était centrée sur la fréquence d'adaptation (Figure 1.3) et qu'on appelle des canaux fréquentiels. Ainsi, l'évaluation de la sensibilité au contraste permet d'appréhender l'ensemble des canaux fréquentiels et la FSC ne représenterait en fait que

l'enveloppe de ces canaux qui se chevaucheraient. Cette sensibilité au contraste est désormais systématiquement testée lors de bilans ophtalmologiques à l'aide de tests plus simples où on évalue, par exemple, la sensibilité au contraste pour une fréquence spatiale donnée (cas de l'échelle de Pelli-Robson ; Pelli et al., 1988) ou l'acuité visuelle pour différents contrastes donnés (cas de l'échelle de Sloan ; Sloan, 1959).

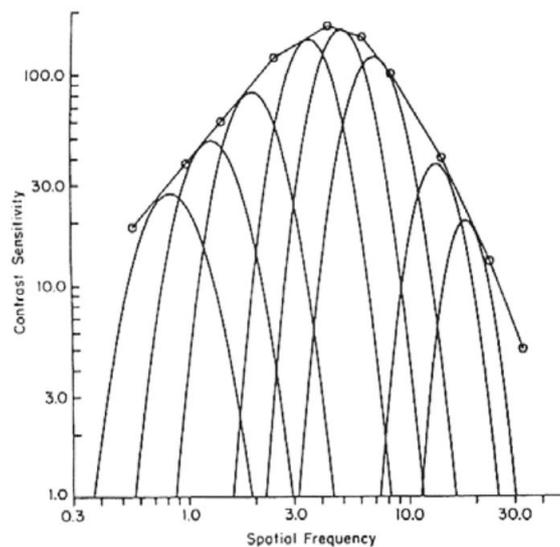


Figure 1.3 : Fonctions de sensibilité au contraste qui représente les différents canaux fréquentiels sélectifs à une bande de fréquences spatiales. Figure adaptée de De Valois et De Valois (1988).

Une particularité très importante du système visuel est que la sensibilité au contraste varie en fonction de l'excentricité rétinienne. En effet, plus l'excentricité rétinienne augmente et plus la sensibilité au contraste diminue (Figure 1.4). Cette diminution est en fait liée à la répartition non homogène des cônes et des bâtonnets sur la rétine dont nous avons parlé précédemment. La densité des cônes est la plus élevée au niveau de la fovéa (au centre de la rétine), tandis que la densité des bâtonnets augmente avec l'excentricité rétinienne (Curcio et al., 1990). Cette répartition explique en partie la baisse de l'acuité visuelle, et donc la baisse de sensibilité au contraste, à mesure que l'on s'éloigne du centre de la rétine. Dans ce contexte, Rovamo, Virsu et Näsänen (1978) ont déterminé la FSC en fonction de la fréquence spatiale et de l'excentricité rétinienne. Ils ont montré que la sensibilité au contraste diminuait rapidement avec l'excentricité rétinienne. Avec un contraste maximum, la plus haute fréquence spatiale perçue passe d'environ 60 cpd en vision centrale à seulement 2 cpd en vision périphérique à 30° d'excentricité rétinienne. Ainsi, la visibilité des fréquences spatiales est étroitement liée au contraste et à l'excentricité rétinienne. Nous devons donc considérer la FSC lorsque l'on s'intéresse aux traitements visuels réalisés en vision centrale et en vision périphérique.

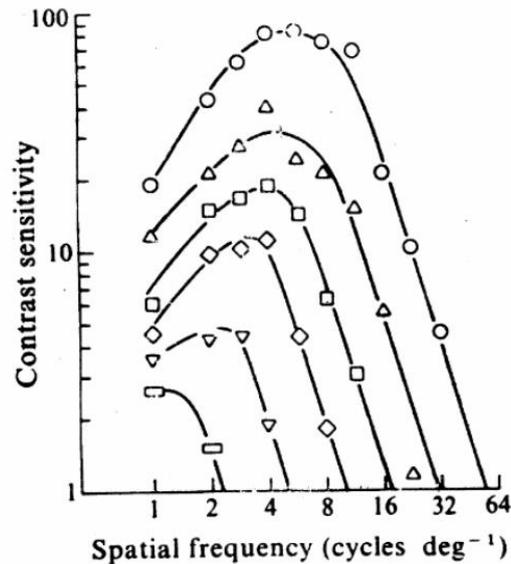


Figure 1.4 : Fonction de sensibilité au contraste chez un observateur normal en fonction de la fréquence spatiale et de l'excentricité rétinienne. Les courbes sont associées aux différentes excentricités en degrés d'angle visuel, de 0° à 30° (O : 0° ; Δ : 1,5° ; □ : 4° ; ◇ : 7,5° ; ▽ : 14° ; □ : 30°). La plus haute fréquence spatiale perçue en cycles par degré (cpd ; la valeur d'intersection pour chaque courbe sur l'axe des abscisses) varie de 64 cpd en présentation fovéale et à seulement 2 cpd en présentation périphérique à 30° d'excentricité. Figure adaptée de Rovamo, Virsu et Näsänen (1978).

Après avoir mis en évidence l'existence de canaux fréquentiels, la question des propriétés temporelles de ces canaux s'est très rapidement posée. Breitmeyer (1975) a par exemple mesuré les temps de réaction pour détecter des réseaux sinusoïdaux à différentes fréquences spatiales et a montré qu'un réseau sinusoïdal en BFS (0,5 cpd) était détecté plus rapidement qu'un réseau sinusoïdal en HFS (11 cpd). Ces résultats psychophysiques viennent supporter ceux des études neurophysiologiques que nous avons présentées dans la partie 1 de ce chapitre et qui suggèrent que les BFS sont véhiculées plus rapidement jusqu'au cortex visuel par la voie magnocellulaire que les HFS par la voie parvocellulaire. Par la suite, de nombreuses études comportementales réalisées sur la perception de formes visuelles globales et locales ont permis de démontrer de façon indirecte une préférence temporelle du traitement des BFS. Ces études sont nées au début du XX^{ème} siècle avec la théorie de la Gestalt. Selon cette théorie, le tout serait supérieur à la somme de ses parties. Autrement dit, la perception d'une forme dans son ensemble domine par rapport à la perception des parties qui la composent. La théorie de la Gestalt a été expérimentalement testée à l'aide du paradigme des formes hiérarchiques. Dans ce paradigme, le stimulus est une forme globale composée d'un groupement de formes locales identiques, qui peuvent être soit congruentes avec la forme globale (la même forme), soit incongruentes (une forme différente). Par exemple, Navon (1977) utilisait une grande lettre globale (e.g., un grand "H"), composée de plusieurs petites lettres locales (e.g., des petits "h" ou des petits "s" selon la condition de congruence, Figure 1.5). La tâche des participants était d'identifier le plus rapidement possible la forme au niveau global ou au niveau local. Ce paradigme montre généralement que l'identification de la forme globale est plus rapide que l'identification des formes locales, que celles-ci soient congruentes ou non avec la forme globale. Ces résultats sont cohérents avec la théorie de la Gestalt et suggèrent que le tout (la

forme globale) est traitée automatiquement et plus rapidement que la somme de ses parties (formes locales). On parle usuellement d'un effet de précedence globale pour décrire ce résultat. Par ailleurs, ce paradigme montre que l'identification des formes locales incongruentes avec la forme globale est plus longue que celle des formes locales congruentes. L'information globale interfère donc sur le traitement de l'information locale soit en facilitant son traitement (dans la condition congruente) soit en gênant son traitement (dans la condition incongruente). Par contre, l'identification de la forme globale reste la même que les formes locales soient congruentes ou incongruentes. On parle dans ce cas d'un effet d'interférence globale. L'effet de précedence globale n'est pas sans rappeler l'hypothèse d'une précedence temporelle du traitement des BFS par rapport au traitement des HFS. En effet, les BFS fournissent une information globale concernant l'input visuel (les formes globales), alors que les HFS fournissent une information sur les détails (les formes locales).

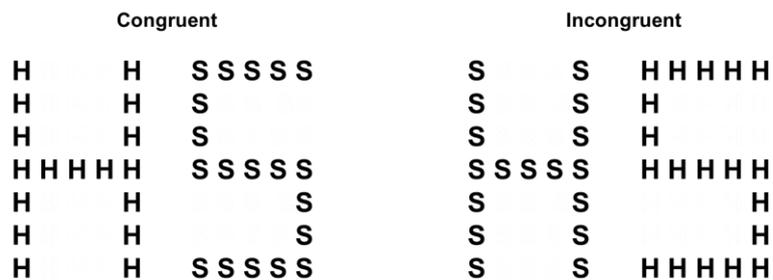


Figure 1.5 : Exemple de formes hiérarchiques utilisées par Navon (1977). A gauche, la forme globale (un grand H et un grand S) est congruente avec les formes locales. A droite, la forme globale (un grand H et un grand S) est incongruente avec les formes locales (des petits S et des petits H).

Plusieurs psychophysiciens ont d'ailleurs cherché à expliciter le lien entre l'effet de précedence globale et le traitement des BFS (Badcock et al., 1990 ; Lamb & Yund 1993 ; Shulman et al., 1986). Par exemple, Shulman et al. (1986) ont tout d'abord montré un lien entre l'identification des formes globale/locale et le traitement des BFS/HFS en utilisant un paradigme psychophysique d'adaptation visuelle. Dans une première session d'adaptation, des réseaux sinusoïdaux contenant soit des BFS, soit des HFS étaient présentés de façon répétitive et prolongée afin de diminuer sélectivement la sensibilité du système visuel à une bande de fréquences spatiales. Dans une deuxième session, les participants devaient identifier le plus rapidement possible l'orientation d'une lettre au niveau global ou au niveau local. Les résultats ont montré que l'adaptation à des réseaux sinusoïdaux en BFS perturbait la détection de l'orientation d'une lettre au niveau global, tandis que l'adaptation à des réseaux sinusoïdaux en HFS perturbait la détection de l'orientation d'une lettre au niveau local. Ce résultat suggère que le niveau global de l'information visuelle serait contenu dans les BFS du signal visuel alors que le niveau local serait contenu dans les HFS. Dans le même contexte, Badcock et al. (1990) ont directement manipulé le contenu en fréquences spatiales des formes hiérarchiques. Ils présentaient à des participants des formes hiérarchiques qui étaient soit non filtrées (toutes les fréquences spatiales étaient disponibles), soit filtrées en HFS (en utilisant un filtre passe-haut

qui retire le signal en BFS). Lorsque les formes n'étaient pas filtrées, les auteurs observaient un effet de précedence globale. La lettre globale était identifiée plus rapidement que les petites lettres locales. Par contre, lorsque la forme hiérarchique était filtrée en HFS, l'effet de précedence globale n'était plus observé. Ce résultat suggère que l'effet de précedence globale reposerait sur le traitement rapide des BFS.

Dans l'ensemble, ces études viennent asseoir l'hypothèse d'une précedence temporelle du traitement des BFS. Qui plus est, ce phénomène est cohérent avec les propriétés anatomiques et fonctionnelles du système visuel où les informations magnocellulaires, comme les BFS, sont véhiculées plus rapidement que les informations parvocellulaires, comme les HFS (Bullier, 2001 ; Van Essen & Deyoe, 1995 ; Laycock et al., 2007). Dans ce contexte, un modèle influent de la perception de scènes proposé pour la première fois par Schyns et Oliva (1994) postule que, par défaut, l'analyse visuelle d'une scène reposerait d'abord sur le traitement rapide des BFS, puis viendrait ensuite celui des HFS. Les auteurs parlent alors d'une séquence de traitement *coarse-to-fine*. Plus précisément, l'analyse rapide de l'information grossière *coarse* portée par les BFS permettrait une première reconnaissance de la scène visuelle qui serait ensuite affinée par l'analyse d'une information plus détaillée *fine* portée par les HFS. Nous allons voir dans la partie qui suit les différents travaux en faveur d'un modèle *coarse-to-fine* de la perception de scènes visuelles.

2.2. Modèle cognitif *coarse-to-fine* de la perception de scènes visuelles

Comme nous venons de le voir, les premiers travaux mettant en évidence la précedence du traitement des BFS par rapport à celui des HFS sont issus de données psychophysiques utilisant des réseaux sinusoïdaux et des formes hiérarchiques comme stimuli. Par la suite, les recherches menées en psychologie cognitive ont exploité ce phénomène afin de modéliser la reconnaissance de stimuli visuels, tels que des scènes complexes, des objets ou des visages. Plus précisément, les propriétés temporelles du traitement des fréquences spatiales ont amené Schyns et Oliva (1994) à proposer l'hypothèse d'une stratégie d'analyse *coarse-to-fine* de l'information visuelle contenue dans des scènes. Les auteurs postulent que l'analyse rapide de l'information grossière *coarse* portée par les BFS permettrait une première reconnaissance de la scène qui serait ensuite affinée par l'analyse d'une information plus détaillée *fine* portée par les HFS. Afin de tester l'hypothèse d'une analyse *coarse-to-fine*, Schyns et Oliva (1994) ont utilisé des images hybrides, c'est-à-dire une image construite en superposant une scène en BFS (en appliquant un filtre passe-bas ; cf. **Encart 1**) à une autre scène en HFS (en appliquant filtre passe-haut ; cf. **Encart 1**). Les deux scènes appartenaient à deux catégories sémantiques différentes. Par exemple, une image hybride pouvait représenter une scène de ville en BFS superposée à une scène de montagne en HFS (Figure 1.6). Une image hybride était présentée en amorce pendant 30 ms ou 150 ms, puis une scène cible non filtrée apparaissait pendant 100 ms. Dans la moitié des essais, la scène cible était en fait la version non filtrée de la scène en BFS ou de la scène en HFS de l'image hybride. Dans l'autre moitié des essais, la scène cible était une nouvelle scène (c'est à dire qu'elle n'était pas présentée au préalable dans l'image hybride, ni dans la scène en BFS, ni dans la scène en HFS). La présentation des images hybrides

était suffisamment courte pour que les participants ne réalisent pas que ces images étaient en fait composées de deux scènes différentes. La tâche des participants était de déterminer si la scène cible non filtrée était présente dans l'image hybride en amorce. Les résultats ont montré que lorsque la durée de présentation des images hybrides était très courte (30 ms), leur perception était alors dominée par l'information en BFS. En d'autres termes, les participants avaient plus tendance à dire que la scène cible était présente dans l'image hybride lorsqu'elle correspondait à la scène en BFS de l'image hybride que lorsqu'elle correspondait à l'image en HFS. Lorsque la durée de présentation des images hybrides était plus longue (150 ms), leur perception était cette fois-ci dominée par les informations en HFS. En conclusion, le contenu en fréquences spatiales préférentiellement utilisé en vue de la reconnaissance visuelle variait selon la durée de présentation de l'image hybride. Ces résultats suggèrent que les BFS seraient utilisées en priorité pour des traitements rapides, tandis que les HFS prendraient le relais lorsqu'un traitement plus long est possible.

Dans une deuxième expérience, ces mêmes auteurs ont testé l'hypothèse d'une stratégie d'analyse *coarse-to-fine* à l'aide de stimuli stimulant directement cette analyse. Chaque stimulus était cette fois-ci une séquence de deux images hybrides présentées rapidement l'une après l'autre (45 ms chacune). Chaque séquence était construite de façon à ce que les BFS de la première image hybride et les HFS de la deuxième image hybride appartiennent à la même scène (par exemple, une ville), et à ce que les HFS de la première image hybride et les BFS de la deuxième image hybride appartiennent à autre scène. Les participants n'étaient pas informés de cette manipulation et leur tâche consistait à catégoriser chaque séquence selon quatre catégories de niveau de base (salon, chambre, ville et autoroute). La réponse des participants permettait ainsi de déterminer s'ils avaient catégorisé l'image hybride en suivant une séquence *coarse-to-fine* ou *fine-to-coarse*. En effet, si les participants donnaient la catégorie de la scène en BFS de la première image hybride et celle en HFS de la deuxième image hybride, alors ils avaient traité le stimulus selon une séquence *coarse-to-fine*. A l'inverse, si les participants donnaient la catégorie de la scène en HFS de la première image hybride et celle en BFS dans la deuxième image hybride, alors ils avaient traité le stimulus selon une séquence *fine-to-coarse*. Les résultats ont montré que 67% des séquences étaient catégorisées selon une séquence de traitement *coarse-to-fine* et seulement 29% selon une séquence *fine-to-coarse* (les 4% restants correspondaient à des erreurs de catégorisation).

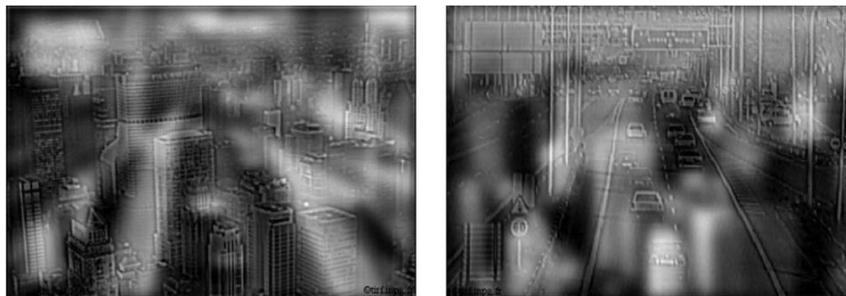


Figure 1.6 : Exemple d'images hybrides utilisées par Schyns et Oliva (1994). A gauche, l'image hybride est composée d'une scène de ville en HFS et d'une scène d'autoroute en BFS. A droite, l'image hybride est composée d'une scène d'autoroute en HFS et d'une scène de ville en BFS. Adapté de Schyns et Oliva (1994).

Par la suite, l'équipe de Carole Peyrin au Laboratoire de Psychologie et NeuroCognition (LPNC) a également testé l'hypothèse d'une analyse *coarse-to-fine* à l'aide de séquences de scènes filtrées (Kauffmann, Chauvin, Guyader & Peyrin, 2015 ; Kauffmann, Chauvin, Pichat & Peyrin, 2015 ; Musel et al., 2012 ; Musel et al., 2014 ; Peyrin et al., 2005 ; Peyrin et al., 2010). Par exemple, Musel et al. (2012) ont utilisé des séquences dynamiques de six images d'une même scène filtrée en différentes fréquences spatiales (Figure 1.7). Les images étaient assemblées en allant soit des BFS au HFS (afin d'imiter explicitement une séquence d'analyse *coarse-to-fine*), soit des HFS au BFS (imitant une séquence d'analyse *fine-to-coarse*). Les participants devaient décider si la scène appartenait à la catégorie des scènes d'intérieur ou à la catégorie des scènes d'extérieur. Les résultats ont montré que les participants catégorisaient la scène plus rapidement lorsque la présentation de la séquence imitait une analyse *coarse-to-fine* que lorsqu'elle imitait une analyse *fine-to-coarse*, suggérant ainsi qu'une analyse successive des fréquences spatiales des BFS aux HFS permettrait une reconnaissance plus rapide et efficace d'une scène.

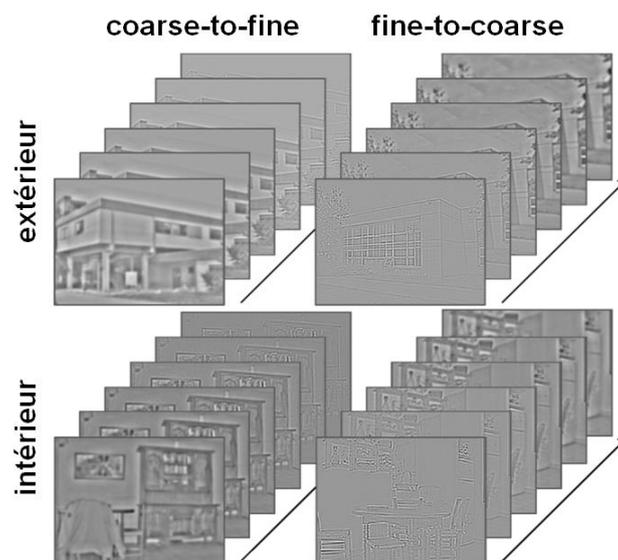


Figure 1.7 : Exemple de séquences dynamiques utilisées par Musel et al. (2012). A gauche, les images étaient assemblées en allant des BFS au HFS afin d'imiter explicitement une séquence d'analyse *coarse-to-fine*. A droite, les images étaient assemblées en allant des HFS au BFS afin d'imiter explicitement une séquence d'analyse *fine-to-coarse*. Figure adaptée de Musel et al. (2012).

Ce résultat a été répliqué plusieurs fois par cette même équipe de recherche avec les mêmes stimuli (Musel et al., 2014 ; Kauffmann, Chauvin, Guyader & Peyrin, 2015 ; Kauffmann, Chauvin, Pichat & Peyrin, 2015). Pourtant, il contraste quelque peu avec les travaux de De Cesarei et Codispoti (2011 ; voir aussi De Cesarei, Mastroia & Codispoti, 2013) qui ont montré que la reconnaissance d'une scène ne dépendrait pas du traitement des BFS avant celui des HFS, comme dans une séquence *coarse-to-fine*, mais plutôt de la réinjection de l'information en fréquences spatiales en lien avec le précédent stimulus d'une séquence. Ces auteurs présentaient des séquences de 16 images filtrées d'une scène dans lesquelles le contenu

de la scène était progressivement révélé via l'ajout progressif d'information en fréquences spatiales d'une image à l'autre. Les séquences commençaient soit par une scène filtrée en BFS et l'information en HFS était progressivement ajoutée d'une image à l'autre, soit par une scène filtrée en HFS et l'information en BFS était progressivement ajoutée. Après chaque image d'une séquence, les participants devaient dire s'ils avaient identifié ou non la scène dans son ensemble à l'aide d'un boîtier réponse (réponse oui *vs.* non). Ainsi, contrairement à l'étude de Musel et al. (2012), l'information en fréquences spatiales de l'image étaient réinjectée dans l'image $n+1$, mais par contre, la tâche ne nécessitait pas réellement de traiter rapidement l'ajout d'un nouveau contenu en fréquences spatiales au contenu précédent. Les résultats ont montré que le taux d'identification correcte augmentait à mesure que l'information en fréquences spatiales était ajoutée et ce, indépendamment de la fréquence spatiale initiale (BFS ou HFS) de la séquence. Mais comme nous l'avons fait remarquer, ces travaux s'intéressent plus à la précision du processus d'identification de scènes qu'à sa rapidité. Ces différences de tâches peuvent donc expliquer les divergences de résultats. Toutefois, ces résultats interrogent sur le mode d'accumulation et d'intégration des fréquences spatiales au cours d'une analyse *coarse-to-fine*. Est-ce qu'un traitement additif des fréquences spatiales amènerait à une catégorisation plus rapide qu'un traitement successif ?

Pour répondre à cette question, Kauffmann, Chauvin, Guyader et al. (2015) ont testé l'avantage de l'analyse *coarse-to-fine* sur l'analyse *fine-to-coarse* selon différents modes d'accumulation des fréquences spatiales. Les auteurs présentaient des séquences dynamiques *coarse-to-fine* et *fine-to-coarse* composées de six versions d'une scène. Dans une condition successive, 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 successivement d'une image à l'autre comme dans l'expérience de Musel et al. (2012). Dans une autre condition additive, 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 comme dans l'expérience de De Cesarei et Codispoti (2011). Dans les deux conditions, les séquences commençaient soit par les BFS (séquences *coarse-to-fine*), soit par les HFS (séquences *fine-to-coarse*). Les résultats ont montré que les participants catégorisaient les séquences *coarse-to-fine* plus rapidement que les séquences *fine-to-coarse* quel que soit le mode d'accumulation (successif ou additif) des fréquences spatiales dans les séquences. Ce résultat suggère que l'ajout additif d'informations en fréquences spatiales n'est pas plus avantageux que le traitement successif d'informations provenant de différentes bandes de fréquences spatiales (allant des BFS aux HFS) lors de la catégorisation des scènes.

Si ces différentes études comportementales suggèrent que l'être humain utilise préférentiellement les BFS pour catégoriser rapidement des scènes, les études computationnelles qui simulent les processus corticaux impliqués dans la perception visuelle valident également cette hypothèse (Héroult, Oliva & Guérin-Dugué, 1997 ; Guérin-Dugué & Oliva, 2000). Par exemple, Héroult, Oliva et Guérin-Dugué (1997) ont montré qu'un système visuel artificiel qui modélise les cellules du cortex visuel par des fonctions de Gabor (i.e. des filtres qui extraient le contenu d'une image en fonction des orientations et des fréquences spatiales) est capable de classer des photographies de scènes en catégories sémantiques

distinctes comme le ferait l'humain en exploitant uniquement les informations en basses et moyennes fréquences spatiales.

Le système visuel humain bénéficierait donc du traitement rapide des BFS d'une scène pour catégoriser très rapidement scènes. Bien qu'une stratégie d'analyse *coarse-to-fine* ait été démontrée expérimentalement à de multiples reprises, il est important de souligner certaines limites quant à ce modèle de perception de scènes.

2.3. Limites du modèle *coarse-to-fine*

Le fait que les BFS soient véhiculées plus rapidement que les HFS à travers le système visuel ne signifie pas pour autant qu'elles sont utilisées en premier pour réaliser une tâche visuelle particulière. En effet, plusieurs études comportementales ont montré un traitement préférentiel des informations en HFS même pour des temps de présentation d'un stimulus très courts, suggérant une flexibilité de l'analyse *coarse-to-fine* (Bonnar, Gosselin & Schyns 2002 ; Campagne et al., 2016 ; Collin, 2006 ; Collin & McMullen, 2005 ; Harel & Bentin, 2009 ; Morrison & Schyns, 2001 ; Oliva & Schyns, 1997 ; Özgen et al., 2005 ; Rotshtein et al., 2010 ; Schyns, Bonnar & Gosselin, 2002 ; Schyns & Oliva, 1994 ; Schyns & Oliva, 1999 ; Wiesmann et al., 2021). Schyns et Oliva (1994) ont par exemple montré dans leur deuxième expérience qu'une analyse *coarse-to-fine* était privilégiée dans la majorité des cas (67%), mais qu'une proportion non négligeable des réponses étaient en fait basée sur une stratégie d'analyse *fine-to-coarse* (29%). Ces mêmes auteurs ont ensuite démontré que différentes tâches de catégorisation de visages appliquées à un même stimulus visuel pouvaient induire une utilisation différente du contenu en fréquences spatiales de ce stimulus (Schyns & Oliva, 1999). Les auteurs présentaient très rapidement, pendant 50 ms, des visages hybrides superposant un visage filtré passe-bas de genre féminin ou masculin, expressif ou neutre, à un visage filtré passe-haut de genre opposé et avec une expression différente. Les participants devaient réaliser trois tâches de catégorisation sur les mêmes visages hybrides : (1) une tâche de catégorisation de genre (homme ou femme), (2) une tâche de catégorisation de la présence d'expression (visage expressif ou non expressif) et (3) une tâche de catégorisation d'expression (visage neutre, joyeux ou colérique). Les résultats ont montré que les BFS de l'image hybride étaient préférentiellement utilisées pour catégoriser l'expression du visage (i.e. neutre, joyeux, colérique) alors que les HFS de l'image hybride étaient préférentiellement utilisées pour décider si un visage était expressif ou non. Aucune préférence particulière n'était observée pour la catégorisation du genre. Dans la lignée de ces travaux, Rotshtein et al. (2010) ont démontré que le contenu en fréquences spatiales utilisé pour catégoriser un stimulus visuel particulier (par exemple, un stimulus appartenant à la catégorie des fleurs) différait en fonction des catégories auxquelles il était comparé. Ces auteurs utilisaient des images hybrides superposant une image en BFS et en HFS. Les participants devaient soit focaliser leur attention sur l'image en BFS (la cible) et donc, ignorer l'image en HFS (le distracteur), soit focaliser leur attention sur l'image en HFS (la cible) et donc, ignorer l'image en BFS (le distracteur). Les auteurs demandaient alors aux participants de réaliser deux tâches de catégorisation : une tâche de catégorisation

“fleur-visage” où ils devaient décider s’ils percevaient une fleur ou une maison dans l’image cible (en BFS ou HFS) et une tâche de catégorisation “fleur-maison” où ils devaient décider s’ils percevaient une fleur ou un visage dans l’image cible (en BFS ou HFS). Le distracteur quant à lui pouvait appartenir soit à une catégorie congruente avec la cible, soit incongruent, soit autre. Les résultats ont montré que l’interférence d’un distracteur incongruent avec la catégorisation de la cible dépendait de leur contenu en fréquences spatiales et de la tâche. Ainsi, les performances de catégorisation “fleur-visage” étaient gênées par les distracteurs incongruents en BFS, tandis que les performances de catégorisation “fleur-maison” étaient gênées par les distracteurs incongruents en HFS. Par ailleurs, les fréquences spatiales préférentiellement utilisées pour catégoriser les stimuli fleurs différaient en fonction de la tâche. Dans la tâche de catégorisation “fleur-visage”, les stimuli fleurs étaient mieux catégorisés lorsqu’ils apparaissaient dans la condition cible en HFS que la condition cible en BFS alors que dans la tâche de catégorisation “fleur-maison”, ces mêmes stimuli étaient mieux catégorisés lorsqu’ils apparaissaient dans la condition cible en BFS que cible en HFS. Cette étude démontre à nouveau une utilisation flexible du contenu en fréquences spatiales en fonction de la tâche visuelle qui est demandée aux participants.

Très récemment, Wiesmann et al. (2021) ont étudié quelles fréquences spatiales étaient préférentiellement utilisées au cours du temps pendant la catégorisation de scènes d’intérieur et d’extérieur et ce, en fonction du niveau de catégorisation (niveau de base vs. surordonné). Les auteurs présentaient des courtes vidéos (333 ms) composées de 40 images successives d’une même scène, mais au contenu en fréquences spatiales différent. Pour chaque image de la vidéo, les auteurs échantillonnaient plusieurs bandes de fréquences spatiales. Par exemple, la première image d’une vidéo pouvait contenir une bande de très basses fréquences combinée à une bande de fréquences moyennes, la seconde image pouvait ensuite contenir une bande de très haute fréquences combinée à une bande de relativement basses fréquences, et ainsi de suite. L’échantillonnage était réalisé de façon aléatoire pour chacune des images composant une vidéo, et ce pour chaque vidéo. L’intérêt d’une telle méthode d’échantillonnage au sein d’une vidéo est de pouvoir tester un maximum de bandes de fréquences spatiales au sein d’un seul stimulus visuel relativement court (seulement 333 ms) sans avoir d’a priori sur la largeur et/ou la fréquence de coupure des bandes de fréquences spatiales comme c’est le cas dans les études présentées jusqu’ici. A titre d’exemple, dans Schyns et Oliva (1994), une scène en BFS comprend toutes les fréquences spatiales inférieures à 2 cpd alors qu’une scène en HFS comprend toutes les fréquences spatiales supérieures à 6 cpd. Les participants devaient catégoriser les vidéos d’intérieur et d’extérieur soit au niveau de base (e.g., bureau, cuisine et chambre pour les scènes d’intérieur ; montagnes, forêts, champs pour les scènes d’extérieur), soit au niveau surordonné (e.g., intérieur ; extérieur). Les auteurs ont ensuite effectué des analyses de régression linéaire multiples pour déterminer quelle combinaison de fréquences spatiales au cours du temps entraînait les meilleures performances de catégorisation. Autrement dit, les auteurs ont relevé les coefficients de corrélation les plus élevés entre la présentation d’une fréquence spatiale à un moment donné et les taux de réussite de la tâche de catégorisation. Pour la tâche de catégorisation au niveau de base, les résultats ont montré que les taux de réussite pour catégoriser des scènes d’intérieur étaient fortement corrélés avec des HFS au début, au milieu et à la fin de la vidéo, tandis que les taux de réussite pour catégoriser scènes

d'extérieur étaient fortement corrélés avec des HFS au début et des BFS à la fin, indiquant une analyse *fine-to-coarse*. Pour la tâche de catégorisation au niveau surordonné, les résultats ont montré que les taux de réussite pour catégoriser des scènes d'intérieur et d'extérieur étaient fortement corrélés avec des BFS au début, puis des HFS au milieu et des BFS à la fin de la présentation. Ces résultats suggèrent une nouvelle fois que l'utilisation préférentielle des fréquences spatiales dépend de la tâche visuelle qui est demandée aux participants (tâche de catégorisation au niveau de base et au niveau surordonné) et soutiennent ainsi la flexibilité de l'analyse *coarse-to-fine*.

Enfin, l'utilisation préférentielle des fréquences spatiales est également fortement sujette à des différences inter-individuelles. Par exemple, Musel et al. (2012) ont montré que si des participants jeunes (moins de 30 ans) avec une vision normale catégorisaient systématiquement plus rapidement une scène suivant une séquence *coarse-to-fine* que *fine-to-coarse*, la même expérience menée auprès de participants âgés (plus de 60 ans) avec une vision normale montrait une tendance privilégier une séquence d'analyse *fine-to-coarse* pour ces derniers. Au sein même d'une population pathologique, des différences peuvent également être observées. Par exemple, Vanmarcke et Wagemans (2016) ont montré que des personnes avec un trouble du spectre autistique élevé étaient plus lents pour catégoriser les HFS d'une image hybride présentée avec un temps court (33 ms) par rapport à des personnes avec un trouble plus faible, le résultat inverse étant observé avec un temps de présentation plus long (132 ms).

Malgré la prédominance apparente du traitement *coarse-to-fine*, les résultats de ces études suggèrent que l'analyse *coarse-to-fine* correspondrait à une analyse « par défaut » qui serait flexible dans le cas où la tâche permettrait d'allouer plus de poids au traitement des HFS. Ainsi, bien qu'il ait été démontré à de nombreuses reprises que les informations en BFS sont disponibles avant les informations en HFS, il est important de considérer qu'elles ne sont pas toujours utilisées en premier pour prendre en charge la reconnaissance visuelle. La sélection des fréquences spatiales dépendrait des interactions entre leurs disponibilités au sein de l'input visuel et des exigences d'une tâche visuelle donnée.

Enfin, l'analyse *coarse-to-fine* que nous avons décrite ci-dessus suggère un traitement des fréquences spatiales par le cerveau selon une propagation uniquement ascendante, des BFS aux HFS. Cette propagation purement ascendante peut être remise en question par la découverte de nombreuses connexions descendantes dans le cerveau. Dans la partie qui suit, nous présenterons les études neurophysiologiques qui mettent en évidence des connexions corticales descendantes et un modèle neurobiologique du traitement visuel qui en découle.

3. Traitements ascendants et descendants de l'information visuelle

3.1. Modèle hiérarchique du cortex visuel

Le long de la voie ventrale, on observe une complexité grandissante des traitements réalisés du cortex visuel primaire jusqu'au cortex inférotemporal. Comme nous l'avons évoqué plus haut, la taille des champs récepteurs des neurones augmente à mesure qu'on avance dans la hiérarchie corticale. Cette complexité grandissante a largement contribué à une conception hiérarchique des traitements visuels réalisés par le cerveau, qui a été proposée pour la première fois par Hubel et Wiesel dans les années 60 (Hubel & Wiesel, 1962). Ces auteurs ont mis en évidence une augmentation de la complexité des champs récepteurs et de la sélectivité des neurones des six couches de l'aire V1 du chat. Ils ont pour cela montré que certains neurones de V1, appelés cellules simples (Hubel & Wiesel, 1959), répondent à la présentation de traits lumineux ayant une orientation spécifique. Par exemple, une cellule qui répond à une barre verticale ne répondra pas à une barre horizontale ou oblique. L'information provenant de plusieurs cellules simples converge ensuite vers des cellules dites complexes qui traitent des informations plus complexes et plus sélectives. Les cellules complexes possèdent un champ récepteur plus large que les cellules simples. Elles sont également sensibles à l'orientation d'un stimulus mais la position précise du stimulus est moins importante. De ce fait, les cellules complexes répondent au mouvement d'un stimulus. La découverte des cellules simples et complexes au sein du cortex visuel a donc amené à l'idée d'une structure hiérarchique du système visuel.

Les premiers véritables modèles hiérarchiques du traitement cortical reposent sur des travaux en neurophysiologie réalisés chez le primate non-humain (Felleman & Van Essen, 1991 ; Van Essen & Deyoe, 1995 ; Van Essen & Maunsell, 1983). Selon ces modèles, le traitement de l'information se ferait de façon ascendante, du premier niveau vers le dernier niveau de la hiérarchie. Les neurones des aires visuelles des premiers niveaux corticaux telles que V1 et V2 effectueraient des traitements locaux sur une information visuelle en deux dimensions (extraction des contours de différentes orientations) qui seraient ensuite traitée par les neurones des aires corticales supérieures codant une information de plus en plus complexe (formes, assemblage) jusqu'à aboutir au traitement d'une forme globale encore plus complexe au niveau le plus haut de la hiérarchie. Cette conception hiérarchique a été testée par David et al. (2006) chez le primate non-humain en utilisant une méthode d'analyse permettant d'estimer la distribution de la sélectivité des neurones de V4. Ces auteurs ont montré que les neurones de V4 étaient sélectifs à des stimuli plus complexes en termes d'orientations et de fréquences spatiales que ceux de V1. Par ailleurs, les neurones des aires extra-striées de primates non-humains sont sensibles à des caractéristiques visuelles qui ne sont pas ou peu traitées dans V1 (Orban, 2008). Par exemple, nous avons vu qu'au niveau de V1, la taille des champs récepteurs des neurones est petite et les neurones répondent à différentes fréquences spatiales et orientations en fonction de leur position dans le champ visuel (Hubel & Wiesel, 1959). Dans les aires intermédiaires de la hiérarchie corticale, la taille des champs récepteurs des neurones augmente et les neurones sont capables de répondre aux contours (Pasupathy & Connor, 1999).

Plus loin dans la hiérarchie dans le cortex inférotemporal, la taille des champs récepteurs des neurones sont très larges et intègrent des contours en formes complexes (Brincat & Connor, 2004).

3.2. Limites d'une conception purement hiérarchique du traitement cortical

Une conception purement hiérarchique du traitement cortical de l'information visuelle est toutefois contestée par plusieurs travaux qui ont étudié la rapidité du processus de reconnaissance visuelle chez l'homme. En effet, il a été montré que moins de 150 ms sont nécessaires pour aboutir à la reconnaissance visuelle chez des sujets humains (Thorpe et al., 1996 ; Torralba & Oliva, 2003) alors qu'environ 10 ms seraient nécessaires pour transmettre l'information entre deux aires successives (Bullier, 1998). En effet, si on considère que le traitement visuel doit passer par les 25 aires organisées hiérarchiquement présentées dans la Figure 1.8 et qu'environ 10 ms sont nécessaires pour transmettre l'information entre deux aires successives (Bullier, 1998), 250 ms seraient alors nécessaires pour aboutir à la reconnaissance visuelle. Ainsi, la rapidité du processus de reconnaissance visuelle est manifestement incompatible avec une transmission ascendante de l'information à travers une hiérarchie corticale. Toutefois, de nombreuses études en neurophysiologie conduites sur le primate non-humain (pour une revue, voir Bullier, 2001) ont montré que les latences des réponses des cellules d'aires de haut niveau dans la hiérarchie sont souvent comparables, et peuvent même précéder celles d'aires de plus bas niveau. Par exemple, l'aire MST (*Medial Superior Temporal*), l'aire FEF (*Frontal Eye Fields*), ou encore l'aire MT (*Medial Temporal*) ont des latences de réponse similaires à celles de V1 (aire visuelle primaire), bien qu'elles soient situées à un niveau plus haut dans la hiérarchie corticale (Figure 1.8). On peut également noter que les aires qui ont les latences les plus courtes sont situées le long de la voie dorsale alors que les aires de la voie ventrale ont globalement des latences plus longues. Cela s'expliquerait par le fait que la voie dorsale reçoit majoritairement des projections de fibres magnocellulaires dont la vitesse de conduction est plus rapide que les projections de fibres parvocellulaires. Dans la voie ventrale, les projections sont majoritairement parvocellulaires bien que des projections magnocellulaires soient présentes en faibles proportions (Merigan & Maunsell, 1993).

Mais la conception uniquement hiérarchique du traitement cortical a finalement été totalement remise en question par de nombreuses études physiologiques menées sur des primates non-humains qui ont mis en évidence des connexions corticales ascendantes, mais aussi descendantes, c'est-à-dire se projetant d'une aire hiérarchiquement supérieure à une aire hiérarchiquement inférieure. Maunsell et Van Essen (1983a) font partie des premiers auteurs à avoir identifié les connexions descendantes au sein du système visuel du primate non-humain. Ils ont observé, par exemple, des connexions réciproques entre l'aire visuelle temporelle moyenne et les aires visuelles striées et extrastriées (V1, V2, V3 et V4). Les travaux menés par Jean Bullier au Centre de Recherche Cerveau et Cognition à Toulouse dans les années 90 mettent également en évidence de nombreuses connexions entre les aires visuelles corticales (Baizer et al., 1991 ; Bullier et al, 1996 ; Morel & Bullier, 1990 ; Salin & Bullier, 1995) dont la

majorité sont descendantes. Ces connexions descendantes suggèrent ainsi la possibilité de nombreux mécanismes de *feedback* au sein du système visuel.

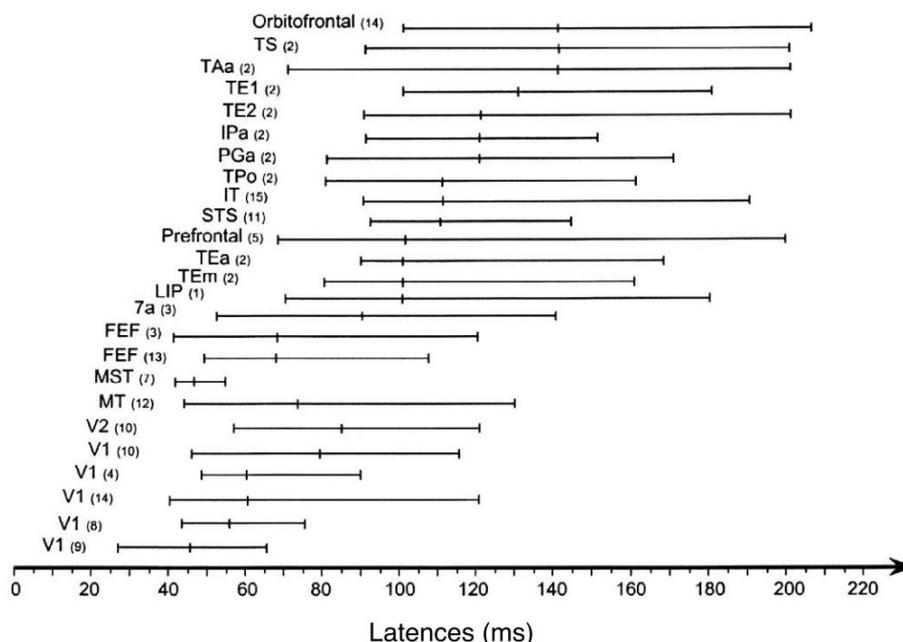


Figure 1.8 : Latences de réponse visuelle de neurones de différentes aires corticales. Pour chaque aire, les barres verticales centrales représentent la latence de réponse médiane (en millisecondes) et les barres de chaque côté représentent les centiles de 10 à 90%. Les chiffres entre parenthèses sont les références citées par Bullier et al. (2001). Les aires sont organisées hiérarchiquement du bas vers le haut (de V1 au cortex orbitofrontal). Les latences de réponses des aires au niveau supérieur dans la hiérarchie ne sont pas nécessairement plus grandes que celles de niveaux inférieurs. Par exemple, l'aire MST (*Medial Superior Temporal*), l'aire FEF (*Frontal Eye Fields*), ou MT (*Medial Temporal*) ont des latences de réponse similaires à celles de V1 (aire visuelle primaire).

3.3. Modèle intégratif du traitement visuel

Afin de considérer les caractéristiques anatomiques et fonctionnelles du système visuel, mais aussi de répondre aux contraintes temporelles de la reconnaissance visuelle, Jean Bullier (2001) a proposé un modèle intégratif du traitement visuel dans lequel il y aurait un mécanisme de rétroinjection de l'information traitée par des aires de haut niveau de la hiérarchie corticale vers les aires de plus bas niveau. Les aires de haut niveau dont les latences de réponse sont courtes pourraient rapidement être activées pour influencer la réponse des aires de plus bas niveau comme V1 et V2. Dans son modèle, la transmission de l'information visuelle ne serait pas seulement ascendante, mais s'effectuerait selon trois vagues d'activation successives (Figure 1.9). La première vague d'activation serait déclenchée par l'arrivée d'un signal magnocellulaire qui arriverait au niveau du cortex visuel via les couches M des corps genouillés latéraux. Ce signal se propagerait très rapidement à travers la voie dorsale jusqu'au cortex pariétal. La deuxième vague d'activation serait déclenchée par les traitements effectués dans

ces régions de haut niveau, qui seraient ensuite rétro-injectés vers V1 et V2 via des connexions descendantes rapides. Enfin, la troisième vague d'activation concernerait l'arrivée d'un signal parvocellulaire au niveau du cortex visuel (dans les aires de bas niveau) via les couches P des corps genouillés latéraux. La rétroinjection des traitements issus de la seconde vague permettrait de guider et de moduler le traitement du signal parvocellulaire entrant et ce signal se propagerait ensuite très rapidement jusqu'au cortex inférotemporal. Ainsi, le traitement ascendant du signal parvocellulaire au sein des aires corticales de bas niveau serait influencé de façon descendante par l'issue du traitement du signal magnocellulaire au sein des aires de haut niveau, ce qui permettrait d'accélérer le traitement visuel. Dans ce modèle, les aires corticales de bas niveau intégreraient le signal magnocellulaire descendant et le signal parvocellulaire ascendant.

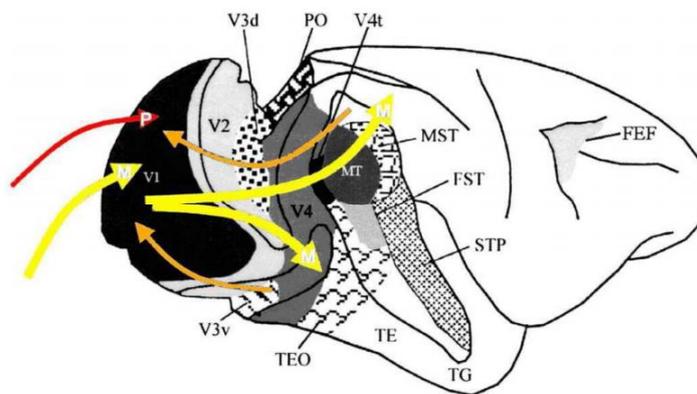


Figure 1.9 : Schématisation du modèle intégratif du traitement visuel de Bullier (2001). Les flèches jaunes représentent la première vague d'activation qui correspond à l'arrivée d'un signal magnocellulaire au niveau du cortex visuel. Les flèches orange représentent la deuxième vague d'activation qui correspond aux connexions descendantes des aires de haut niveau vers les aires visuelles de bas niveau. La flèche rouge représente la troisième vague d'activation qui correspond à l'arrivée plus tardive du signal parvocellulaire au niveau du cortex visuel.

A noter que le modèle de Bullier (2001) ne mentionne pas certaines spécificités du système visuel qu'il semble quand même important de préciser. Par exemple, des travaux neurophysiologiques menés chez le primate non-humain ont mis en évidence des réponses neuronales très rapides de la voie ventrale, plus précisément du cortex inférotemporal, entre 80 à 100 ms après la présentation d'un stimulus (Perrett et al., 1982 ; Vogels, 1999). Ces données, sans pour autant contredire le modèle de Bullier (2001), permettent de rappeler le nombre considérable de projections des fibres magnocellulaires sur le cortex inférotemporal au sein de la voie ventrale et posent également la question d'un traitement très rapide de l'information visuelle au sein de cette voie. Enfin, le modèle de Bullier (2001) ne précise pas la nature de l'information visuelle véhiculée lors des différentes vagues d'activation. Comme nous l'avons mentionné précédemment, les voies magnocellulaire et parvocellulaire véhiculent différents types d'informations visuelles. La voie magnocellulaire véhicule une information en BFS alors que la voie parvocellulaire véhicule plutôt une information en HFS. Ainsi, le premier traitement

ascendant rapide de l'information magnocellulaire pourrait correspondre à la propagation d'une information en BFS, dont les résultats du traitement seraient ensuite rétroinjectés pour venir influencer le traitement ascendant plus lent d'une information en HFS.

3.4. Théorie de la hiérarchie inversée

L'existence de connexions descendantes est également au cœur de la Théorie de la hiérarchie inversée (*reverse hierarchy theory*) proposée par Hochstein et Ahissar (2002). Dans cette théorie, les connexions corticales sont bidirectionnelles, c'est-à-dire que l'information peut être transmise de façon ascendante, des aires de bas niveau aux aires de haut niveau, comme dans une hiérarchie classique, et de façon descendante, des aires de haut niveau aux aires de bas niveau, la hiérarchie étant alors inversée. Ces connexions sont à la base de deux modes d'analyses visuels complémentaires (Figure 1.10) : *vision at a glance* et de *vision with scrutiny*. Le premier mode *vision at a glance* correspond aux traitements ascendants qui seraient essentiellement implicites et inconscients. Une fois arrivés dans les aires de haut niveau, ces traitements rapides deviendraient alors explicites et conscients et permettrait d'aboutir à une première représentation de l'essentiel d'une scène visuelle mais sans précision (e.g., de la forêt, des immeubles ; appelé le *gist*). Cette première reconnaissance du *gist* s'effectuerait grâce à une attention divisée sur l'ensemble de la scène qui nécessite l'activation des larges champs récepteurs des aires de haut niveau. Le deuxième mode *vision with scrutiny* correspond aux connexions descendantes (*top-down*) qui permettrait l'accès à des représentations plus précises. Ces connexions descendantes sont explicites et conscientes et se propageraient le long de la hiérarchie corticale jusqu'aux aires de bas niveau afin d'intégrer l'information détaillée à la reconnaissance du *gist*. L'intégration des détails s'effectueraient grâce à une attention focalisée sur des parties de la scène visuelle (e.g., nombre de fenêtres sur une maison) qui nécessite l'activation des petits champs récepteurs dans des aires de bas niveau. Une fois de plus dans ce modèle, les auteurs n'associent pas leurs concepts aux traitements des fréquences spatiales, mais nous pouvons facilement faire le rapprochement entre le traitement des BFS et des HFS et les concepts de *vision at a glance* et de *vision with scrutiny*, respectivement.

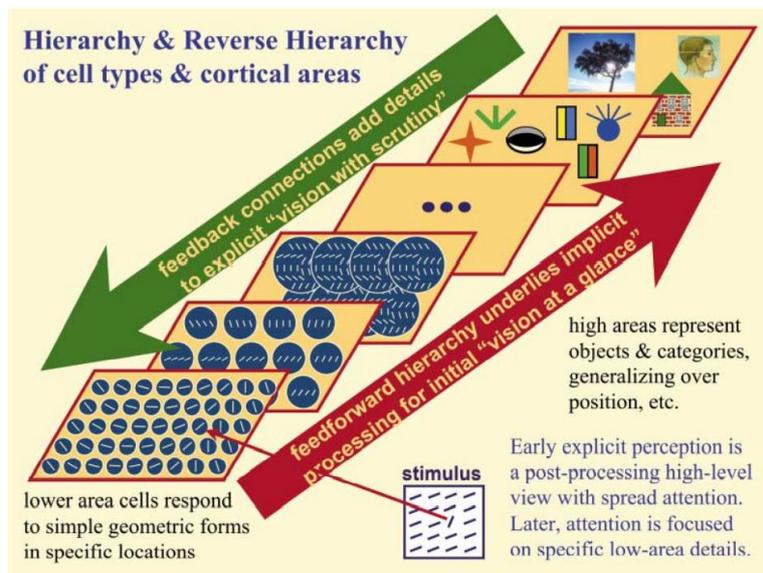


Figure 1.10 : Modèle de hiérarchie inversée de Hochstein et Ahissar (2002).

Ainsi la perception visuelle serait fortement conditionnée par des influences d'aires corticales de haut niveau sur les aires corticales de bas niveau. Ces influences descendantes sont au centre des théories prédictives de la reconnaissance visuelle qui dominent actuellement dans les modèles de la reconnaissance visuelle.

4. Traitements visuels prédictifs et rôle des fréquences spatiales

4.1. Théorie du codage prédictif appliquée à la perception visuelle

La théorie du codage prédictif (ou du traitement prédictif) du cerveau postule que le cerveau infère un modèle interne du monde extérieur qui est utilisé pour générer en permanence des prédictions afin d'anticiper les entrées sensorielles. L'anticipation revient à utiliser le passé pour prédire le présent afin d'interpréter les entrées sensorielles bruitées, et même à remplacer un stimulus masqué ou absent (Rao & Ballard, 1999 ; Friston, 2005). Par exemple, mon modèle interne sait que mes lunettes sont habituellement posées sur l'étagère de la salle de bain. Grâce à l'anticipation de cette information sensorielle, je peux facilement reconnaître mes lunettes, même si elles sont partiellement cachées par d'autres objets ou si ma vision est floue en l'absence de port de lunettes. Ainsi, le codage prédictif permet d'expliquer comment, grâce aux prédictions, les entrées sensorielles bruitées sont facilement et rapidement reconnues. D'un point de vue opérationnel, le cerveau générerait en permanence des prédictions sur la base d'une information sensorielle rudimentaire. Nous verrons dans la suite de ce manuscrit de thèse que c'est le contenu en BFS d'une entrée sensorielle visuelle qui serait utilisé pour générer des

prédictions. Les prédictions seraient générées par les aires corticales de haut niveau et influenceraient ensuite le traitement du signal visuel dans les aires de plus bas niveau. Le signal visuel entrant serait comparé au signal prédit, et seul le signal ne correspondant pas aux prédictions (erreur de prédiction) serait transmis vers les aires corticales de plus haut niveau, permettant ainsi de mettre à jour les prédictions et de réduire ainsi les coûts de traitement des entrées sensorielles.

Le codage prédictif s'inscrit dans un cadre étendu au-delà de la perception visuelle et concerne de nombreux champs des neurosciences cognitives. Dans le domaine de la perception visuelle, Rao et Ballard (1999) ont proposé un modèle du traitement visuel dans lequel il y aurait un mécanisme de retour (*feedback*) de l'information traitée par des aires de haut niveau vers les aires de plus bas niveau (mécanisme qui pourrait s'apparenter à celui de rétroinjection proposé deux années plus tard dans le modèle intégratif du traitement visuel de Bullier, 2001). Le modèle de Rao et Ballard (1999) introduit en plus les termes de prédiction et d'erreur de prédiction. Selon ce modèle, chaque niveau de la hiérarchie corticale tente de prédire les traitements des aires corticales des niveaux inférieurs via des connexions descendantes (*top-down*). L'erreur de prédiction, c'est-à-dire le signal qui ne correspond pas aux prédictions, serait transmise vers les aires de haut niveau via des connexions ascendantes permettant de mettre à jour les modèles internes et donc, les prédictions à venir. Les prédictions et erreurs de prédictions seraient générées simultanément dans toute la hiérarchie, de façon à ce que les informations descendantes influencent les estimations des aires de bas niveaux et les informations ascendantes influencent les estimations des aires de haut niveau.

4.2. Modèle proactif de la reconnaissance visuelle d'objets

Influencé par la mise en évidence des connexions descendantes dans les travaux de Bullier (2001) et s'inspirant directement de la théorie du codage prédictif, Moshe Bar (Bar, 2003 ; Bar, 2007) propose un modèle proactif de la reconnaissance visuelle d'objets qui repose sur un mécanisme cortical facilitant de façon *top-down* la reconnaissance visuelle (Figure 1.11). Plus précisément, ce modèle postule que l'information en BFS contenue dans un stimulus visuel de type objet serait projetée très rapidement et directement, grâce à des "raccourcis" anatomiques magnocellulaires, des aires visuelles primaires au cortex préfrontal (plus particulièrement au cortex orbitofrontal). Bien que rudimentaire, cette représentation visuelle grossière de l'objet permettrait au cortex préfrontal de faire des prédictions sur l'interprétation la plus probable du stimulus visuel. Ces prédictions seraient alors projetées au cortex temporal pour être intégrées au traitement visuel ascendant. Ce mécanisme *top-down* faciliterait alors la reconnaissance visuelle en limitant considérablement le nombre de représentations d'objets à prendre en compte. Par ailleurs, il serait particulièrement utile dans le cas où une réponse rapide est nécessaire.

Plusieurs études menées par Moshe Bar et ses collaborateurs viennent expérimentalement appuyer ce modèle (Bar et al., 2006 ; Kveraga, Ghuman & Bar 2007). Dans une première étude en magnétoencéphalographie (MEG), Bar et al. (2006) demandaient à des participants de reconnaître des dessins d'objets. Ces dessins étaient présentés très rapidement

(pendant 26 ms) et suivis d'un masque afin de gêner leur reconnaissance. Leur reconnaissance était toutefois possible en répétant la présentation des dessins. Les auteurs ont comparé l'activité cérébrale induite par les essais pour lesquels les objets avaient été reconnus à celle induite par les essais pour lesquels ils n'avaient pas été reconnus. Les résultats ont montré que la reconnaissance effective de l'objet (comparé à la présentation initiale de l'objet lorsqu'il n'était pas encore reconnu) s'accompagnait d'une activité du cortex orbitofrontal 50 ms avant l'activité du cortex temporal. Par ailleurs, les auteurs ont observé une synchronie temporelle entre l'activité du cortex orbitofrontal et celle du cortex temporal, suggérant une forte connectivité fonctionnelle entre ces deux régions. Dans une seconde étude en imagerie par résonance magnétique fonctionnelle (IRMf), Bar et al. (2006) ont démontré que l'activité du cortex orbitofrontal était plus importante lorsque les dessins d'objets étaient filtrés en BFS qu'en HFS. Cette étude suggère donc que le cortex orbitofrontal serait très rapidement impliqué dans le processus de reconnaissance d'objet en recevant spécifiquement des informations visuelles en BFS, bien avant l'implication du cortex temporal. En outre, la synchronicité fonctionnelle entre ces deux régions a amené les auteurs à postuler que les traitements réalisés au sein du cortex orbitofrontal seraient susceptibles d'influencer la reconnaissance finale de l'objet au sein du cortex temporal via des connexions descendantes. Ainsi, le réseau cortical responsable de la reconnaissance visuelle inclurait des régions corticales bien au-delà du traditionnel cortex visuel, le cortex orbitofrontal étant alors responsable d'un mécanisme visuel de facilitation *top-down*.

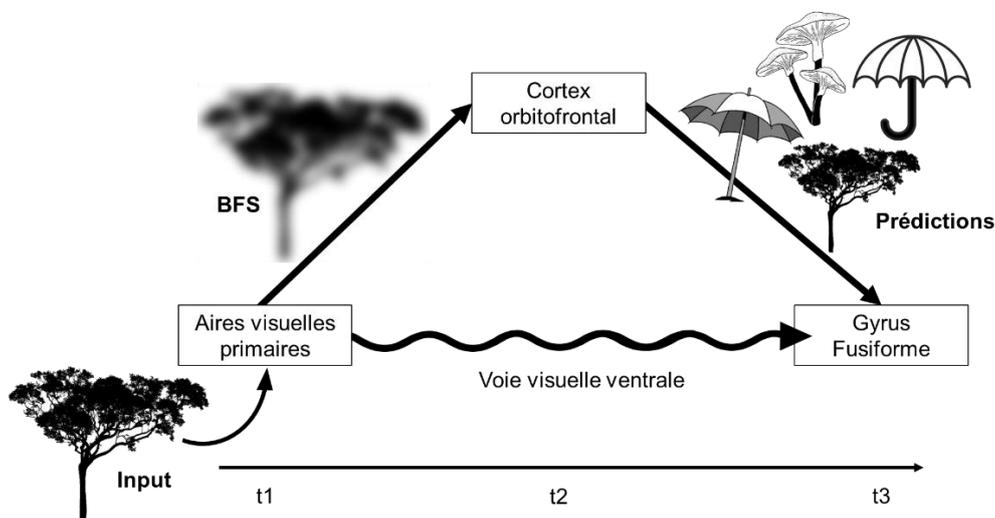


Figure 1.11 : Modèle proactif de la reconnaissance visuelle d'objets de Moshe Bar. Au temps 1 (t_1), l'information contenue dans les BFS de l'objet est envoyée rapidement vers le cortex orbitofrontal, permettant d'activer des prédictions au temps 2 (t_2). Parallèlement, l'information visuelle se propage le long de la voie ventrale jusqu'au gyrus fusiforme dans le cortex inférotemporal. Au temps 3 (t_3), les prédictions sont ensuite projetées sur le gyrus fusiforme et intégrées à l'analyse ascendante afin de faciliter la reconnaissance. Figure inspirée de Bar et al. (2006) et Bar (2007).

Le cortex orbitofrontal ne peut être le site privilégié d'un tel mécanisme que si son activité se développe bien avant celle des régions du cortex temporal associées à la reconnaissance visuelle. Pour Kveraga et al. (2007), cela serait possible grâce à la voie magnocellulaire dont les fibres à conduction rapide activeraient rapidement le cortex orbitofrontal. Afin de tester cette hypothèse, Kveraga et al. (2007) ont mené une étude en IRMf où ils présentaient des dessins d'objets dont les caractéristiques visuelles étaient manipulées afin d'activer sélectivement la voie magnocellulaire (les dessins étaient achromatiques et faiblement contrastés) ou la voie parvocellulaire (les dessins étaient en rouge et vert et isoluminants). Les participants devaient décider si l'objet représenté sur le dessin était plus grand ou moins grand qu'une boîte à chaussures. Les résultats ont effectivement montré que les stimuli magnocellulaires activaient plus fortement le cortex orbitofrontal que les stimuli parvocellulaires alors que ces derniers activent plus fortement le cortex temporal (au niveau du gyrus fusiforme) que les stimuli magnocellulaires. Les auteurs se sont ensuite intéressés à la connectivité fonctionnelle entre le cortex orbitofrontal et le gyrus fusiforme en étudiant la relation temporelle entre les activations de ces deux régions grâce à des modèles causaux dynamiques (*dynamic causal modelling*, DCM). Les résultats ont montré que le traitement des stimuli magnocellulaires avait pour effet d'augmenter la connectivité entre le cortex orbitofrontal et le gyrus fusiforme.

Ainsi, les travaux menés par cette équipe de recherche apportent des arguments en faveur d'un modèle proactif de la reconnaissance visuelle d'objets (Bar, 2003 ; Bar, 2007). Les informations magnocellulaires d'un stimulus visuel (comme, par exemple, les BFS) seraient rapidement envoyées au cortex orbitofrontal. Des connaissances stockées à propos du stimulus y seraient alors activées et utilisées pour générer des prédictions sur l'interprétation la plus probable du stimulus visuel. Le résultat de ces prédictions serait alors renvoyé aux aires visuelles au cortex temporal lors du processus final de reconnaissance, guidant alors le traitement d'informations parvocellulaires arrivant plus lentement (comme, par exemple, les HFS).

4.3. Bases neurales de l'analyse *coarse-to-fine* de scènes visuelles

Les travaux de Moshe Bar sont cohérents avec une stratégie *coarse-to-fine* de l'analyse de l'information visuelle où le traitement des BFS précéderait celui des informations HFS, mais ils suggèrent également que le résultat du traitement initial des BFS est susceptible d'influencer le traitement des HFS via des influences descendantes du cortex orbitofrontal sur les régions du cortex temporal inférieur responsables de la reconnaissance visuelle. Carole Peyrin et ses collaborateurs se sont alors posés la question d'influences descendantes des BFS sur les HFS bien avant le cortex temporal. En effet, les travaux de Bullier (2001) sur le primate non humain suggèrent une rétroinjection de l'information traitée par des aires de haut niveau de la hiérarchie corticale vers les aires de plus bas niveau comme V1 et V2. Cette équipe de recherche a donc mené une première étude en IRMf (Peyrin et al., 2010) afin d'identifier le réseau cortical impliqué dans l'analyse *coarse-to-fine* et son adéquation avec le mécanisme de rétroinjection décrit par Bullier (2001). Dans cette étude, les auteurs présentaient successivement deux scènes

qui pouvaient appartenir ou non à la même catégorie. Dans une condition expérimentale, la première scène était filtrée en BFS et la seconde en HFS (afin d'imiter une séquence d'analyse *coarse-to-fine*). Dans une seconde condition expérimentale, la première scène était cette fois-ci filtrée en HFS et la seconde en BFS (imitant une séquence d'analyse *fine-to-coarse*). Les participants devaient décider si les deux scènes appartenaient à la même catégorie ou non. Les principaux résultats en IRMf ont montré que les séquences *coarse-to-fine* activaient plus fortement le cortex frontal inférieur, le cortex temporal, ainsi que le cortex visuel primaire que les séquences *fine-to-coarse*. Les auteurs ont ensuite étudié le décours temporel de ces activations en mesurant l'activité électroencéphalographique (EEG) des mêmes participants dans des conditions expérimentales similaires. Les résultats en EEG ont montré que lors de l'appariement des deux scènes d'une séquence *coarse-to-fine* (en comparaison à une séquence *fine-to-coarse*), l'activité du cortex frontal inférieur était associée au traitement de la première scène en BFS d'une séquence *coarse-to-fine* (en comparaison avec le traitement de première scène en HFS d'une séquence *fine-to-coarse*), tandis que l'activité du cortex visuel primaire était associée au traitement de la seconde scène en HFS d'une séquence *coarse-to-fine* (en comparaison avec le traitement de la seconde scène en BFS d'une séquence *fine-to-coarse*). En se référant au mécanisme de rétroinjection proposé par Bullier (2001), les auteurs ont postulé que l'augmentation de l'activité du cortex visuel primaire lorsque le traitement des HFS était spécifiquement précédé de celui des BFS correspondait au traitement rapide des BFS par des aires corticales de haut niveau (en particulier, le cortex frontal inférieur) de manière à ce que celles-ci guident, par de rapides feedbacks, le cortex visuel primaire lors du traitement ultérieur des HFS. Malheureusement, la méthode utilisée dans cette étude ne permettait pas de tester explicitement si le cortex frontal inférieur influençait l'activité du cortex visuel primaire.

Kauffmann, Chauvin, Pichat et Peyrin (2015) ont donc mené une étude en IRMf afin d'étudier directement la connectivité fonctionnelle entre les aires cérébrales impliquées dans l'analyse *coarse-to-fine*. Les auteurs ont ici utilisé les séquences dynamiques *coarse-to-fine* et *fine-to-coarse* qui leur avait précédemment permis de démontrer la prédominance d'une stratégie d'analyse *coarse-to-fine* lors de la reconnaissance rapide de scènes (Kauffmann, Chauvin, Guyader et Peyrin 2015). En accord avec les précédents résultats de Peyrin et al. (2010), la catégorisation des séquences *coarse-to-fine* (en comparaison à la catégorisation des séquences *fine-to-coarse*) s'accompagnait d'une activation du cortex frontal inférieur (au niveau du cortex orbitofrontal), du cortex occipito-temporal (au niveau du gyrus fusiforme et du gyrus parahippocampique) et du cortex occipital (au niveau du cortex visuel primaire). Les auteurs ont ensuite réalisé une analyse de connectivité fonctionnelle (*dynamic causal modeling*) afin d'étudier directement la relation temporelle entre les activations de ces régions et donc, leurs interactions. Cette analyse a révélé que la catégorisation des séquences *coarse-to-fine* augmentait la connectivité du cortex orbitofrontal de l'hémisphère droit sur le cortex visuel primaire et le gyrus fusiforme de l'hémisphère droit. Cette étude suggère ainsi qu'un traitement initial des BFS au sein du cortex frontal inférieur pourrait influencer de façon descendante les traitements, en particulier celui des HFS, réalisés au sein du cortex occipito-temporal comme cela avait été précédemment observé par Kveraga et al. (2007), mais aussi au sein du cortex visuel primaire comme le suggéraient les travaux de Peyrin et al. (2010). Malheureusement, une fois encore, si cette étude a permis de démontrer une influence directe de l'activité cortex

orbitofrontal sur l'activité du cortex visuel primaire et du cortex occipito-temporal, les auteurs n'étaient pas en mesure de démontrer que l'analyse des BFS influence réellement celle des HFS. Par contre, un résultat intéressant qui découle des travaux menés par l'équipe de Carole Peyrin concerne les différences de traitement interhémisphériques lors de l'analyse *coarse-to-fine* et plus globalement, lors de l'analyse des fréquences spatiales.

4.4. Spécialisation hémisphérique du traitement des fréquences spatiales et du traitement visuel prédictif

Les différences de traitement entre les hémisphères ont suscité un grand nombre de recherches chez l'homme. Le résultat le plus stable reste la dominance de l'hémisphère gauche pour le traitement du langage chez les droitiers (Broca, 1865 ; Geschwind & Galaburda, 1984 ; Geschwind & Levitsky, 1968 ; Witelson, 1983 ; Witelson & Kigar, 1988). Toutefois, la dominance hémisphérique ne concerne pas seulement le langage comme en témoignent les nombreuses observations du comportement de patients cérébrolésés ou de patients au cerveau dédoublé, mais aussi de personnes avec une vision normale. Les recherches réalisées auprès de personnes dites saines dans le domaine de la vision ont été possible grâce au développement du paradigme expérimental de présentation de stimuli en champs visuels divisés. Cette technique repose sur le principe de controlatéralité de la projection des informations du champ visuel sur le cerveau. L'information présentée dans un hémichamp visuel se projette sur le cortex visuel de l'hémisphère controlatéral. En présentant à un observateur un stimulus soit à gauche, soit à droite d'une fixation centrale, il est alors possible de l'adresser à un seul hémisphère, à condition que la présentation soit brève pour empêcher les mouvements oculaires qui redirigeraient l'information vers l'hémisphère non désiré. On est ainsi en mesure de comparer les différences de performances de l'observateur selon que le stimulus ait été présenté dans le champ visuel gauche/hémisphère droit (CVG/HD) ou dans le champ visuel droit/hémisphère gauche (CVD/HG).

Grâce à cette méthode comportementale, Carole Peyrin et ses collaborateurs ont mis en évidence une spécialisation hémisphérique des fréquences spatiales (Peyrin et al., 2003 ; Peyrin, Chokron et al., 2006 ; Peyrin, Mermillod et al., 2006). Dans leur étude princeps (Peyrin et al., 2003), les auteurs présentaient rapidement (100 ms) des petites scènes (4° d'angle visuel) filtrées en BFS et HFS soit dans le champ visuel gauche, soit dans le champ visuel droit. Les participants devaient décider si la scène présentée était une autoroute ou une ville indépendamment de son contenu en fréquences spatiales. Les résultats ont montré que les participants étaient plus rapides pour catégoriser une scène en BFS dans le CVG/HD que dans le CVD/HG. Inversement, ils étaient plus rapides pour catégoriser une scène en HFS dans le CVD/HG que dans le CVG/HD. Ces résultats suggèrent une supériorité de l'hémisphère droit pour catégoriser des BFS et un avantage de l'hémisphère gauche pour catégoriser des HFS. Ces mêmes auteurs ont ensuite adapté leur expérience comportementale de catégorisation de scènes filtrées en BFS et HFS dans une étude en IRMf afin d'identifier les aires cérébrales impliquées dans cette spécialisation hémisphérique (Peyrin et al., 2004). Dans cette étude, les auteurs utilisaient les mêmes petites scènes filtrées en BFS et HFS qu'ils présentaient en vision centrale,

la technique de l'IRMf permettant d'observer directement dans le cerveau l'hémisphère traitant l'information visuelle. Les auteurs comparaient ensuite directement les activations d'un hémisphère à celles de l'autre hémisphère pour chaque fréquence spatiale. Les résultats ont montré que le gyrus occipital moyen était plus activé dans l'hémisphère droit que dans l'hémisphère gauche lors de la catégorisation des scènes en BFS, alors que cette même région était plus activée dans l'hémisphère gauche que dans l'hémisphère droit lors de la catégorisation des scènes en HFS. Des résultats similaires ont été observés dans l'étude en IRMf d'Idaka et al. (2004) où les auteurs utilisaient des images de visages et de maisons filtrés en BFS et HSF. Ces résultats ont ensuite été répliqués par Musel et al. (2013) qui ont utilisé des scènes filtrées en BFS et HFS de plus grande taille. Ces trois études en IRMf mettent donc en évidence une spécialisation hémisphérique du traitement des fréquences spatiales dès le cortex occipital.

Peyrin et al. (2005) se sont ensuite demandés si ce traitement asymétrique des fréquences spatiales au niveau occipital déterminerait une dominance hémisphérique lors de l'analyse *coarse-to-fine* de scènes visuelles. En effet, lors d'une analyse *coarse-to-fine*, le traitement initial de l'information en BFS pourrait, par exemple, pré-activer l'hémisphère droit pour un traitement ultérieur de l'information en HFS dans ce même hémisphère. Les auteurs ont donc testé cette hypothèse grâce à une étude en IRMf où ils comparaient directement les activations d'un hémisphère à celles de l'autre hémisphère lors de la catégorisation de séquences dynamiques *coarse-to-fine* ou *fine-to-coarse*. Pour rappel, ces stimuli, également utilisés dans Peyrin et al. (2010), consistent à présenter successivement deux scènes qui peuvent appartenir ou non à la même catégorie. Dans une condition expérimentale, la première scène est filtrée en BFS et la seconde en HFS (afin d'imiter une séquence d'analyse *coarse-to-fine*). Dans une seconde condition expérimentale, la première scène est cette fois-ci filtrée en HFS et la seconde en BFS (imitant une séquence d'analyse *fine-to-coarse*). Les participants devaient décider si les deux scènes appartenaient ou non à la même catégorie. Les résultats ont montré que le cortex occipito-temporal droit était plus activé que la gauche lors de l'analyse *coarse-to-fine*, alors que cette même région était plus activée dans l'hémisphère gauche que dans le droit lors d'une analyse *fine-to-coarse*. Ces résultats suggèrent que ces deux types de séquence d'analyse des fréquences spatiales (*coarse-to-fine* et *fine-to-coarse*) peuvent coexister dans le système visuel, chacune prédominant dans un hémisphère. L'hémisphère spécialisé dans le traitement d'une séquence d'analyse particulière serait alors déterminé par la bande de fréquences spatiales apparaissant en premier. On peut alors imaginer que la voie corticale ventrale de l'hémisphère droit ou de l'hémisphère gauche serait préférentiellement impliquée en fonction de la tâche visuelle à réaliser. En effet, nous avons vu précédemment que si la reconnaissance visuelle de scènes suivrait par défaut une logique d'analyse *coarse-to-fine*, de nombreuses études psychophysiques montrent qu'une logique de traitement *fine-to-coarse* est parfois préférée en fonction des demandes de la tâche de reconnaissance (REF).

Si l'hémisphère droit prédomine dans l'analyse *coarse-to-fine* de scènes visuelles, il est alors possible qu'il soit celui qui prédomine dans les traitements visuels prédictifs. C'est effectivement ce qui a été observé dans l'étude en IRMf de Kauffmann, Chauvin, Pichat et Peyrin (2015). Ces auteurs ont observé que la catégorisation des séquences *coarse-to-fine* augmentait la connectivité du cortex orbitofrontal de l'hémisphère droit sur le cortex visuel primaire et le gyrus fusiforme de l'hémisphère droit. Il est important de préciser au sujet de cette

étude que les auteurs ont en fait observé des activations bilatérales de ces régions mais que ce résultat n'avait pas été observé pour la même analyse de la connectivité fonctionnelle réalisée sur ces mêmes régions de l'hémisphère gauche. Cette étude suggère donc que les influences descendantes du cortex frontal inférieur sur des aires visuelles inférieures seraient spécifiques aux régions de l'hémisphère droit et donc, que la génération des prédictions prédominerait dans ce même hémisphère.

Malheureusement, l'ensemble des études que nous avons présentées dans les précédentes parties ne permettaient pas de tester une réelle influence des BFS sur l'analyse de HFS et posent globalement la question de la façon dont les BFS et HFS sont intégrées en un seul percept. Dans la partie qui suit, nous allons donc décrire les différents travaux comportementaux et en neuroimagerie qui ont notamment montré que les BFS pouvaient influencer directement le traitement des HFS afin de faciliter la reconnaissance visuelle.

4.5. Intégration des fréquences spatiales lors de la perception de scènes visuelles

La problématique de l'intégration des fréquences spatiales lors de la catégorisation de scènes a été adressée par Mu et Li (2013) dans une étude combinant mesures comportementales et en EEG, puis par Kauffman, Bourgin, Guyader et Peyrin (2015) dans une étude combinant des mesures comportementales et en IRMf, grâce à un paradigme expérimental d'interférence sémantique du traitement des BFS sur le traitement des HFS. Pour cela, les auteurs utilisaient des images hybrides comme stimuli expérimentaux, c'est-à-dire des images qui superposaient deux scènes, l'une filtrée en BFS et l'autre en HFS (Figure 1.12). Les deux scènes qui composaient une image hybride étaient soit sémantiquement congruentes (e.g., une scène de ville en BFS appartenant à la catégorie des scènes "artificielles" superposée à une autre scène de ville en HFS appartenant donc à cette même catégorie), soit sémantiquement incongruentes (e.g., une scène de ville en BFS appartenant à la catégorie des scènes "artificielles" superposée à une scène de plage en HFS appartenant à la catégorie des scènes "naturelles"). Les auteurs manipulaient également la similarité physique entre les deux scènes qui composent l'image hybride. Ainsi, dans chaque condition de congruence sémantique, les deux scènes qui composaient l'image hybride étaient soit physiquement similaires (i.e. que les deux scènes se superposaient spatialement - pixel par pixel - et leurs spectres d'amplitude étaient similaires), soit physiquement dissimilaires. Les participants devaient ignorer la scène en BFS et porter leur attention sur la scène en HFS. Ils devaient alors catégoriser la scène en HFS comme étant soit une scène "artificielle", soit une scène "naturelle". Des hypothèses peuvent alors être formulées sur la base du modèle proactif de la reconnaissance visuelle (Bar, 2003 ; Bar, 2007). Si l'information en BFS est rapidement utilisée pour générer des prédictions sur la catégorie de la scène et guider le traitement des HFS, alors une information sémantique en BFS différente de celle en HFS devrait conduire à des prédictions erronées et donc à interférer avec la catégorisation de la scène en HFS par rapport à des informations sémantiques en BFS et HFS congruentes. On parle ainsi d'un effet d'interférence sémantique de l'information en BFS sur

le processus de catégorisation de la scène en HFS. Cet effet d'interférence peut être considéré comme étant la résultante d'une intégration des deux informations en un tout.

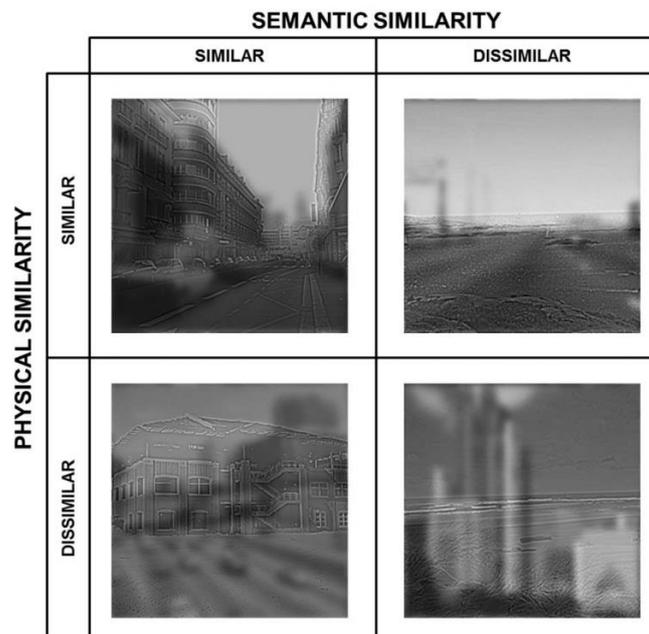


Figure 1.12 : Exemple de stimuli hybrides utilisés dans Kauffmann, Bourgin et al. (2015) et Kauffmann et al. (2017). Les deux scènes appartenant à une image hybride pouvaient être sémantiquement congruentes ou sémantiquement incongruentes, et physiquement similaires ou physiquement dissimilaires. Adapté de Kauffmann, Bourgin et al. (2015).

Les résultats comportementaux de ces deux études en neuroimagerie ont effectivement montré que la catégorisation des scènes en HFS était fortement gênée par un contenu en BFS sémantiquement incongruent qui devaient pourtant être ignoré. Les participants faisaient plus d'erreurs de catégorisation de la scène en HFS lorsque la scène en BFS à ignorer était sémantiquement incongruente que sémantiquement congruente. Par ailleurs, dans l'étude de Mu et Li (2013), cet effet d'interférence était plus marqué lorsque les deux scènes étaient physiquement similaires, ce résultat n'ayant pas été répliqué par Kauffmann, Bourgin et al. (2015). Au niveau cérébral, les résultats EEG de Mu et Li (2013) ont montré que l'interférence sémantique des BFS sur les HFS était associé à une composante frontale négative (N1) seulement 122 ms après l'apparition de l'image hybride et ce, seulement lorsque les deux scènes étaient physiquement similaires. Ce résultat suggère que ce court laps de temps suffirait pour que les informations sémantiques et physiques contenues dans les BFS influencent les HFS au sein du cortex frontal. De façon cohérente, les résultats IRMf de Kauffmann, Bourgin et al. (2015) ont montré une activation plus importante du cortex frontal inférieur (au niveau du cortex orbitofrontal) et du cortex inférotemporal (au niveau du gyrus fusiforme et du gyrus parahippocampique) lorsque la scène en BFS était sémantiquement incongruente que sémantiquement congruente, et ce indépendamment de la similarité physique entre les deux scènes. Une analyse de connectivité fonctionnelle (*psycho-physiological interactions*, PPI) a

révélé que l'interférence sémantique des BFS augmentait la connectivité fonctionnelle entre le gyrus fusiforme et le gyrus frontal inférieur de l'hémisphère droit. Pour ces auteurs, ces résultats suggèrent un traitement automatique de l'information en BFS, même lorsqu'elle n'est pas explicitement utile à la catégorisation. Ainsi, lorsque les deux scènes sont sémantiquement incongruentes, ce traitement automatique conduirait le cortex frontal inférieur à générer des prédictions erronées concernant la catégorie de la scène en HFS. Ces prédictions interféreraient sur le traitement des HFS au sein du cortex inférotemporal, résultant en une augmentation des erreurs de catégorisation et en une augmentation de la connectivité entre ces deux régions de l'hémisphère droit pour résoudre le conflit. Ces deux régions interagiraient plus fortement afin de traiter les ambiguïtés entre les informations en BFS et HFS. Qui plus est, l'hémisphère droit jouerait alors un rôle prépondérant dans le processus d'intégration des fréquences spatiales grâce à sa spécialisation dans le traitement rapide des BFS et donc, de son implication plus rapide que l'hémisphère gauche.

Ces deux études suggèrent que des mécanismes d'influence des BFS sur les HFS pourraient être initiés par des régions frontales, et donc de nouveau le possible rôle de ces régions dans la gestion des prédictions issues du traitement des BFS. Par la suite, Kauffmann et al. (2017) se sont posés la question d'une potentielle interférence sémantique des HFS sur la catégorisation d'une scène en BFS. En effet, il existe des preuves comportementales d'un traitement préférentiel des informations en HFS même pour des temps de présentation d'un stimulus très courts (Bonnar, Gosselin & Schyns 2002 ; Campagne et al., 2016 ; Harel & Bentin, 2009 ; Morrison & Schyns, 2001 ; Oliva & Schyns, 1997 ; Özgen et al., 2005 ; Rotshtein et al., 2010 ; Schyns, Bonnar & Gosselin, 2002 ; Schyns & Oliva, 1994 ; Schyns & Oliva, 1999 ; Wiesmann et al., 2021). Cependant, le rôle respectif de l'information en HFS et son impact sur le traitement des BFS est peu connu. Kauffmann et al. (2017) ont alors réalisé une étude comportementale dans laquelle la moitié des participants devaient catégoriser la scène en HFS de l'image hybride et ignorer la scène en BFS. Cette condition expérimentale était similaire à celle utilisée par Mu et Li (2013) et Kauffmann, Bourgin et al. (2015). L'autre moitié des participants devait cette fois-ci catégoriser la scène en BFS de l'image hybride et ignorer la scène en HFS. Les deux scènes qui composaient une image hybride étaient soit sémantiquement congruentes, soit sémantiquement incongruentes. Les auteurs manipulaient également la similarité physique entre les deux scènes qui composent l'image hybride. Ainsi, dans chaque condition de congruence sémantique, les deux scènes qui composaient l'image hybride étaient soit physiquement similaires (i.e. que les deux scènes se superposaient spatialement - pixel par pixel - et leurs spectres d'amplitude étaient similaires), soit physiquement dissimilaires. Les résultats ont tout d'abord montré que les participants qui catégorisaient la scène en BFS de l'image hybride étaient plus rapides que ceux qui catégorisaient la scène en HFS. Ce résultat appuie à nouveau l'hypothèse d'une analyse *coarse-to-fine*, à savoir que l'analyse de l'information grossière *coarse* portée par les BFS permettrait une reconnaissance rapide de la scène. Par ailleurs, les auteurs ont observé un effet d'interférence de la scène à ignorer pour les deux groupes expérimentaux. L'interférence sémantique des BFS sur la catégorisation des scènes en HFS, qui réplique les résultats de Mu et Li (2013) et Kauffmann, Bourgin et al. (2015), suggère que les BFS seraient analysées très rapidement et permettraient de générer des prédictions sur le contenu de la scène, prédictions susceptibles de moduler la catégorisation de

la scène en HFS. Plus surprenant, les auteurs ont observé que l'effet d'interférence sémantique des HFS sur la catégorisation de la scène en BFS était supérieure à celui des BFS sur la catégorisation de la scène en HFS. Selon les auteurs, lors de la catégorisation d'images complexes et incohérentes telles que les images hybrides, il serait difficile, voire même contre-productif, d'ignorer l'information en HFS au profit des seules représentations grossières issues de l'analyse des BFS. Ainsi, lorsqu'un participant doit catégoriser une scène en BFS, sa réponse pourrait alterner entre les besoins de la tâche, c'est-à-dire porter son attention sur les BFS qui permettent de catégoriser très rapidement la scène en guidant l'analyse tardive des HFS via des mécanismes prédictifs descendants, et la stratégie involontaire déclenchée par la nature complexe du stimulus, c'est-à-dire privilégier l'analyse et l'intégration ascendante des HFS qui interfèrent alors sur le traitement des BFS.

En résumé, cette étude suggère donc que le système visuel s'appuie sur une première représentation grossière de la scène visuelle issue du signal en BFS pour générer rapidement des prédictions sur le contenu de la scène, et les HFS jouent un rôle prépondérant en cas d'incertitude sur le stimulus visuel. Les résultats de cette étude ont également montré que la similarité physique des deux scènes modulait l'effet d'interférence sémantique sans toutefois reproduire les résultats de Mu et Li (2013) ou Kauffmann, Bourgin et al. (2015). Lorsque les participants devaient catégoriser une scène en BFS, l'effet d'interférence sémantique des HFS était plus fort lorsque les deux scènes dans l'image hybride étaient physiquement similaires que dissimilaires. Par contre, les BFS n'interféraient pas plus fortement sur les HFS lorsqu'elles étaient physiquement similaires que lorsqu'elles étaient physiquement dissimilaires. Bien que différents facteurs puissent rendre compte des divergences de résultats entre les différentes études qui ont étudié l'influence de la similarité physique entre deux scènes sur leur interférence sémantique respective (base de stimuli, méthode d'investigation, puissance statistique, etc.), toutes ces études suggèrent que le traitement des propriétés physiques des BFS et des HFS dans les scènes pourrait jouer un rôle important dans leur intégration lors de la catégorisation de scènes.

Plus récemment, en étudiant la contribution relative du traitement des BFS et des HFS lors du décours temporel de la reconnaissance des visages, Petras et al. (2019) ont indirectement démontré que le traitement rapide des BFS influence celui plus lent des HFS. Dans cette étude, les auteurs ont entraîné un classifieur à discriminer des configurations d'enregistrements EEG évoqués par des stimuli en BFS ou en HFS dénués de sens (i.e. des visages humains et de singes dont le spectre de phase avait été modifié par permutation aléatoire). Lorsque le classifieur était capable de reconnaître si la configuration EEG était liée au traitement des BFS ou des HFS, il était alors utilisé pour attribuer un label BFS ou HFS à des enregistrements EEG (évoqués par d'autres stimuli, des stimuli tests, qui contenaient cette fois-ci toutes les fréquences spatiales). Ces stimuli tests pouvaient être des visages intacts ou des visages dont la phase avait été brouillée (i.e. des visages non identifiables mais équivalents aux visages intacts en termes de spectre d'amplitude). Le classifieur était testé toutes les 10 ms à partir de la présentation d'un stimulus test. Il était alors possible de savoir au cours du temps si le signal EEG ressemblait plus à un traitement des BFS ou des HFS. Les résultats ont montré qu'au début du traitement d'un visage intact (entre 130 et 200 ms après sa présentation), le classifieur attribuait plus souvent le label BFS que le label HFS. Ce résultat n'était pas observé pour un visage dont la

phase avait été brouillée. En revanche, à partir de 200 ms, le classifieur n'attribuait plus le label BFS aux visages intacts alors qu'il attribuait cette fois-ci plus souvent le label HFS aux visages dont la phase avait été brouillée. Ces résultats suggèrent un traitement précoce des BFS lorsque le stimulus est porteur de sens (visages intacts) afin d'obtenir une représentation fiable du visage qui permet ensuite de réduire le traitement non nécessaire des HFS. Cette étude, bien que réalisée sur des stimuli visuels de type visage, suggère donc une influence du traitement précoce des BFS sur le traitement des HFS.

Ainsi, les études présentées dans cette partie soutiennent l'hypothèse d'une analyse *coarse-to-fine* de l'information visuelle, et elles démontrent en particulier que les BFS pouvaient influencer directement le traitement des HFS en vue de faciliter la reconnaissance visuelle. Une synthèse de ces deux principes a été rapportée une première fois dans une revue écrite par Kauffmann, Ramanoël et Peyrin (2014), puis formalisée sous le nom de modèle Fréquentiel Prédicatif (Modèle FP ; Figure 1.13) par Alexia Roux-Sibilon lors de sa thèse de doctorat sous la direction de Carole Peyrin (Roux-Sibilon, 2020).

4.6. Modèle fréquentiel prédictif de la perception de scènes

Dans le modèle Fréquentiel Prédicatif (Modèle FP) résumé dans la Figure 1.13, le système visuel décompose d'abord l'image rétinienne en différentes fréquences spatiales comme un analyseur de Fourier. L'analyse visuelle suivrait alors une analyse par défaut *coarse-to-fine*. Les fibres magnocellulaires acheminent le signal BFS de l'image rétinienne au système de reconnaissance visuelle en premier (temps t1) permettant une première représentation globale et rudimentaire (ou *gist*) du stimulus visuel. Les fibres parvocellulaires acheminent le signal HFS de l'image rétinienne au système de reconnaissance visuelle ensuite (temps t2) permettant une analyse plus détaillée et minutieuse du stimulus visuel. Le modèle comprend un système prédictif localisé dans des aires corticales de haut niveau. Celles-ci répondent très rapidement grâce à des projections de fibres magnocellulaires. Les fibres magnocellulaires y acheminent très rapidement le signal BFS de l'image rétinienne (temps t1). Le système prédictif dispose ainsi du *gist* du stimulus visuel qui lui permet de générer des prédictions sur l'interprétation la plus probable de l'image rétinienne. Grâce aux nombreuses connexions descendantes des aires corticales de haut niveau sur des aires de plus bas niveau, le système prédictif envoie ses prédictions au système de reconnaissance afin qu'elles soient intégrées au traitement ascendant de l'information visuelle provenant de l'image rétinienne. Ces prédictions guident en particulier le traitement plus lent du signal HFS. Le signal prédit sur la base des BFS est comparé au signal HFS entrant. Le signal en HFS ne correspondant pas aux prédictions induit une erreur de prédiction qui est alors transmise au système prédictif pour mettre à jour les prédictions. En ce sens, les HFS joueraient un rôle prépondérant en cas d'incertitude sur le signal visuel.

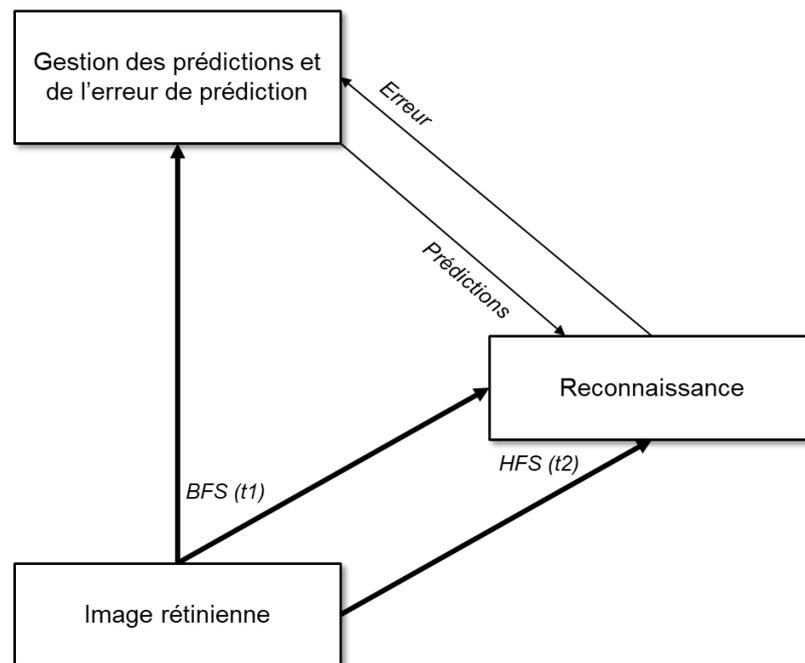


Figure 1.13 : Modèle Fréquentiel Prédicatif. Figure adaptée de la thèse d'Alexia Roux-Sibilon (2020).

Les études ont donc montré que le système visuel s'appuie sur une première représentation grossière de la scène visuelle issue du signal en BFS pour générer rapidement des prédictions sur le contenu de la scène. Cependant, la majorité de ces travaux manipulant les fréquences spatiales ont été menés à l'aide de stimuli visuels présentés en vision centrale et de fait, ne considèrent pas une caractéristique particulièrement importante du traitement rétino-cortical qui est la distinction entre la vision centrale et la vision périphérique. Étant donné que les signaux en BFS proviennent principalement du champ visuel périphérique, il semble important de considérer cette propriété dans les modèles de la reconnaissance visuelle lorsque ceux-ci reposent sur l'analyse rapide d'une information de basse résolution spatiale. Cette lacune nous amène à nous interroger sur le rôle de la vision périphérique pour catégoriser rapidement une scène et elle figure comme la problématique de cette thèse.

5. Problématique et objectifs du travail de thèse

Nous avons vu à plusieurs reprises dans ce chapitre théorique que le système visuel bénéficierait du traitement rapide des BFS. Il disposerait donc d'une représentation globale et rudimentaire (*gist*) de la scène particulièrement utile pour catégoriser une scène dans le cas où une réponse rapide - bien que peu précise - serait nécessaire. Parallèlement, cette représentation servirait également au système visuel à générer très rapidement des prédictions sur l'interprétation la plus probable de la scène qui seraient alors utilisées pour influencer et guider les traitements visuels plus minutieux mais plus lents. Cela permettrait, par exemple, de limiter considérablement le nombre de représentations visuelles à prendre en compte lors de l'analyse plus minutieuse de la scène visuelle. Nous avons également vu dans ce chapitre théorique que

la résolution spatiale des informations visuelles n'est pas uniforme dans le champ visuel mais est déterminée par les propriétés des cellules rétiniennes. La densité de cellules utilisées pour traiter les HFS est très élevée dans la fovéa, tandis que la densité de cellules utilisées pour traiter les BFS augmente avec l'excentricité rétinienne. La précedence temporelle des BFS et leur prédominance en vision périphérique nous amène alors à nous interroger sur le rôle de la vision périphérique pour catégoriser rapidement une scène, mais aussi sur son rôle dans le processus de génération de prédictions. Cette question est d'autant plus pertinente que plusieurs recherches ont mis en évidence l'importance de la vision périphérique, voir même sa supériorité sur la vision centrale, lors de la catégorisation rapide de scènes (Boucart et al., 2013 ; Geuzebroek & van den Berg, 2018 ; Larson & Loschky, 2009 ; Loschky et al., 2019).

Ainsi, l'objectif principal des travaux expérimentaux menés au cours de cette thèse était d'approfondir nos connaissances sur le rôle de la vision périphérique lors de la catégorisation visuelle de scènes en considérant les différences de traitement entre la vision centrale et la vision périphérique. Pour notre problématique de recherche, nous avons décidé d'opposer, même grossièrement, la vision centrale, c'est-à-dire la portion du champ visuel autour de la fixation (comprenant la fovéa et la parafovéa) *versus* la vision périphérique, c'est-à-dire le reste du champ visuel. Nous avons fait l'hypothèse générale que le système visuel utiliserait très rapidement les informations de basses résolution (les BFS) disponibles en vision périphérique afin d'accéder à une première représentation de la scène et de générer des prédictions sur son contenu. Ces prédictions guideraient ensuite la reconnaissance de l'information présente en vision centrale. Afin de tester notre hypothèse, nous avons tout d'abord mené des études comportementales sur des participants sains qui seront présentées dans les Chapitres 2, 3 et 4.

L'objectif du **Chapitre 2** était d'étudier spécifiquement l'intégration des BFS et HFS en vision périphérique qui n'a encore jamais été étudiée jusqu'ici. Nous avons étudié ce processus d'intégration en nous focalisant sur la manière dont les informations visuelles contenues dans les BFS influencent le traitement des HFS et inversement. Nous avons pour cela utilisé un paradigme expérimental étudiant l'interférence sémantique lors de la catégorisation de scènes dans des images hybrides (Kauffmann et al., 2017). Les images hybrides superposaient deux scènes, l'une filtrée en BFS et l'autre en HFS, qui étaient soit sémantiquement congruentes (e.g., une scène de ville en BFS appartenant à la catégorie des scènes "artificielles" superposée à une autre scène de ville en HFS appartenant donc à cette même catégorie), soit sémantiquement incongruentes (e.g., une scène de ville en BFS appartenant à la catégorie des scènes "artificielles" superposée à une scène de plage en HFS appartenant à la catégorie des scènes "naturelles"). Les participants devaient soit catégoriser la scène en BFS tout en ignorant la scène en HFS, soit catégoriser la scène en HFS tout en ignorant la scène en BFS. Cette dernière condition avait pour objectif de tester spécifiquement l'hypothèse selon laquelle le système visuel utiliserait très rapidement les BFS issues du champ visuel périphérique afin de générer des prédictions sur l'input visuel qui guideraient le traitement des HFS. Afin de comparer ce mécanisme de codage prédictif en vision centrale et en vision périphérique, nous avons présenté les scènes hybrides soit dans le champ visuel central, soit dans le champ visuel périphérique.

Cependant, l'utilisation d'images hybrides comporte certaines limites (stimuli ambigus et non écologiques) que nous avons décidé de prendre en considération. Dans la suite de cette

thèse, nous nous sommes donc plus particulièrement intéressés à l'intégration des BFS et des HFS en considérant leur disponibilité dans le champ visuel (les BFS majoritairement en vision périphérique et les HFS majoritairement en vision centrale). Ceci nous a tout d'abord amené dans le **Chapitre 3** à étudier distinctement la catégorisation de scènes en vision centrale et en vision périphérique afin de mieux comprendre de quelle façon ces différentes régions du champ visuel contribuent respectivement à la reconnaissance rapide de scènes. Dans ce chapitre, nous avons particulièrement testé l'hypothèse selon laquelle l'analyse visuelle d'une scène suivrait une séquence de traitement *coarse/peripheral-to-fine/central*. Pour cela, nous nous sommes inspirés des séquences dynamiques *coarse-to-fine* et *fine-to-coarse* utilisées par Kauffmann, Chauvin, Guyader et Peyrin, (2015). Nous avons utilisé de grandes photographies de scènes à partir desquelles nous avons construit cinq stimuli révélant différentes parties d'une même scène à travers des anneaux circulaires de différentes excentricités rétinienne. Les participants devaient catégoriser des séquences construites en présentant successivement les stimuli d'une même scène, de la périphérie au centre (séquence *peripheral-to-central*) ou du centre à la périphérie (séquence *central-to-peripheral*).

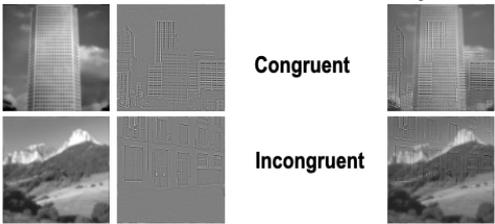
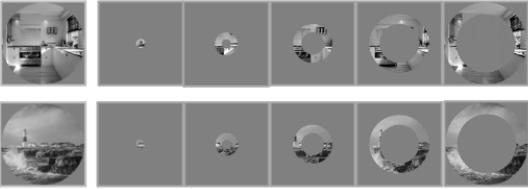
Dans le **Chapitre 4**, nous avons ensuite directement étudié comment les informations visuelles disponibles en vision périphérique influencent la catégorisation de scènes en vision centrale. Nous avons une fois de plus exploité le paradigme d'interférence sémantique. Nous avons utilisé de larges photographies de scènes d'intérieur et d'extérieur à partir desquelles nous avons extrait un disque contenant l'information présente en vision centrale et un anneau contenant l'information présente en vision périphérique. Nous avons ensuite construit des stimuli composés à la fois du disque central et de l'anneau périphérique. Le disque central et l'anneau périphérique pouvaient appartenir soit à la même catégorie sémantique (sémantiquement congruent), soit à une catégorie sémantique différente (sémantiquement incongruent). Les participants devaient soit catégoriser le disque central tout en ignorant l'anneau périphérique, soit catégoriser l'anneau périphérique tout en ignorant le disque central.

Enfin, dans le **Chapitre 5** de cette thèse, notre objectif était toujours d'étudier l'influence de la vision périphérique sur les traitements réalisés en vision centrale, mais avec une approche originale empruntée à la neuropsychologie. Il s'agissait d'étudier quelles sont les conséquences d'une perte de la vision périphérique sur les traitements réalisés en vision centrale pour en déduire son rôle dans des conditions normales. Nous avons pour cela testé des patients atteints d'un glaucome. Cette pathologie ophtalmique affecte en particulier la rétine périphérique et représente ainsi un bon modèle cognitif pathologique d'un système de reconnaissance visuelle dans lequel la vision périphérique a moins d'importance. Nous avons mené une expérience comportementale auprès de patients atteints d'un glaucome primitif à angle ouvert (GPAO) et de participants avec une vision normale appariés en âge afin de contrôler des différences imputables au vieillissement normal. Nous avons demandé aux participants de catégoriser en vision centrale des scènes filtrées en BFS et HFS, ainsi que des images hybrides de scènes. Nous avons comparé les performances des patients à celles des participants contrôles. Les résultats obtenus auprès de patients GPAO nous ont permis de préciser quels mécanismes impliqués dans la reconnaissance de scènes en vision centrale dans des conditions de vision normale (i.e. pour les participants avec une vision normale) sont en fait imputables à la vision périphérique. Tout en exploitant le modèle pathologique du glaucome

pour étudier le rôle de la vision périphérique, les résultats de cette expérience nous ont également permis de préciser les déficits cognitifs dont les patients souffrent en vision centrale.

Le Tableau 1.1 ci-dessous détaille les différentes expériences réalisées au cours de cette thèse et propose un aperçu des principaux stimuli visuels utilisés.

Tableau 1.1 : Ensemble des expériences réalisées au cours de cette thèse et aperçu des principaux stimuli visuels utilisés.

CHAPITRE 2 – ÉTUDE DE L'INTÉGRATION DES FRÉQUENCES SPATIALES EN VISION CENTRALE VERSUS PÉRIPHÉRIQUE LORS DE LA CATÉGORISATION DE SCÈNES			
<p>Expérience 1 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides physiquement dissimilaires</p> <p>Expérience 2 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides physiquement similaires</p> <p>Expérience 3 - Test de l'influence de la similarité physique (scènes hybrides physiquement similaire et dissimilaires)</p>	BFS	HFS	Hybride
	<p>Congruent</p> <p>Incongruent</p>		
CHAPITRE 3 – ETUDE DE LA CONTRIBUTION RESPECTIVE DE LA VISION CENTRALE ET DE LA VISION PÉRIPHÉRIQUE LORS DE LA CATÉGORISATION DE SCÈNES			
<p>Expériences 4 et 5 - Importance de la vision périphérique par rapport à la vision centrale</p> <ul style="list-style-type: none"> - Expériences 4 - Catégorisation de larges scènes et de parties de scènes de différentes excentricités rétinienne - Expériences 5 - Catégorisation de larges scènes dynamiques en expansion ou en contraction <p>Expériences 6, 7 et 8 - Importance de la vision périphérique par rapport à la vision centrale : Contrôle de biais méthodologiques</p> <ul style="list-style-type: none"> - Expériences 6 - Catégorisation de larges scènes et de parties de scènes de différentes excentricités rétinienne - Expériences 7 - Catégorisation de larges scènes dynamiques en expansion ou en contraction - Expériences 8 - Contrôle d'un effet de masquage latéral rétroactif 			
	<p>Disque entier Disque central Anneau 1 Anneau 2 Anneau 3 Anneau 4</p> <p>-----> <i>central-to-peripheral</i></p> <p>←----- <i>peripheral-to-central</i></p>		
CHAPITRE 4 – INFLUENCE DES INFORMATIONS SEMANTIQUES ET PHYSIQUES EN VISION PÉRIPHÉRIQUE SUR LA CATÉGORISATION DES SCÈNES EN VISION CENTRALE			
<p>Expérience 9 - Influences mutuelles des informations provenant de la vision périphérique et de la vision centrale lors de la catégorisation de scènes</p> <p>Expérience 10 - Nature de l'interférence de la vision périphérique sur la catégorisation de scènes en vision centrale</p>	Congruent	Incongruent	
			
CHAPITRE 5 – ETUDE DU RÔLE DE LA VISION PÉRIPHÉRIQUE À TRAVERS UN MODÈLE PATHOLOGIQUE : LE GLAUCOME PRIMAIRE À ANGLE OUVERT			
<p>Expérience 11 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides dans la vision centrale de patients atteints d'un glaucome</p>	BFS	HFS	Hybride
			

Chapitre 2 - Etude de l'intégration des fréquences spatiales en vision centrale versus périphérique lors de la catégorisation de scènes

Les études présentées dans le chapitre théorique ont donc montré que le système visuel s'appuie sur une première représentation grossière de la scène visuelle issue du signal en BFS pour générer rapidement des prédictions sur le contenu de la scène, et que les HFS jouent un rôle prépondérant en cas d'incertitude sur le stimulus visuel. Cependant, la majorité de ces travaux manipulant les fréquences spatiales ont été menées à l'aide de stimuli visuels présentés en vision centrale. Ils posent alors la question de la façon dont les BFS et HFS sont intégrés en un seul percept en vision périphérique.

Notre objectif dans ce premier chapitre expérimental était donc d'étudier comment l'intégration des fréquences spatiales évolue entre la vision centrale et la vision périphérique et plus précisément, d'étudier pour la première fois comment les informations visuelles contenues dans les BFS et les HFS s'influencent mutuellement en vision périphérique lors de la catégorisation de scènes. Nous avons pour cela testé l'interférence des informations sémantiques contenues dans les BFS sur le traitement des informations sémantiques contenues dans les HFS, et inversement, en vision centrale et en vision périphérique. Nous avons mené trois expériences comportementales (**Expériences 1, 2 et 3**) auprès de participants avec une vision considérée comme normale. Nous avons utilisé les images hybrides de Kauffmann et al. (2017) créées en superposant une scène filtrée en BFS et une autre scène filtrée en HFS. Les deux scènes formant une image hybride pouvaient soit être sémantiquement congruentes (e.g., une scène naturelle en BFS superposée à une autre scène naturelle en HFS), soit être sémantiquement incongruentes (e.g., une scène naturelle en BFS superposée à une scène artificielle en HFS). Dans une condition expérimentale, les participants devaient ignorer la scène en HFS (scène distractive) et catégoriser la scène en BFS (scène cible). A l'inverse, dans une autre condition expérimentale, les participants devaient ignorer la scène en BFS (scène distractive) et catégoriser la scène en HFS (scène cible). Ces deux conditions expérimentales nous ont permis d'évaluer comment le contenu sémantique d'une bande de fréquences spatiales influence l'autre. La seconde condition permet spécifiquement de tester l'hypothèse selon laquelle le système visuel utiliserait très rapidement les BFS issues du champ visuel périphérique afin de générer des prédictions sur l'input visuel qui guideraient le traitement des HFS. Afin de comparer le processus d'intégration des fréquences spatiales en vision centrale et celui qui intervient en vision périphérique, nous avons présenté les scènes hybrides soit dans le champ visuel central, soit dans le champ visuel périphérique (aléatoirement dans le champ visuel droit ou gauche). Ces expériences étaient donc également l'occasion de tester l'hypothèse d'une prédominance de l'hémisphère droit dans le traitement visuel prédictif.

Dans les études de Kauffmann et al. (2015, 2017), les deux scènes qui composent l'image hybride étaient soit physiquement similaires, soit physiquement dissimilaires. Dans l'**Expérience 1** de cette thèse, nous avons préféré étudier l'effet d'interférence sémantique en n'utilisant que les images hybrides physiquement dissimilaires. En effet, deux scènes physiquement similaires se caractérisent par la superposition spatiale - pixel par pixel - de l'information visuelle des deux scènes et par la similarité de leurs spectres d'amplitude. A cause de la perte de résolution spatiale en vision périphérique, nous pensons que la superposition spatiale entre les deux scènes de l'image hybride allait fortement gêner la distinction entre ces deux scènes en vision périphérique et donc la catégorisation de la scène cible, qu'elle soit sémantiquement congruente ou non congruente. Même cela devrait également être difficile pour les images physiquement dissimilaires, nous pensons que la catégorisation de la scène cible demeurerait toutefois réalisable. Contre toute attente, les participants de l'**Expérience 1** ont rapporté n'éprouver aucune difficulté à distinguer les deux scènes physiquement dissimilaires qui composent l'image hybride, que ce soit en vision centrale ou en vision périphérique. Dans l'**Expérience 2** de cette thèse, nous avons donc étudié l'effet d'interférence sémantique en utilisant cette fois-ci les images hybrides physiquement similaires. Dans l'**Expérience 3**, nous avons testé directement l'effet de la similarité physique sur l'interférence sémantique en incluant à la fois les images hybrides physiquement similaires et dissimilaires dans le même design expérimental.

Nous avons conscience que, dans ce chapitre, la description de la méthode et l'ensemble des résultats présentés peuvent apparaître particulièrement lourds pour le lecteur. C'est pourquoi, nous avons décidé de résumer d'ores et déjà les principaux résultats des trois expériences dans le Tableau 2.1 auquel le lecteur pourra d'emblée se reporter pour retenir les principaux résultats et suivre la progression des expériences.

Tableau 2.1 : Synthèse des résultats des **Expériences 1, 2 et 3**. Une croix est reportée dans le tableau lorsque l'hypothèse n'est pas infirmée. Les cases noires représentent les conditions expérimentales non testées dans l'expérience.

	Expérience 1 Hybrides physiquement DISSIMILAIRES		Expérience 2 Hybrides physiquement SIMILAIRES		Expérience 3 Hybrides physiquement DISSIMILAIRES et SIMILAIRES	
	%ER	TRs	%ER	TRs	%ER	TRs
Avantage du traitement des BFS ou HFS						
BFS au CENTRE						X
BFS en PERIPHERIE						X
BFS hybrides au CENTRE						X
BFS hybrides en PERIPHERIE	X		X		X	X
Interférence Sémantique (IntSem) du distracteur BFS ou HFS						
IntSem BFS au CENTRE	X		X		X	X
IntSem BFS en PERIPHERIE	X		X		X	X
IntSem BFS PERIPHERIE > CENTRE	X		X		X	
IntSem HFS au CENTRE	X		X		X	X
IntSem HFS en PERIPHERIE			X		X	X
IntSem HFS CENTRE > PERIPHERIE	X		X		X	X
CENTRE IntSem HFS > IntSem BFS			X		X	X
PERIPHERIE IntSem BFS > IntSem HFS	X				X	
Dominances inter-hémisphériques (comparaison entre champs visuels)						
Avantage des BFS dans CVG/HD						
Avantage des HFS dans CVD/HG						
Avantage des BFS par / HFS dans CVG/HD	X					
Avantage des HFS par / BFS dans CVD/HG						
IntSem BFS dans CVG/HD > CVD/HG						
IntSem HFS dans CVD/HG > CVG/HD						

1. **Expérience 1 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides physiquement dissimilaires**

1.1. Méthode

1.1.1. Participants

Quarante étudiants de l'Université Grenoble Alpes (23 femmes ; âge moyen : $20,85 \pm 2,47$) ont participé à cette expérience. Ils ont été répartis en deux groupes ($n = 20$) qui devaient respectivement catégoriser les BFS ou les HFS de l'image hybride. Nous avons fixé la taille de l'échantillon de chaque groupe en nous basant sur les résultats obtenus par Kauffmann et al. (2017). Dans cette expérience où les images hybrides n'étaient présentées qu'en vision centrale, l'interaction entre la congruence sémantique et le contenu en fréquences spatiales de la scène cible était significative pour l'analyse des pourcentages d'erreurs moyens avec une taille d'effet estimée par le calcul de l'éta-carré partiel (η^2) de 0.29. Nous avons ensuite converti la valeur de l'éta-carré partiel en d de Cohen afin de définir la taille d'échantillon. Ainsi, nous avons défini la taille de l'échantillon grâce à une analyse de puissance a priori pour test t de Student à échantillon indépendant réalisée avec GPower 3.1 (Faul et al., 2007), avec une taille d'effet attendue de 1.28 (d de Cohen obtenu), une puissance désirée de 0.99 et un seuil alpha de 0.05. Grâce à cette analyse, nous avons obtenu une taille d'échantillon minimale de 20 participants par groupe expérimental, soit un total de 40 participants. Nous avons ensuite exclu un participant de chaque groupe des analyses statistiques car leurs données différaient significativement des autres participants. Ces *outliers* ont été supprimées à l'aide du package *Outliers* utilisé sur Rstudio. Les analyses statistiques ont donc été conduites sur 38 participants (23 femmes ; âge moyen : $20,87 \pm 2,51$).

Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Tous les participants ont donné leur consentement libre et éclairé avant de participer à cette expérience qui a été menée en conformité avec le code d'éthique de l'*American Psychological Association* (APA) et approuvée par un comité d'éthique (CERGA, Comité d'éthique de la recherche Grenoble Alpes, IRB00010290-2018-10-16-50). A la fin de l'expérience, chaque participant recevait un bon de 0,25 point par quart d'heure d'expérience à valoir lors des examens universitaires. Cette procédure, mise en place et votée par l'UFR SHS de l'Université Grenoble Alpes, permet également aux étudiants de compléter leur formation pratique aux méthodes expérimentales utilisées en psychologie.

1.1.2. Stimuli

Nous avons utilisé les images hybrides précédemment utilisées dans l'étude de Kauffmann et al. (2017). Les images hybrides ont été construites en superposant une scène filtrée en BFS à une scène filtrée en HFS. Les scènes sont des petites photographies en couleurs de l'environnement visuel (résolution 256×256 pixels) sélectionnées au sein de la base

d'images du projet LabelMe du *MIT Computer Science and Artificial Intelligence Laboratory* (Oliva & Torralba, 2001). Les 2687 scènes de cette base d'images ont d'abord été classifiées en deux catégories distinctes : les scènes 'naturelles' (e.g., plages, champs, montagnes, etc.) et les scènes 'artificielles' (e.g., villes, autoroutes, etc.). La similarité physique des scènes a ensuite été évaluée sur la base (1) de la corrélation entre les valeurs d'intensité des pixels (corrélation IP) et (2) de la corrélation entre les valeurs d'intensité des pixels des spectres d'amplitude des scènes (corrélation SA). Deux scènes d'une image hybride étaient considérées comme physiquement similaires si les deux coefficients IP et SA étaient supérieurs à 0.6 (ce qui correspondait à seulement 1,59% des paires de scènes) et deux scènes étaient considérées comme physiquement dissimilaires si le coefficient de corrélation IP était inférieur à 0.01 et le coefficient de corrélation SA était inférieur à 0.6 (ce qui correspondait à seulement 1,37% des paires de scènes). Dans l'**Expérience 1** de cette thèse, nous avons seulement utilisé des images hybrides qui combinent deux scènes avec des propriétés physiques dissimilaires (Figure 2.1).

Les scènes qui composent les images hybrides ont toutes été converties en 256 niveaux de gris et filtrées afin de préserver soit les informations en BFS, soit les informations en HFS. La manipulation des images a été réalisée dans l'environnement MATLAB (Mathworks Inc., Sherborn, MA, USA). Pour appliquer la procédure de filtrage (voir **Encart 1**), la taille angulaire des images a été fixée à $6^\circ \times 6^\circ$ d'angle visuel. Les images filtrées ont été obtenues en multipliant la transformée de Fourier des images originales avec des filtres gaussiens. L'écart-type des filtres gaussiens dépendait de la fréquence de coupure pour une atténuation du signal de 3 dB. Pour les scènes en BFS, les fréquences spatiales inférieures ou égales à 24 cpi (2 cpd) étaient conservées. Pour les scènes en HFS, les fréquences spatiales supérieures ou égales à 36 cpi (6 cpd) étaient conservées. Les images une fois filtrées ont ensuite été normalisées pour obtenir une luminance moyenne de chaque image de 0,5 pour des valeurs d'intensité des pixels allant de 0 à 1 (c'est-à-dire une luminance moyenne de 128 sur une échelle de 256 niveaux de gris). A l'issue du filtrage des scènes, le contraste de luminance des scènes en BFS était plus élevé que celui des scènes en HFS (contraste *Root-Mean-Square* ou RMS moyen \pm écart-type pour les scènes BFS 'naturelles' : $0,24 \pm 0,06$; scènes BFS 'artificielles' : $0,24 \pm 0,05$; scènes HFS 'naturelles' : $0,06 \pm 0,02$; scènes HFS 'artificielles' : $0,07 \pm 0,02$). Pour finir, les images hybrides ont été créées en combinant une scène en BFS avec une autre scène en HFS.

Les deux scènes pouvaient être soit sémantiquement congruentes (condition Congruent ; par exemple, une scène naturelle en BFS superposée à une autre scène naturelle en HFS), soit sémantiquement incongruente (condition Incongruent ; par exemple, une scène naturelle en BFS superposée à une scène artificielle en HFS). Nous avons utilisé au total 80 images hybrides : 40 images hybrides sémantiquement congruentes (20 images hybrides superposant deux scènes artificielles et 20 images hybrides superposant deux scènes naturelles ; corrélation IP : 0.01 ± 0.00 ; corrélation SA : 0.43 ± 0.06) et 40 images hybrides sémantiquement incongruentes (superposant une scène artificielle et une scène artificielle ; corrélation IP : 0.01 ± 0.00 ; corrélation SA : 0.44 ± 0.05). La similarité physique (valeurs de corrélations IP et SA) était comparable entre les paires sémantiquement congruentes et incongruentes.

Dans cette expérience, nous avons également utilisé les 160 scènes filtrées qui composent les images hybrides (80 scènes en BFS et 80 scènes en HFS).

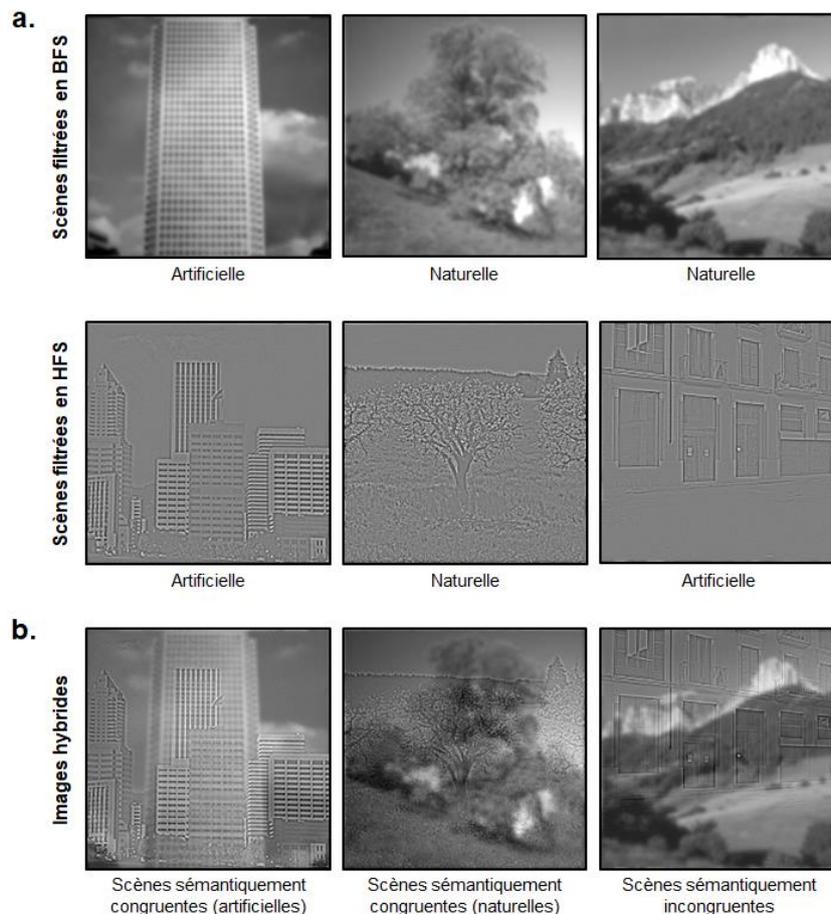


Figure 2.1 : Exemples de stimuli utilisés dans l'**Expérience 1** : **(a)** Scènes filtrées en BFS et en HFS ; **(b)** Images hybrides créées en superposant une scène filtrée en BFS et une autre scène filtrée en HFS. Les deux scènes sont physiquement dissimilaires. Elles pouvaient être soit sémantiquement congruentes (e.g., une scène naturelle en BFS superposée à une autre scène naturelle en HFS), soit sémantiquement incongruentes (e.g., une scène naturelle en BFS superposée à une scène artificielle en HFS).

1.1.3. Procédure

L'expérience a été réalisée dans un box expérimental du Laboratoire de Psychologie et NeuroCognition (LPNC) de l'Université de Grenoble Alpes. Les participants étaient installés dans l'obscurité face à un écran d'ordinateur. Ils posaient leur menton sur une mentonnière installée sur l'axe sagittal médian du participant, à 55,5 cm de l'écran afin de respecter la taille angulaire de 6° d'angle visuel des stimuli. L'expérience était programmée grâce au logiciel E-prime 2.0 (E-prime Psychology Software Tools Inc., Pittsburgh, USA) et elle était présentée sur un écran DELL ULTRASHARP (60 Hz) de 30 pouces (61,4 × 40 cm), avec un affichage de résolution 2560 × 1600 pixels.

Tous les participants réalisaient trois sessions expérimentales. Dans la **Session 1**, nous avons utilisé les scènes en BFS et HFS qui composaient les images hybrides. Nous les avons présentées individuellement en vision centrale (Figure 2.2a). Les participants devaient catégoriser les scènes comme appartenant soit à la catégorie 'naturelle', soit à la catégorie

'artificielle'. Cette session expérimentale avait pour but de nous assurer que les scènes en BFS et HFS des images hybrides étaient correctement catégorisées en dehors des effets d'interférence sémantique, mais aussi de comparer le traitement des BFS et HFS.

Dans la **Session 2**, nous avons utilisé les images hybrides que nous avons présentées en vision centrale (Figure [2.2a](#)). Les participants étaient répartis en deux groupes (n = 19 dans chaque groupe). Le premier groupe devait catégoriser la scène en BFS tout en ignorant la scène en HFS (condition Cible BFS/Distracteur HFS) et le deuxième groupe devait catégoriser la scène en HFS tout en ignorant la scène en BFS (condition Cible HFS/Distracteur BFS). Cette session expérimentale avait pour but d'étudier l'intégration des fréquences spatiales en vision centrale en nous assurant de répliquer les résultats de Kauffmann et al. (2017).

Dans la **Session 3**, nous avons à nouveau utilisé les images hybrides que nous avons présentées en vision périphérique (Figure [2.2b](#)), soit dans le champ visuel droit (CVD), soit dans le champ visuel gauche (CVG). L'excentricité des bords internes et externes des images hybrides était fixée respectivement à 2° et 8° d'excentricité rétinienne à partir du point de fixation central. Les participants qui devaient catégoriser la scène en BFS dans la Session 2 catégorisaient à nouveau la scène en BFS et les participants qui devaient catégoriser la scène en HFS dans la Session 2 catégorisaient à nouveau la scène en HFS. Cette session expérimentale avait pour but d'étudier l'intégration des fréquences spatiales en vision périphérique.

Dans les **Sessions 2 et 3**, chaque image hybride était présentée une fois dans sa version originale et une fois dans une version retournée en miroir (i.e. retournée sur l'axe vertical). En effet, un objet positionné à droite d'une scène (e.g., un arbre, un immeuble) n'aura pas la même excentricité si l'image est présentée dans le champ visuel droit ou gauche. Ainsi, en retournant les images en miroir, les objets d'une scène étaient présentés avec la même excentricité dans les deux champs visuels.

Pour chaque session expérimentale, un essai se déroulait comme suit (Figure [2.2](#)) : un point de fixation central (noir sur un fond gris de luminance 0,5) était présenté pendant 500 ms et suivi de la présentation du stimulus (Session 1 : scènes filtrées en BFS et en HFS ; Sessions 2 et 3 : images hybrides physiquement dissimilaires) sur un fond gris (luminance 0,5) pendant 100 ms, puis d'un écran gris (luminance 0,5) de 2000 ms au cours duquel les participants pouvaient répondre. Les participants devaient catégoriser chaque scène cible comme appartenant soit à la catégorie 'naturelle', soit à la catégorie 'artificielle' par l'intermédiaire des touches 1 et 2 du clavier d'ordinateur placé sur l'axe sagittal médian des participants. Pour chaque groupe expérimental, la moitié des participants appuyait sur la touche 1 avec l'index de la main droite pour la catégorie 'naturelle' et sur la touche 2 avec le majeur de la main droite pour la catégorie 'artificielle'. Les touches de réponse étaient contrebalancées pour l'autre moitié des participants. Les participants avaient pour consigne de répondre le plus correctement et rapidement possible.

La Session 1 comprenait 160 essais (40 scènes en BFS appartenant à la catégorie 'naturelle', 40 scènes en BFS appartenant à la catégorie 'artificielle' filtrées en BFS, 40 scènes en HFS appartenant à la catégorie 'naturelle' et 40 scènes en HFS appartenant à la catégorie 'artificielle'). La Session 2 comprenait 160 essais (Congruent : 40 images hybrides 'naturelle' dont 20 retournées en miroir, 40 images hybrides 'artificielle' dont 20 retournées en miroir ; Incongruent : 80 images hybrides dont 40 retournées en miroir). Enfin, la Session 3 comprenait

320 essais (condition Congruent : 80 images hybrides dans le CVD dont 40 retournées en miroir, et 80 images hybrides dans le CVG dont 40 retournées en miroir ; condition Incongruent : 80 images hybrides dans le CVD dont 40 retournées en miroir, et 80 images hybrides dans le CVG dont 40 retournées en miroir). L'expérience totale durait environ 50 minutes. Pour chaque essai, nous enregistrons l'exactitude de la réponse et le temps de réponse (en millisecondes). Avant de commencer chaque session expérimentale, les participants réalisaient quelques essais d'entraînement avec des stimuli non utilisés dans l'expérience afin de se familiariser avec la tâche et les stimuli.

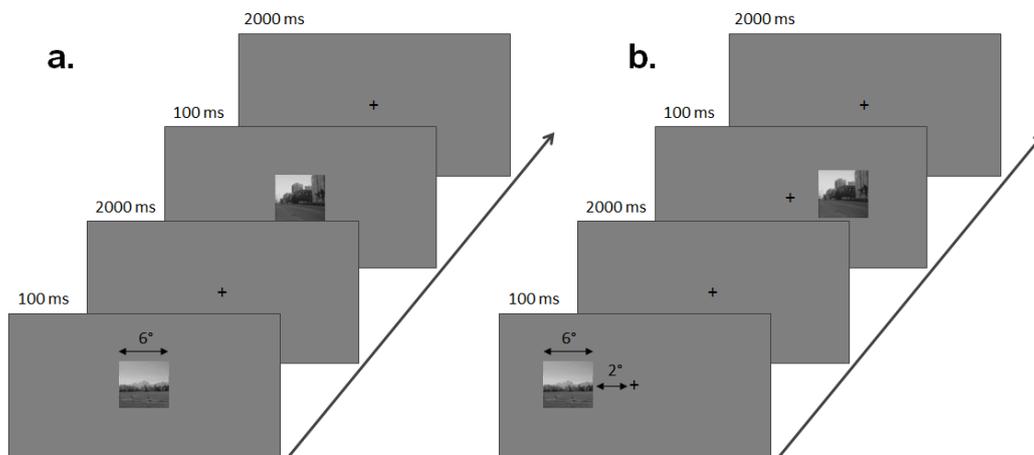


Figure 2.2 : (a) Schéma de la procédure expérimentale des Sessions 1 et 2 : Catégorisation d'une image filtrée ou d'une image hybride en vision centrale. (b) Schéma de la procédure expérimentale de la Session 3 : Catégorisation d'images hybrides en vision périphérique.

1.2. Hypothèses

- **Avantage du traitement des BFS sur les HFS lors de la catégorisation de scènes**

Pour la **Session 1**, sur la base du postulat théorique d'une préférence temporelle de l'analyse des BFS sur l'analyse des HFS (Breitmeyer, 1975 ; voir également Navon, 1977 ; Shulman & Wilson, 1987 ; Badcock et al., 1990) et de l'utilisation préférentielle d'une analyse *coarse-to-fine* des fréquences spatiales lors de la catégorisation de scènes (e.g., Kauffmann, Chauvin, Guyader & Peyrin, 2015 ; Musel et al., 2012 ; Schyns & Oliva, 1994), nous nous attendions à ce que la catégorisation d'une scène en BFS en vision centrale soit plus rapide que la catégorisation d'une scène en HFS. Suivant cette même hypothèse, dans les **Sessions 2** et **3**, la catégorisation de la scène en BFS qui compose une image hybride devrait être plus rapide que la catégorisation de la scène en HFS de cette même image hybride. Les modèles théoriques ne nous permettent pas de formuler d'hypothèses concernant un avantage des BFS sur l'exactitude des réponses de catégorisation (% Erreurs) car aucune des études que nous avons rapportées dans notre chapitre théorique n'en faisaient état, mais les travaux de Kauffmann et al. (2017) montrent que cet avantage s'observe également sur cette mesure.

- ***Interférence sémantique entre BFS et HFS lors de la catégorisation de scènes hybrides***

Plus important au regard de notre problématique de recherche, les **Sessions 2 et 3** nous ont permis d'étudier comment le traitement des BFS et celui des HFS s'influencent mutuellement en vision centrale et en vision périphérique. Pour cela, nous avons étudié l'effet d'interférence sémantique en fonction du contenu en fréquences spatiales de la scène cible, mais aussi de son champ visuel de présentation.

Nous avons vu dans le **Chapitre 1** de cette thèse que les signaux en BFS proviennent principalement du champ visuel périphérique et que les signaux en HFS proviennent principalement du champ visuel central. Nous avons donc fait l'hypothèse que le système visuel utiliserait très rapidement les BFS issues du champ visuel périphérique afin de former une première représentation de la scène et de générer des prédictions sur son contenu, prédictions qui guideraient le traitement des HFS. Dans notre expérience, une scène distractive en BFS devrait donc interférer plus fortement sur la catégorisation de la scène en HFS en vision périphérique qu'en vision centrale. Cela devrait se traduire pour la condition Cible HFS/Distracteur BFS par une interaction Congruence sémantique \times Position spatiale.

De plus, nous avons souligné précédemment que face à des images hybrides, il serait difficile, voire même contre-productif, d'ignorer les détails en HFS au profit des seules représentations grossières en BFS. On peut donc également s'attendre à ce que la scène distractive en HFS interfère sur la catégorisation de la scène en BFS. Sur la base des précédents travaux de Kauffmann et al. (2017), on peut plus précisément s'attendre à ce que l'interférence des HFS en vision centrale dépasse celle des BFS (interaction Congruence sémantique \times Fréquences spatiales pour la condition Centre). Néanmoins, étant donné la perte de résolution spatiale en vision périphérique, une scène distractive en HFS devrait interférer moins fortement sur la catégorisation de la scène en BFS en vision périphérique qu'en vision centrale. Cela devrait se traduire pour la condition Cible BFS/Distracteur HFS par une interaction Congruence sémantique \times Position spatiale pour la condition Cible BFS/Distracteur HFS. Le corollaire est que si nous observons une interférence des HFS sur les BFS même en vision périphérique, cet effet devrait être toutefois plus faible que l'interférence des BFS sur les HFS dans cette même région du champ visuel.

En d'autres termes, nous nous attendions à une interaction double entre la congruence sémantique (Congruent, Incongruent), le contenu en fréquences spatiales de la scène cible/distractive (Cible HFS/Distracteur BFS, Cible BFS/Distracteur HFS) et la position spatiale des images hybrides dans le champ visuel (Centre, Périphérie).

- ***Différences interhémisphériques***

Nous nous sommes également intéressés à la façon dont chaque hémisphère cérébral serait impliqué dans le processus d'intégration des fréquences spatiales. En effet, nous avons vu dans le **Chapitre 1** que les hémisphères cérébraux traitent différemment les fréquences spatiales. Les scènes en BFS sont catégorisées plus rapidement dans le champ visuel gauche (CVG, initialement traité par l'hémisphère droit) que dans le champ visuel droit, alors que les

scènes en HFS sont catégorisées plus rapidement dans le champ visuel droit (CVD, initialement traité par l'hémisphère gauche) que dans le champ visuel gauche (Iidaka et al., 2004 ; Peyrin et al., 2003 ; Peyrin et al., 2004 ; Peyrin, Chokron et al., 2006 ; Peyrin, Mermillod et al., 2006). Sur la base de ces travaux, nous nous attendions dans la **Session 3** à ce que les performances de catégorisation de la scène en BFS contenue dans une image hybride soient meilleures lorsque l'image hybride était présentée dans le CVG que dans le CVD et inversement, à ce que les performances de catégorisation de la scène en HFS contenue dans une image hybride soient meilleures lorsque l'image hybride est présentée dans le CVD que dans le CVG. En d'autres termes, nous nous attendions à une interaction entre le contenu en fréquences spatiales de la scène cible et le champ visuel de présentation des images hybrides.

De plus, des études en neuroimagerie montrent que l'hémisphère droit serait particulièrement impliqué dans l'analyse de type *coarse-to-fine* d'une scène (Peyrin et al., 2010 ; Kauffmann et al., 2015). Ces résultats suggèrent que c'est surtout l'hémisphère droit qui utiliserait très rapidement les BFS afin de former une première représentation de la scène et de générer des prédictions. Ainsi, nous nous attendions toujours dans la **Session 3** à ce que l'effet d'interférence sémantique des BFS (distracteur) sur les HFS (scène cible) attendu en vision périphérique soit plus important lorsqu'une image hybride était présentée dans le CVG que dans le CVD (interaction Congruence sémantique \times champ visuel pour la condition Cible HFS/Distracteur BFS). En suivant cette logique, si nous observons un effet d'interférence des HFS (distracteur) sur les BFS (scène cible) en vision périphérique, on peut également imaginer que celui-ci soit plus important lorsqu'une image hybride est présentée dans le CVD que dans le CVG (interaction Congruence sémantique \times Champ visuel pour la condition Cible BFS/Distracteur HFS). En d'autres termes, nous nous attendions à une interaction double entre la congruence sémantique (Congruent, Incongruent), le contenu en fréquences spatiales de la scène cible/distractrice (Cible HFS/Distracteur BFS, Cible BFS/Distracteur HFS) et le champ visuel de présentation des images hybrides (CVG, CVD).

1.3. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale de chaque session, nous avons calculé les pourcentages d'erreurs moyens (%ERm) et les temps de réponses correctes moyens en millisecondes (TRm). Afin de réduire l'effet des valeurs extrêmes sur les résultats, les temps de réponses atypiques pour chaque participant dans chaque condition ont été supprimés à l'aide de la méthode des boîtes à moustaches. Cette méthode permet de mettre en évidence pour chaque participant et dans chaque condition les données supérieures ou inférieures à 3 fois l'intervalle interquartile. A l'aide de cette méthode, nous avons seulement supprimé 1,49% des temps de réponse pour la Session 1, 1,39% des temps de réponse pour la Session 2 et 0,71% des temps de réponse pour la Session 3. Le seuil de significativité a été fixé à 0,05. La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2). Afin de faciliter la lecture de cette partie, nous avons consigné les résultats relatifs à nos hypothèses dans le Tableau [2.1](#).

1.3.1. Session 1 : Catégorisation de scènes filtrées en BFS et en HFS en vision centrale

Pour l'analyse de la **Session 1**, un test *t* de Student pour échantillons appariés a été réalisé sur les valeurs de %ERm et les TRm (Tableau 2.2) avec comme variable intra-sujets la Fréquence spatiale de la scène à catégoriser (BFS ; HFS).

Tableau 2.2 : Résultats de la Session 1 de l'**Expérience 1**. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS).

BFS		HFS	
%ERm	TRm	%ERm	TRm
3,91 ± 0,50	637 ± 15	3,52 ± 0,44	642 ± 14

Quelle que soit la condition expérimentale, les taux d'erreurs étaient bas (inférieurs à 4%). Contrairement à notre hypothèse d'un avantage du traitement des BFS sur le traitement des HFS, la différence de performances de catégorisation entre les BFS et les HFS n'était pas significative, ni pour les %ERm ($t(37) = 0.86, p = .396$; Figure 2.3a), ni pour les TRm ($t(37) = -1,44, p = .157$; Figure 2.3b).

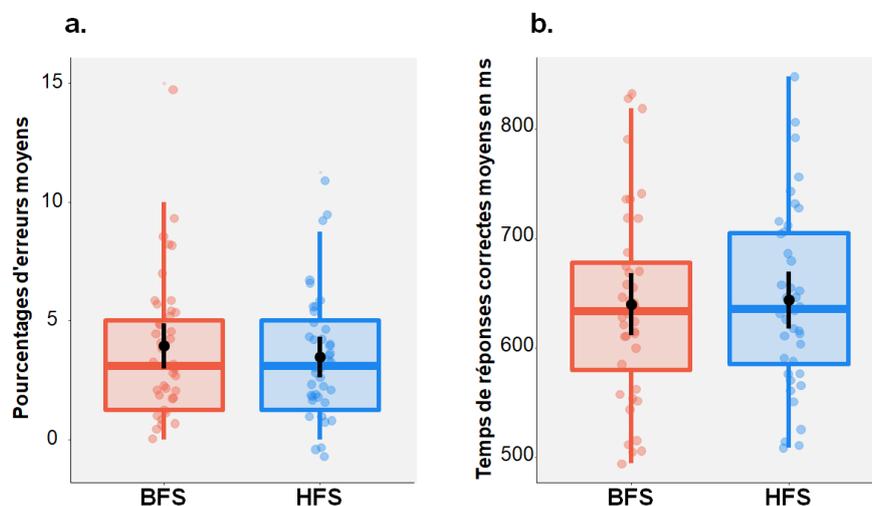


Figure 2.3 : Résultats de la Session 1 de l'**Expérience 1** (catégorisation de scènes filtrées en BFS et en HFS en vision centrale) : **(a) Pourcentages d'erreurs moyens** et **(b) Temps de réponses correctes moyens en ms** en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

1.3.2. Sessions 2 et 3 : Catégorisation d'images hybrides en vision centrale et en vision périphérique

Dans cette partie, nous avons réuni les résultats des **Sessions 2 et 3** dans la même analyse afin de comparer directement les effets obtenus en vision centrale et en vision périphérique. Ainsi, une analyse de la variance (ANOVA) de type $2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm (Tableau 2.3) avec comme facteur intra-sujet la Congruence sémantique (Congruent ; Incongruent) et la Position spatiale des images hybrides dans le champ visuel (Centre ; Périphérie) et comme facteur inter-sujet la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS).

Tableau 2.3 : Résultats des Sessions 2 et 3 de l'Expérience 1. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction du contenu en Fréquences spatiales de la cible et du distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et de la Position spatiale de l'image hybride dans le champ visuel (Centre ; Périphérie).

		Centre		Périphérie	
		%ERm	TRm	%ERm	TRm
Cible BFS/Distracteur HFS	Congruent	6,91 ± 1,26	739 ± 25	8,82 ± 1,23	722 ± 21
	Incongruent	14,09 ± 1,94	780 ± 27	13,22 ± 2,77	758 ± 23
Cible HFS/Distracteur BFS	Congruent	7,43 ± 1,26	735 ± 25	14,14 ± 1,23	735 ± 21
	Incongruent	12,96 ± 1,94	752 ± 27	27,47 ± 2,77	759 ± 23

- Analyse des pourcentages d'erreurs moyens

Concernant tout d'abord notre hypothèse d'un avantage du traitement des BFS lors de la catégorisation de scène, l'ANOVA menée sur les %ERm a révélé un effet principal de la Fréquence spatiale Cible/Distracteur ($F(1,36) = 5,25, p = .028, \eta^2 = .13$). Les participants faisaient en moyenne moins d'erreurs pour catégoriser les scènes cibles en BFS ($10,76 \pm 1,46\%$) que pour catégoriser les scènes cibles en HFS ($15,50 \pm 1,46\%$). La Fréquence spatiale Cible/Distracteur interagissait avec la Position spatiale des images hybrides ($F(1,36) = 31,67, p < .001, \eta^2 = .47$). Les comparaisons planifiées ont montré que, quel que soit la condition de congruence sémantique, les participants faisaient moins d'erreurs pour catégoriser les scènes cibles en BFS que les scènes cibles en HFS lorsqu'elles étaient présentées en vision périphérique ($F(1,36) = 16,15, p < .001$; BFS : $11,02 \pm 2,43\%$; HFS : $20,81 \pm 2,43\%$), mais pas lorsqu'elles étaient présentées en vision centrale (BFS : $10,49 \pm 2,06\%$; HFS : $10,19 \pm 2,06\%$; $F(1,36) < 1, p = .887$).

Concernant ensuite nos hypothèses d'interférence sémantique, l'ANOVA a bien révélé un effet principal de la Congruence sémantique ($F(1,36) = 38,77, p < .001, \eta^2 = .52$; Figure 2.4). Les participants faisaient en moyenne plus d'erreurs pour catégoriser des images hybrides incongruentes ($16,93 \pm 2,14\%$) que congruentes ($9,32 \pm 1,08\%$). Cet effet indique un effet d'interférence sémantique, c'est-à-dire que lorsque deux scènes étaient sémantiquement incongruentes, la scène distractive à ignorer était quand même traitée au niveau sémantique et

interférait sur la catégorisation de la scène cible. La Fréquence spatiale Cible/Distracteur n'interagissait pas avec la Congruence sémantique ($F(1,36) = 2,21, p = .145$) suggérant un effet d'interférence quel que soit le contenu en fréquences spatiales de la scène distractive à ignorer.

Cependant, comme attendu, nous avons observé que ces effets d'interférence étaient modulés par la position des images hybrides dans le champ visuel (interaction double entre la Congruence sémantique, la Fréquence spatiale Cible/Distracteur et la Position spatiale, $F(1,36) = 9,96, p = .003, \eta^2 = .22$). Les comparaisons planifiées ont révélé une interférence des BFS sur les HFS aussi bien en vision centrale (Congruent : $7,43 \pm 1,26\%$; Incongruent : $12,96 \pm 1,94\%$; $F(1,36) = 13,66, p < .001$) qu'en vision périphérique (Congruent : $14,14 \pm 1,23\%$; Incongruent : $27,47 \pm 2,77\%$; $F(1,36) = 27,16, p < .001$). L'interaction significative entre la Congruence sémantique et la Position spatiale pour la condition Cible HFS/Distracteur BFS ($F(1,36) = 10,86, p = .002$) suggère que l'effet d'interférence sémantique des BFS sur les HFS était plus fort en vision périphérique qu'en vision centrale. Par contre l'interférence des HFS sur les BFS n'était significative qu'en vision centrale (Centre Congruent : $6,91 \pm 1,26\%$; Centre Incongruent : $14,08 \pm 1,94\%$; $F(1,36) = 23,01, p < .001$; Périphérie Congruent : $8,88 \pm 1,23\%$; Périphérie Incongruent : $13,22 \pm 2,77\%$; $F(1,36) = 2,97, p = .09$). Ainsi, en vision périphérique, seules les BFS interféraient sur les HFS. En vision centrale, l'interaction Congruence sémantique \times Fréquence spatiale n'était pas significative ($F(1,36) < 1, p = .442$). Ce dernier résultat suggère que l'interférence de la scène distractive en HFS n'était pas plus forte que l'interférence de la scène distractive en BFS en vision centrale, ce qui ne reproduit pas les résultats de Kauffmann et al. (2017).

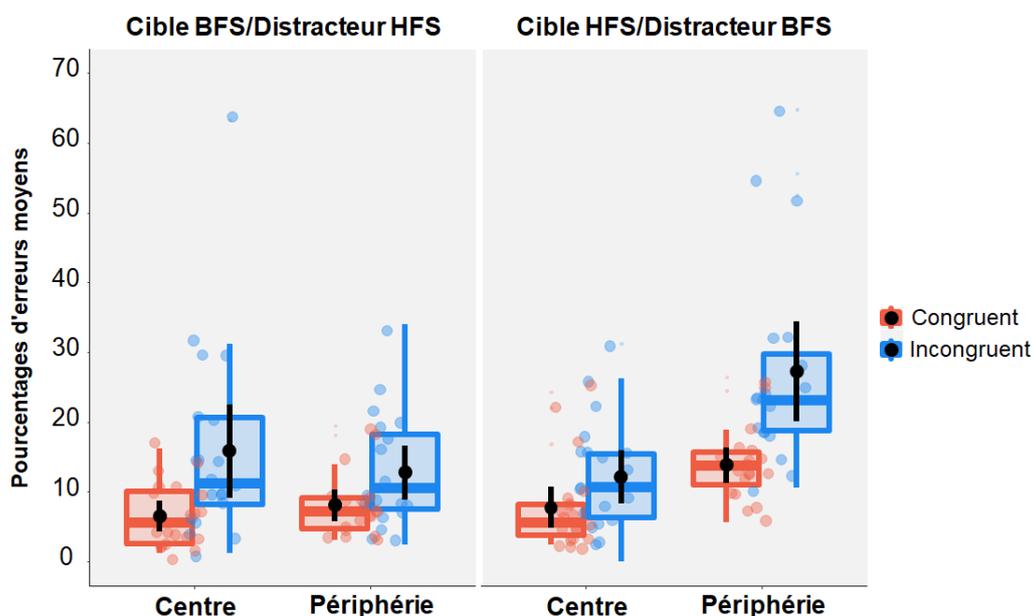


Figure 2.4 : Résultats des Sessions 2 et 3 de l'Expérience 1 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : **Pourcentages d'erreurs moyens** en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et de la Position spatiale des images hybrides dans le champ visuel (Centre ; Périphérie). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

- Analyse des temps de réponses correctes

Concernant tout d'abord notre hypothèse d'un avantage du traitement des BFS, l'ANOVA menée sur les TRm n'a pas montré d'effet principal de la Fréquence spatiale Cible/Distracteur ($F(1,36) < 1, p = .886$) et ce facteur n'interagissait pas avec la Position spatiale des images hybrides dans le champ visuel ($F(1,36) = 1,50, p = .228$).

Concernant ensuite nos hypothèses d'interférence sémantique, l'effet principal de la Congruence sémantique était significatif ($F(1,36) = 29,96, p < .001, \eta^2 = .45$; Figure 2.5). Les participants étaient plus rapides pour catégoriser des images hybrides congruentes (733 ± 22 ms) qu'incongruentes (762 ± 24 ms). Cet effet indique une nouvelle fois un effet d'interférence sémantique. Ce facteur n'interagissait pas avec la Fréquence spatiale Cible/Distracteur ($F(1,36) = 2,84, p < .100$). Enfin, l'interaction double Congruence sémantique \times Fréquence spatiale Cible/Distracteur \times Position spatiale n'était pas significative ($F(1,36) < 1, p = .503$).

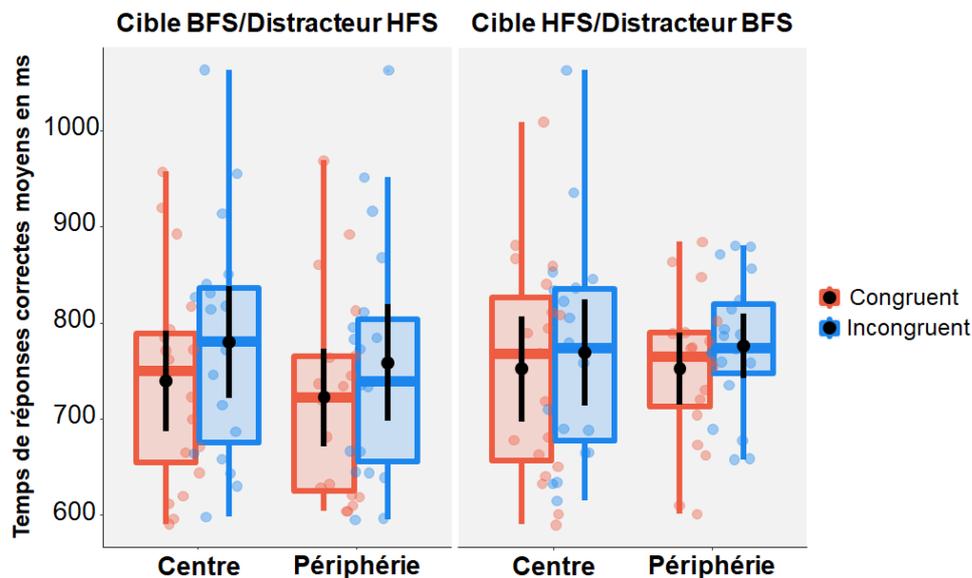


Figure 2.5 : Résultats des Sessions 2 et 3 de l'**Expérience 1** (catégorisation d'images hybrides en vision centrale et en vision périphérique) : **Temps de réponses correctes moyens en ms** en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et de la Position spatiale des images hybrides dans le champ visuel (Centre ; Périphérie). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

1.3.3. Session 3 : Catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche

Dans cette partie, nous avons analysé uniquement les résultats de la **Session 3** afin de comparer directement les effets obtenus entre le champ visuel droit (CVD) et le champ visuel

gauche (CVG). Cette analyse nous permettait de répondre à notre dernière hypothèse sur les différences interhémisphériques dans le processus d'intégration des fréquences spatiales¹. Ainsi, une ANOVA de type $2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm (Tableau 2.4) avec comme facteur intra-sujet la Congruence sémantique (Congruent ; Incongruent) et le Champ visuel/Hémisphère de présentation des images hybrides (CVD/HG ; CVG/HD) et comme facteur inter-sujet la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS).

Tableau 2.4 : Résultats de la Session 3 de l'Expérience 1. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction du contenu en Fréquences spatiales de la cible et du distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et du Champ visuel de présentation des images hybrides (CVD ; CVG).

		CVD		CVG	
		%ERm	TRm	%ERm	TRm
Cible BFS/Distracteur HFS	Congruent	10,26 ± 1,40	720 ± 22	7,37 ± 1,45	724 ± 21
	Incongruent	13,75 ± 3,01	760 ± 24	12,70 ± 2,68	757 ± 24
Cible HFS/Distracteur BFS	Congruent	14,08 ± 1,40	734 ± 22	14,21 ± 1,45	736 ± 21
	Incongruent	26,32 ± 3,01	753 ± 24	28,62 ± 2,68	765 ± 24

L'ANOVA menée sur les %ERm n'a pas révélé d'effet principal du Champ visuel de présentation des images hybrides ($F(1,36) < 1$, $p = .606$; Figure 2.6), mais ce facteur interagissait comme attendu avec la Fréquence spatiale Cible/Distracteur ($F(1,36) = 4,81$, $p = .035$, $\eta^2 = .12$). Les comparaisons planifiées ont montré que les participants faisaient en moyenne moins d'erreurs pour catégoriser les scènes cibles en BFS que les scènes cibles en HFS à la fois lorsqu'elles étaient présentées dans le CVD ($F(1,36) = 9,13$, $p = .005$) et dans le CVG ($F(1,36) = 23,26$, $p < .001$). Toutefois, l'interaction suggère que cette différence était plus

¹ Nous avons mené une **première analyse préliminaire** où nous avons inclus la variable contrôle Miroir (i.e. chaque image hybride était présentée une fois dans une version originale et une fois dans une version retournée en miroir) afin de nous assurer que les performances obtenues dans la condition Miroir étaient équivalentes à celles obtenues dans la condition Originale quel que soit le champ visuel de présentation des stimuli (CVD ; CVG). L'ANOVA menée sur les %ERm et sur les TRm n'a révélé ni d'interaction Miroir × Champ visuel (%ERm : $F(1,36) < 1$; TRm : $F(1,36) < 1$), ni d'interaction Miroir × Champ visuel × Fréquence spatiale Cible/Distracteur (%ERm : $F(1,36) = 4,12$, $p = .06$; TRm : $F(1,36) < 1$), ni d'interaction Miroir × Champ visuel × Fréquence spatiale Cible/Distracteur × Congruence sémantique (%ERm : $F(1,36) < 1$; TRm : $F(1,36) < 1$). Dans une **deuxième analyse préliminaire**, nous avons inclus la variable contrôle Sexe des participants (Femme ; Homme) afin de nous assurer que cette variable n'influence pas les différences de performance entre les hémisphères. En effet, il existe un large débat dans la littérature concernant la façon dont le sexe des participants influence les différences de traitement interhémisphériques (Hausmann et al., 2000 ; Voyer, 1996). Concernant spécifiquement la catégorisation de scènes en BFS et HFS, Peyrin, Chokron et al. (2006) ont montré que les hommes étaient plus latéralisés que les femmes. L'analyse de la variance (ANOVA) menée sur les %ERm et sur les TRm n'a révélé ni d'interaction Sexe × Champ visuel (%ERm : $F(1,36) = 2,23$, $p = .139$; TRm : $F(1,36) < 1$), ni d'interaction Sexe × Champ visuel × Fréquence spatiale Cible/Distracteur (%ERm : $F(1,36) = 1,47$, $p = .233$; TRm : $F(1,36) < 1$), ni d'interaction Sexe × Champ visuel × Fréquence spatiale Cible/Distracteur × Congruence sémantique (%ERm : $F(1,36) = 1,01$, $p = .321$; TRm : $F(1,36) < 1$). Ces variables n'ont donc pas été considérées dans la suite des analyses.

grande dans le CVG (BFS : $10,03 \pm 2,36\%$; HFS : $21,41 \pm 2,36\%$) que dans le CVD (BFS : $12,01 \pm 2,71\%$; HFS : $20,20 \pm 2,71\%$). Par contre, il n'y avait pas de différence significative entre le CVD et le CVG, à la fois pour catégoriser les scènes cibles en BFS ($F(1,36) = 3,68, p = .063$) et les scènes cibles en HFS ($F(1,36) = 1,40, p = .245$).

Enfin, l'interaction double attendue Champ visuel de présentation des stimuli \times Fréquence spatiale Cible/Distracteur \times Congruence sémantique n'était pas significative ($F(1,36) < 1, p = .902$).

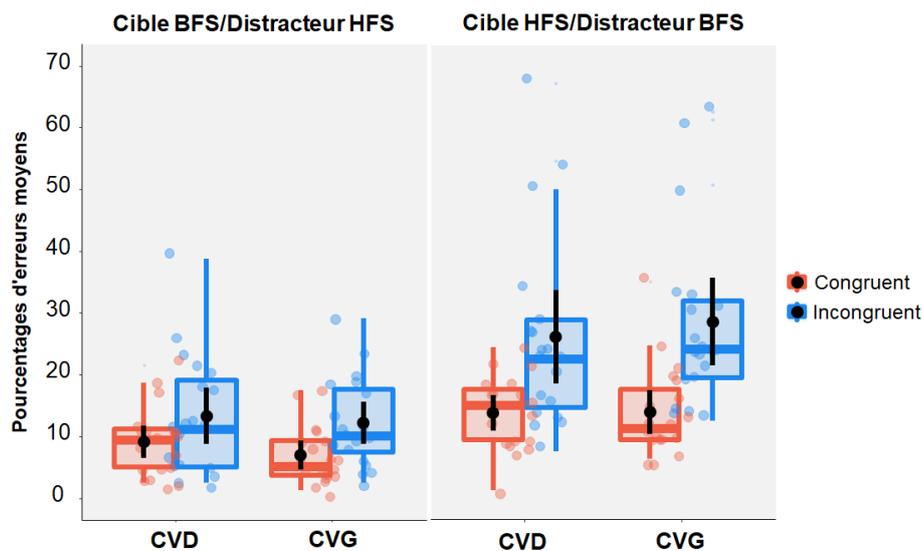


Figure 2.6 : Résultats de la Session 3 de l'Expérience 1 (catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche) : **Pourcentage d'erreurs moyens** en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et du Champ visuel de présentation des images hybrides (CVD ; CVG). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

L'ANOVA menée sur les TRm n'a pas révélé d'effet principal du Champ visuel de présentation des stimuli ($F(1,36) < 1, p = .362$; Figure 2.7), ni d'interaction Champ visuel \times Fréquence spatiale Cible/Distracteur ($F(1,36) < 1, p = .410$), ni d'interaction Champ visuel de présentation des stimuli \times Fréquence spatiale Cible/Distracteur \times Congruence sémantique ($F(1,36) < 1, p = .390$).

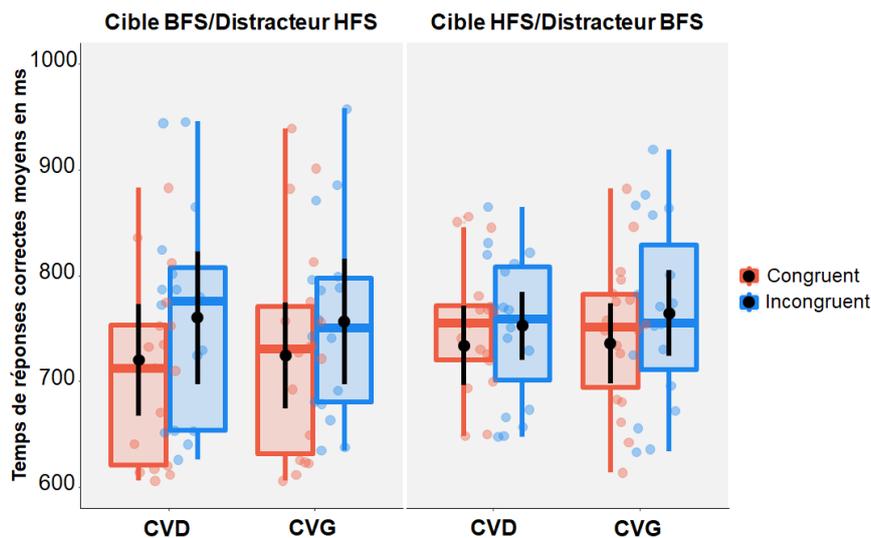


Figure 2.7 : Résultats de la Session 3 de l'Expérience 1 (catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche) : Temps de réponses correctes moyennes en ms en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et du Champ visuel de présentation des images hybrides (CVD ; CVG). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

1.4. Discussion

Cette expérience avait pour objectif d'étudier comment l'intégration des fréquences spatiales évolue entre la vision centrale et la vision périphérique et plus précisément, comment les informations visuelles contenues dans les BFS et les HFS s'influencent mutuellement en vision centrale et en vision périphérique lors de la catégorisation de scènes. Pour cela, nous avons utilisé un paradigme d'interférence sémantique qui repose sur la catégorisation d'images hybrides. Nous avons utilisé les images hybrides physiquement dissimilaires de Kauffmann et al. (2017). Ces images superposent une scène filtrée en BFS et une autre scène filtrée en HFS. Nous avons manipulé la congruence sémantique entre les deux scènes des images hybrides. Les deux scènes pouvaient soit être sémantiquement congruentes (e.g., une scène naturelle en BFS superposée à une autre scène naturelle en HFS), soit être sémantiquement incongruentes (e.g., une scène naturelle en BFS superposée à une scène artificielle en HFS). Ainsi, ce paradigme expérimental nous a permis d'évaluer l'effet d'interférence sémantique de la scène à ignorer sur la scène à catégoriser. Plus précisément, ce paradigme expérimental nous a permis d'étudier comment une bande de fréquence spatiale, qui doit être ignorée, influence l'autre bande de fréquence spatiale, qui doit être traitée, en fonction de leur position dans le champ visuel (central vs. périphérique, à gauche ou à droite). Nous avons également testé la catégorisation des scènes en BFS et HFS qui composent les images hybrides afin de nous assurer que ces scènes étaient correctement catégorisées en dehors des effets d'interférence sémantique, mais aussi de comparer le traitement des BFS et HFS.

- **Avantage des BFS sur les HFS lors de la catégorisation de scènes**

Les résultats de la **Session 1** (catégorisation de scènes filtrées en BFS et en HFS en vision centrale) ont montré que les participants étaient capables de catégoriser très précisément les scènes filtrées en BFS et en HFS. Par contre, nous n'avons pas observé de catégorisation plus rapide d'une scène en BFS qu'en HFS. On remarque toutefois que les pourcentages d'erreurs extrêmement bas dans cette session (inférieurs à 4%) indiquent que la tâche de catégorisation était trop facile à réaliser. Cet effet plafond pourrait ainsi augmenter la difficulté à détecter l'effet attendu sur les temps de réponses correctes. Autrement dit, nous aurions pu observer l'effet attendu dans le cas où la tâche aurait été plus contraignante.

Par ailleurs, les scènes filtrées étaient des stimuli non ambigus, présentés en vision centrale et pendant un temps relativement long (100 ms) si on considère l'ensemble des travaux qui ont été menés sur la reconnaissance ultra-rapide de scènes visuelles. Rappelons que les travaux de Schyns et Oliva (1994) suggèrent que les BFS domineraient la perception visuelle lorsque peu de temps est alloué au traitement (par exemple, seulement 30 ms) alors que les HFS seraient privilégiées dès lors que plus de temps de traitement serait possible. Dans notre expérience, le poids alloué au traitement des HFS pourrait alors augmenter au même niveau que celui des BFS. D'ailleurs, nous avons émis cette hypothèse principalement sur la base des travaux Kauffmann et al. (2017) dont la présente expérience est en partie une réplique. Mais il faut noter que ces auteurs présentaient en fait leurs scènes filtrées soit pendant 100 ms, soit pendant 30 ms, imposant ainsi une contrainte de temps de traitement plus importante dans la condition de présentation à 30 ms, et que leurs résultats n'ont en fait montré un avantage des BFS sur les HFS que lorsque les scènes filtrées étaient présentées pendant 30 ms (pour les pourcentages d'erreurs moyens uniquement). Ainsi, l'utilisation préférentielle d'une bande de fréquence spatiale dépendrait des contraintes d'une tâche visuelle donnée, comme le temps de présentation des stimuli.

De la même façon, les résultats de la **Session 2** (catégorisation d'images hybrides en vision centrale) n'ont pas révélé de différence significative entre la catégorisation des scènes cibles en BFS et des scènes cibles en HFS en vision centrale, même si les images hybrides sont considérées comme des stimuli plus complexes, augmentant la difficulté de la tâche. Cependant, nous avons choisi d'utiliser uniquement des images hybrides physiquement dissimilaires dans cette expérience, réduisant ainsi la forte ambiguïté présente dans des images hybrides physiquement similaires.

Pris ensemble, les résultats des **Sessions 1** et **2** montrent que la catégorisation de scènes en vision centrale est aussi efficace lorsqu'elle est basée sur les BFS et sur les HFS. Nous pouvons toutefois supposer qu'avec une tâche plus contraignante (en réduisant le temps de présentation, en utilisant des stimuli plus ambigus tels que des images hybrides physiquement similaires), la catégorisation de scènes serait plus efficace lorsqu'elle est basée sur les BFS que sur les HFS.

Les résultats de la **Session 3** (catégorisation d'images hybrides en vision périphérique) ont montré que les participants étaient cette fois-ci plus précis pour catégoriser les scènes cibles en BFS d'une image hybride par rapport aux scènes cibles en HFS de cette même image hybride. Ce résultat semble cohérent avec les propriétés anatomiques du système visuel. Etant

donné que les BFS sont majoritairement extraites de la vision périphérique, il n'est donc pas surprenant d'observer un avantage des BFS par rapport aux HFS en vision périphérique.

- ***Effets d'interférence sémantique entre BFS et HFS lors de la catégorisation de scènes hybrides***

Les résultats des **Sessions 2 et 3** (catégorisation d'images hybrides en vision centrale et en vision périphérique) ont montré que les performances de catégorisation de la scène cible étaient meilleures lorsque les deux scènes d'une image hybride étaient sémantiquement congruentes qu'incongruentes.

Cet effet d'interférence sémantique était modulé par la position des images hybrides dans le champ visuel lorsque la scène cible était en HFS et la scène distractive en BFS. En effet, conformément à notre hypothèse, la scène distractive en BFS interférait plus fortement sur la catégorisation de la scène cible en HFS en vision périphérique qu'en vision centrale. Ainsi, le système visuel utiliserait très rapidement les BFS à la fois disponibles en vision centrale et en vision périphérique afin de générer des prédictions, mais de façon plus importante en vision périphérique. Ces prédictions seraient validées ou non par les informations plus lentes en HFS. Plus précisément, dans le cas où les deux scènes d'une image hybride sont sémantiquement congruentes, ces prédictions permettraient de faciliter le traitement des HFS, surtout en vision périphérique où les BFS sont majoritairement extraites. A l'inverse, dans le cas où les deux scènes d'une image hybride sont sémantiquement incongruentes, ces prédictions pourraient être erronées et gêneraient la catégorisation des HFS. Autrement dit, lorsque la catégorisation d'une image hybride s'effectue en vision périphérique, le système visuel augmenterait le poids de l'interférence des prédictions issues de l'analyse en BFS (dans la tâche de catégorisation d'une scène cible en HFS) et réduirait le poids d'une analyse détaillée portée par les HFS.

Par ailleurs, sur la base des résultats de Kauffmann et al. (2017) obtenus en vision centrale uniquement, nous nous attendions à ce qu'une scène distractive en HFS interfère également sur la catégorisation d'une scène cible en BFS et que cet effet soit plus grand que celui des BFS sur les HFS. En effet, selon les auteurs, lors de la catégorisation d'images complexes et incohérentes telles que les images hybrides, il serait difficile, voire même contre-productif, d'ignorer l'information en HFS au profit des seules représentations grossières issues de l'analyse des BFS. L'effet d'interférence sémantique très fort des HFS sur les BFS suggérerait une influence irrépressible des informations en HFS lors de la catégorisation d'images hybrides, malgré les informations a priori plus pertinentes contenues dans les scènes cibles en BFS. En cas d'incertitude ou d'ambiguïté sur le stimulus visuel, comme c'est le cas avec des images hybrides, le traitement des détails contenus dans les HFS pourraient être plus pertinents pour démêler les interprétations et pour effectuer une tâche de catégorisation de scènes. Néanmoins, étant donné la perte de résolution spatiale en vision périphérique, nous avons fait l'hypothèse que la scène distractive en HFS interfère moins fortement en vision périphérique qu'en vision centrale. Nous avons effectivement observé un effet d'interférence des HFS sur les BFS seulement en vision centrale, suggérant ainsi des effets d'interférence plus importants dans cette partie du champ visuel qu'en vision périphérique. Par contre,

contrairement aux précédents résultats de Kauffmann et al. (2017), l'interférence des HFS sur les BFS observée en vision centrale n'était pas supérieure à celle des BFS sur les HFS. Alors comment pouvons-nous expliquer ces différences de résultats entre notre expérience et celle de Kauffmann et al. (2017) ? Une explication pourrait être que nous avons utilisé uniquement des images hybrides physiquement dissimilaires alors que Kauffmann et al. (2017) utilisaient à la fois des images hybrides physiquement dissimilaires et physiquement similaires. Les auteurs ont montré que l'effet d'interférence sémantique des HFS sur la catégorisation des scènes en BFS était plus grand lorsque les deux scènes d'une image hybride étaient physiquement similaires (là où l'ambiguïté est à son maximum) que dissimilaires. Ce dernier résultat suggère que la similarité physique entre les deux scènes d'une image hybride augmenterait le poids de l'interférence des HFS sur la catégorisation des BFS et pourrait donc expliquer pourquoi nous n'avons pas observé le même résultat que Kauffmann et al. (2017) en vision centrale. En conséquence, nous avons réalisé l'**Expérience 2** de cette thèse afin de considérer cette fois-ci des images hybrides physiquement similaires.

- **Différences interhémisphériques**

Dans cette expérience, nous nous sommes également intéressés à la façon dont chaque hémisphère cérébral serait impliqué dans le processus d'intégration des fréquences spatiales. Sur la base des précédents travaux montrant que l'hémisphère droit est plus performant pour catégoriser des scènes en BFS et que l'hémisphère gauche est plus performant pour catégoriser des scènes en HFS (Iidaka et al., 2004 ; Peyrin et al., 2003 ; Peyrin et al., 2004; Peyrin, Chokron et al., 2006 ; Peyrin, Mermillod et al., 2006), nous avons tout d'abord fait l'hypothèse de différences interhémisphériques lors de la catégorisation des scènes cibles en BFS et HFS, indépendamment de la congruence sémantique de la scène distractrice. Ainsi, les performances catégorisation d'une scène cible en BFS devrait être meilleures lorsqu'elle est présentée dans le CVG (dont initialement traitée par l'hémisphère droit) que dans le CVD (dont initialement traitée par l'hémisphère gauche), et inversement pour une scène cible en HFS. De plus, les travaux d'IRMf de Kauffmann, Bourgin et al. (2015) ayant montré que l'hémisphère droit est tout particulièrement engagé dans l'effet d'interférence des BFS sur les HFS, nous nous attendions à ce que l'effet d'interférence des BFS (scène distractrice) sur les HFS (scène cible) soit plus important lorsqu'une image hybride était présentée dans le CVG que dans le CVD.

Les résultats de la **Session 3** ont effectivement montré une interaction entre le champ visuel de présentation des stimuli et le contenu en fréquences spatiales de la scène cible. Cependant, il n'y avait pas de différences significatives entre les champs visuels, que ce soit pour catégoriser les scènes en BFS que pour catégoriser les scènes en HFS. Par contre, les participants étaient plus précis pour catégoriser les scènes cibles en BFS que les scènes cibles en HFS à la fois lorsqu'elles étaient présentées dans le CVD/HG et dans le CVG/HD et l'avantage des BFS étant plus grand dans le CVG/HD que dans le CVD/HG. Ce dernier résultat suggère que l'hémisphère droit serait plus performant que l'hémisphère gauche dans le traitement des BFS par rapport au traitement des HFS et soutient les précédentes études qui ont montré une supériorité de l'hémisphère droit dans le traitement des BFS (Peyrin et al., 2003, 2004 ; Kauffmann, Bourgin et al., 2015). En revanche, nos résultats n'ont ni montré de

supériorité de l'hémisphère gauche lors de la catégorisation de la scène cible en HFS, ni de supériorité de l'hémisphère droit dans l'effet d'interférence des BFS sur les HFS. Alors comment expliquer cette absence de spécialisation hémisphérique ?

Une première interprétation pourrait concerner les différences de latéralisation hémisphérique entre les hommes et les femmes. En effet, de nombreuses études ont montré que la latéralisation hémisphérique des fonctions cognitives (e.g., la parole, l'orientation spatiale, la reconnaissance des visages) était plus prononcée chez les hommes que chez les femmes. Le traitement hémisphérique des femmes aurait quant à lui tendance à être symétrique (Hausmann, Slabberkoorn et al., 2000). Par exemple, Voyer (1996) a comparé 396 asymétries cérébrales observées chez des hommes et des femmes, extraites de 266 études différentes utilisant une présentation visuelle, auditive ou tactile de stimuli verbaux ou non-verbaux. Cette étude comparative a montré que les asymétries cérébrales visuelles et auditives étaient différentes entre les hommes et les femmes dans le sens d'une latéralisation hémisphérique plus prononcée chez les hommes que chez les femmes. Peyrin, Chokron et al. (2006) ont quant à eux montré que la latéralisation hémisphérique du traitement des fréquences spatiales dans des tâches de catégorisation de scènes était plus prononcée chez les hommes que chez les femmes. Dans cette dernière expérience, les participants devaient catégoriser des scènes filtrées en HFS et en BFS présentées dans le CVD et dans le CVG. Les résultats pour les hommes ont montré une spécialisation hémisphérique classique du traitement des fréquences spatiales (i.e. une supériorité du CVD/HG dans le traitement des scènes en HFS et une supériorité du CVG/HD dans le traitement des BFS). En revanche, les résultats pour les femmes n'ont révélé aucune différence entre champs visuels confirmant ainsi que l'organisation cérébrale fonctionnelle des femmes est moins latéralisée que celle des hommes (Voyer, 1996).

Concernant nos résultats, une interprétation en termes de différences hommes-femme semble peu probable car nous n'avons pas observé d'effet du sexe sur les autres variables manipulées. Une interprétation qui nous semble plus plausible est qu'il est difficile d'observer des différences interhémisphériques sur la seule base de données comportementales, comme c'est le cas dans nos expériences. En effet, les différences interhémisphériques correspondent à une réalité neuroanatomique, difficile à identifier à travers des données comportementales issues de la catégorisation d'images hybrides présentées en champs visuels divisés. En ce sens, l'absence de différences statistiques dans nos données comportementales ne signifie pas nécessairement qu'au niveau neuroanatomique, les deux hémisphères soient engagés de la même façon dans le processus d'intégration des fréquences spatiales. Nous pouvons suggérer que c'est davantage notre paradigme expérimental qui ne contraignait pas suffisamment les différences interhémisphériques plutôt que de conclure sur une absence de résultats. Ce point sera abordé dans la Discussion générale de ce Chapitre 2.

- **Conclusion**

Dans l'ensemble, les résultats de cette expérience suggèrent que les BFS peuvent être utilisées par le système visuel pour générer des prédictions et guider une tâche de catégorisation. Conformément à nos hypothèses, la génération de prédictions issue des BFS utilisées pour guider le traitement des HFS serait plus importante en vision périphérique qu'en vision centrale.

Les HFS quant à elles ne pourraient être ignorées du processus de catégorisation en vision centrale, et en particulier en cas d'incertitude ou d'ambiguïté sur le stimulus visuel (cas des images hybrides).

Dans une deuxième expérience (**Expérience 2**), nous avons décidé de tester la catégorisation d'images hybrides physiquement similaires afin d'essayer de répliquer les résultats de l'étude de Kauffmann et al. (2017) non observés dans cette **Expérience 1** : (1) une précedence temporelle des BFS en vision centrale et (2) un effet d'interférence des HFS qui surpasse celui des BFS en vision centrale.

2. Expérience 2 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides physiquement similaires

La méthode expérimentale et les hypothèses étaient identiques à celles de l'**Expérience 1**, mais nous avons utilisé cette fois-ci des images hybrides physiquement similaires.

2.1. Méthode

2.1.1. Participants

Quarante étudiants de l'Université Grenoble Alpes (33 femmes ; âge moyen : $20,69 \pm 1,72$) ont participé à cette expérience. Ils ont été répartis en deux groupes ($n = 20$) qui devaient respectivement ignorer les HFS ou les BFS de l'image hybride. La taille de l'échantillon de chaque groupe a été calculée comme dans l'**Expérience 1**. Nous avons obtenu une taille d'échantillon minimale de 20 participants par groupe expérimental, soit un total de 40 participants. Nous avons exclu des analyses statistiques un participant (groupe Cible HFS) dont les données différaient significativement des autres participants (package *Outliers* sur Rstudio. Les analyses statistiques ont donc été conduites sur 39 participants (33 femmes ; âge moyen : $20,75 \pm 1,74$). Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Comme dans l'**Expérience 1**, tous les participants ont donné leur consentement libre et éclairé et l'expérience a été approuvée par un comité d'éthique (CERGA).

2.1.2. Stimuli et procédure

L'expérience était en tout point identique à celle de l'**Expérience 1**, à l'exception que nous avons substitué les images hybrides physiquement similaires aux images hybrides physiquement dissimilaires (Figure 2.8). Les deux scènes pouvaient être soit sémantiquement congruentes, soit sémantiquement incongruente. Nous avons utilisé au total 80 images hybrides

physiquement similaires : 40 images hybrides sémantiquement congruentes (20 paires de scènes superposant deux scènes artificielles et 20 images hybrides superposant deux scènes naturelles ; corrélation IP : 0.75 ± 0.08 ; corrélation SA : 0.67 ± 0.03) et 40 images hybrides sémantiquement incongruentes (superposant une scène artificielle et une scène artificielle ; corrélation IP : 0.73 ± 0.06 ; corrélation SA : 0.66 ± 0.03). Comme pour l'**Expérience 1**, nous avons utilisé les 160 scènes filtrées qui composent les images hybrides physiquement similaires (80 scènes en BFS et 80 scènes en HFS).

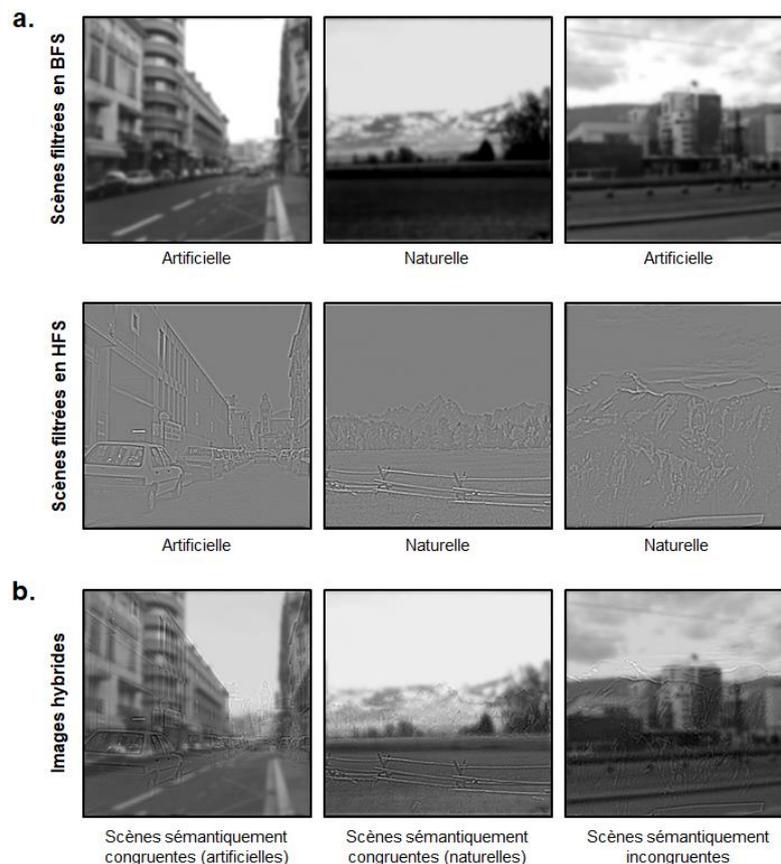


Figure 2.8 : Exemple de stimuli utilisés dans l'**Expérience 2** : (a) Scènes filtrées en BFS et en HFS ; (b) Images hybrides créées en superposant une scène filtrée en BFS et une autre scène filtrée en HFS. Les deux scènes sont physiquement similaires. Elles pouvaient être soit sémantiquement congruentes (e.g., une scène naturelle en BFS superposée à une autre scène naturelle en HFS), soit sémantiquement incongruentes (e.g., une scène naturelle en BFS superposée à une scène artificielle en HFS).

Tous les participants réalisaient trois sessions expérimentales. Dans la **Session 1**, nous avons présenté en vision centrale les scènes en BFS et HFS qui composaient les images hybrides. Nous les avons présentées individuellement en vision centrale. Les participants devaient catégoriser les scènes comme appartenant soit à la catégorie 'naturelle', soit à la catégorie 'artificielle'. Dans la **Session 2**, nous avons utilisé les images hybrides que nous avons présentées en vision centrale. Les participants étaient répartis en deux groupes ($n = 20$ dans le

premier groupe et $n = 19$ dans le deuxième groupe). Le premier groupe devait catégoriser la scène en BFS tout en ignorant la scène en HFS (condition Cible BFS/Distracteur HFS), et le deuxième groupe devait catégoriser la scène en HFS tout en ignorant la scène en BFS (condition Cible HFS/Distracteur BFS). Dans la **Session 3**, nous avons à nouveau utilisé les images hybrides que nous avons présentées en vision périphérique, soit dans le champ visuel droit (CVD), soit dans le champ visuel gauche (CVG). Les participants qui devaient catégoriser la scène en BFS dans la Session 2 catégorisaient à nouveau la scène en BFS et les participants qui devaient catégoriser la scène en HFS dans la Session 2 catégorisaient à nouveau la scène en HFS.

2.2. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale de chaque session, nous avons calculé les pourcentages d'erreurs moyen (%ERm) et les temps de réponses correctes moyens en millisecondes (TRm). Les temps de réponses supérieurs ou inférieurs à trois fois l'intervalle interquartile ont été supprimés (soit 1,69% des réponses pour la Session 1, 0,78% pour la Session 2 et 1,03% pour la Session 3). Le seuil de significativité a été fixé à 0,05. La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2). Afin de faciliter la lecture de cette partie, nous avons consigné les résultats relatifs à nos hypothèses dans le Tableau [2.1](#).

2.2.1. Session 1 : Catégorisation de scènes filtrées en BFS et en HFS en vision centrale

Pour l'analyse de la **Session 1**, un test t de Student pour échantillons appariés a été réalisé sur les valeurs de %ERm et les TRm (Tableau [2.5](#)) avec comme variable intra-sujets la Fréquence spatiale de la scène à catégoriser (BFS ; HFS).

Tableau 2.5 : Résultats de la Session 1 de l'**Expérience 2**. Pourcentages d'erreurs moyens (%ERm), temps de réponse correcte moyens (en millisecondes) et erreurs-types en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS).

BFS		HFS	
%ERm	TRm	%ERm	TRm
3,72% \pm 0,45	648 \pm 13	3,62% \pm 0,41	655 \pm 12

Quelle que soit la condition expérimentale, les taux d'erreurs étaient bas (inférieurs à 4%). Contrairement à notre hypothèse d'un avantage du traitement des BFS sur le traitement des HFS, la différence de performances de catégorisation entre les BFS et les HFS n'était pas significative, ni pour les %ERm ($t(37) = 0,18, p = .860$; Figure [2.9a](#)), ni pour les TRm ($t(37) = -1,64, p = .110$; Figure [2.9b](#)).

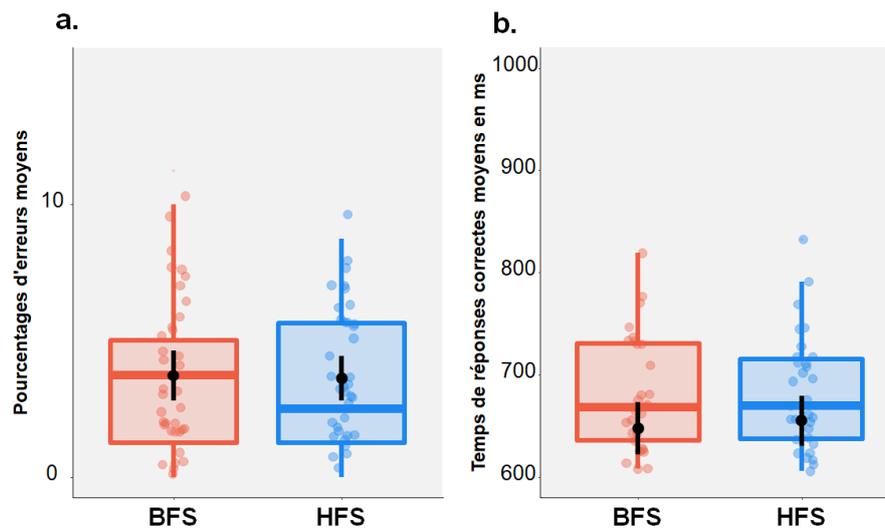


Figure 2.9 : Résultats de la Session 1 de l'**Expérience 2** (catégorisation de scènes filtrées en BFS et en HFS en vision centrale) : **(a) Pourcentages d'erreurs moyens** et **(b) Temps de réponses correctes moyens en ms** en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

2.2.2. Sessions 2 et 3 : Catégorisation d'images hybrides en vision centrale et en vision périphérique

Dans cette partie, nous avons réuni les résultats des **Sessions 2 et 3** dans la même analyse afin de comparer directement les effets obtenus en vision centrale et en vision périphérique. Ainsi, une ANOVA de type $2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm (Tableau 2.6) avec comme facteur intra-sujet la Congruence sémantique (Congruent ; Incongruent) et la Position spatiale des images hybrides dans le champ visuel (Centre ; Périphérie) et comme facteur inter-sujet la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS).

Tableau 2.6 : Résultats des Sessions 2 et 3 de l'**Expérience 2**. Pourcentages d'erreurs moyens, temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et de la Position spatiale des images hybrides dans le champ visuel (Centre ; Périphérie).

		Centre		Périphérie	
		%ERm	TRm	%ERm	TRm
Cible BFS/Distracteur HFS	Congruent	5,44 ± 2,02	718 ± 26	8,69 ± 2,35	693 ± 23
	Incongruent	38,38 ± 2,54	812 ± 27	31,75 ± 2,20	758 ± 28
Cible HFS/Distracteur BFS	Congruent	13,03 ± 2,07	786 ± 27	21,97 ± 2,41	770 ± 23
	Incongruent	28,03 ± 2,60	837 ± 28	45,46 ± 2,26	805 ± 29

- **Analyse des pourcentages d'erreurs moyens**

Concernant tout d'abord notre hypothèse d'un avantage du traitement des BFS lors de la catégorisation de scène, l'ANOVA menée sur les %ERm a révélé un effet principal de la Fréquence spatiale Cible/Distracteur était significatif ($F(1,37) = 9,89, p = .003, \eta^2 = .21$). Les participants faisaient moins d'erreurs pour catégoriser les scènes cibles en BFS ($21,06 \pm 1,34\%$) par rapport aux scènes cibles en HFS ($27,12 \pm 1,38\%$). La Fréquence spatiale Cible/Distracteur interagissait avec la Position spatiale des images hybrides ($F(1,37) = 21,73, p < .001, \eta^2 = .37$). Comme pour l'**Expérience 1**, les comparaisons planifiées ont montré que, quel que soit la condition de congruence sémantique, les participants faisaient en moyenne moins d'erreurs pour catégoriser les scènes cibles en BFS que les scènes cibles en HFS lorsqu'elles étaient présentées en vision périphérique ($F(1,37) = 33,12, p < .001$; BFS : $20,22 \pm 2,32\%$; HFS : $33,72 \pm 2,38\%$), mais pas lorsqu'elles étaient présentées en vision centrale (BFS : $21,90 \pm 2,61\%$; HFS : $20,53 \pm 2,68\%$; $F(1,37) < 1, p = .605$).

Concernant ensuite nos hypothèses d'interférence sémantique, l'ANOVA a tout d'abord révélé un effet principal de la Congruence sémantique ($F(1,37) = 167,55, p < .001, \eta^2 = .82$). Les participants faisaient plus d'erreurs pour catégoriser des images hybrides sémantiquement incongruentes ($35,90 \pm 2,07\%$) que sémantiquement congruentes ($12,28 \pm 1,65\%$). La Fréquence spatiale Cible/Distracteur interagissait avec la Congruence sémantique ($F(1,37) = 5,76, p = .022, \eta^2 = .13$). Les comparaisons planifiées ont montré que l'effet d'interférence était significatif à la fois pour la condition Cible BFS/Distracteur HFS (Congruent : $7,06 \pm 2,30\%$; Incongruent : $35,01 \pm 2,20\%$; $F(1,37) = 120,80, p < .001$) et pour la condition Cible HFS/Distracteur BFS (Congruent : $17,50 \pm 2,36\%$; Incongruent : $36,74 \pm 2,98\%$; $F(1,37) = 54,21, p < .001$). Toutefois, l'effet d'interférence était plus grand dans la condition Cible BFS/Distracteur HFS que dans la condition Cible HFS/Distracteur BFS. Comme attendu, nous avons observé que ces effets d'interférence étaient modulés par la position des images hybrides dans le champ visuel (interaction double entre la Congruence sémantique, la Fréquence spatiale Cible/Distracteur et la Position spatiale des images hybrides, $F(1,37) = 18,53, p < .001, \eta^2 = .33$; Figure 2.10). Les comparaisons planifiées ont révélé une interférence des BFS sur les HFS aussi bien en vision centrale (Congruent : $13,03 \pm 2,07\%$; Incongruent : $28,03 \pm 2,60\%$; $F(1,37) = 28,92, p < .001$) qu'en vision périphérique (Congruent : $21,97 \pm 2,41\%$; Incongruent : $45,46 \pm 2,26\%$; $F(1,37) = 52,29, p < .001$). L'interaction significative entre la Congruence sémantique et la position spatiale pour la condition Cible HFS/Distracteur BFS ($F(1,37) = 7,72, p = .008$) suggère, comme pour l'**Expérience 1**, que l'effet d'interférence sémantique des BFS sur les HFS était plus fort en vision périphérique qu'en vision centrale. De plus, les comparaisons planifiées ont aussi révélé une interférence des HFS sur les BFS à la fois en vision centrale (Congruent : $5,44 \pm 2,02\%$; Incongruent : $38,38 \pm 2,54\%$; $F(1,37) = 146,78, p < .001$) qu'en vision périphérique (Congruent : $8,69 \pm 2,35\%$; Incongruent : $31,75 \pm 2,20\%$; $F(1,37) = 53,07, p < .001$). L'interaction significative entre la Congruence sémantique et la position spatiale pour la condition Cible BFS/Distracteur HFS ($F(1,37) = 11,00, p = .002$) suggère que l'effet d'interférence sémantique des HFS sur les BFS était plus fort en vision centrale qu'en vision périphérique.

Enfin, en vision centrale, l'interaction Congruence sémantique × Fréquence spatiale Cible/Distracteur était significative ($F(1,37) = 21,21, p < .001$). Plus précisément, lorsque les images hybrides étaient présentées en vision centrale, l'effet d'interférence sémantique était significatif à la fois pour la condition cible en BFS/Distracteur en HFS ($F(1,37) = 146,78, p < .001$) et pour la condition cible en HFS/Distracteur en BFS ($F(1,37) = 28,92, p < .001$). Toutefois, l'interaction significative suggère que l'effet d'interférence sémantique était plus grand lorsque le distracteur était en HFS (Congruent : $5,44 \pm 2,02\%$; Incongruent : $38,38 \pm 2,95\%$) qu'en BFS (Congruent : $13,03 \pm 2,07\%$; Incongruent : $28,03 \pm 2,60\%$). En revanche, en vision périphérique, l'interaction Congruence sémantique × Fréquence spatiale Cible/Distracteur n'était pas significative ($F(1,37) < 1, p = .926$), indiquant que l'interférence de la scène distractive en BFS n'était pas plus grande que l'interférence de la scène en HFS.

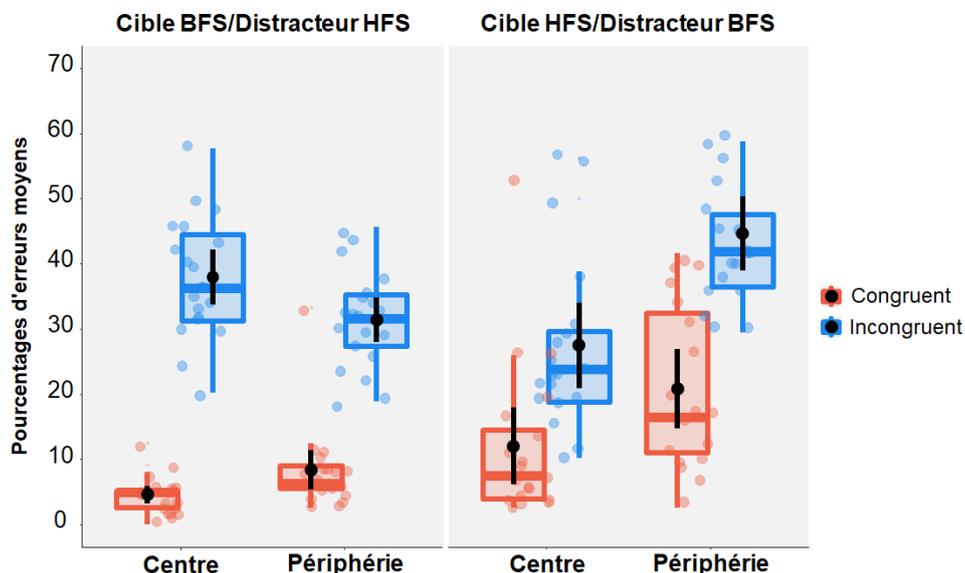


Figure 2.10 : Résultats des Sessions 2 et 3 de l'**Expérience 2** (catégorisation d'images hybrides en vision centrale et en vision périphérique) : **Pourcentages d'erreurs moyennes** en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et de la Position spatiale des images hybrides dans le champ visuel (Centre ; Périphérie). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

- **Analyse des temps de réponses correctes**

Concernant tout d'abord notre hypothèse d'un avantage du traitement des BFS lors de la catégorisation de scène, l'ANOVA menée sur les TRm n'a pas montré ni d'effet principal de la Fréquence spatiale Cible/Distracteur ($F(1,37) = 2,41, p = .061$), ni d'interaction Fréquence spatiale Cible/Distracteur × Champ visuel ($F(1,37) < 1, p = .463$).

Concernant ensuite nos hypothèses d'interférence sémantique, l'effet principal de la Congruence sémantique était significatif ($F(1,37) = 113,39, p < .001, \eta^2 = .75$; Figure 2.11).

Les participants étaient plus rapides pour catégoriser des images hybrides congruentes (742 ± 23 ms) qu'incongruentes (803 ± 27 ms). La Fréquence spatiale Cible/Distracteur interagissait avec la Congruence sémantique ($F(1,37) = 9,86, p = .003$). Les comparaisons planifiées ont montré que l'effet d'interférence était significatif à la fois pour la condition Cible BFS/Distracteur HFS (Congruent : 706 ± 32 ms ; Incongruent : 785 ± 37 ms ; $F(1,37) = 97,56, p < .001$) et pour la condition Cible HFS/Distracteur BFS (Congruent : 778 ± 33 ms ; Incongruent : 821 ± 38 ; $F(1,37) = 27,49, p < .001$). Toutefois, l'interaction suggère que l'effet d'interférence était plus grand dans la condition Cible BFS/Distracteur HFS que dans la condition Cible HFS/Distracteur BFS. En revanche, nous n'avons pas observé d'interaction double Congruence sémantique \times Fréquence spatiale Cible/Distracteur \times Position spatiale ($F(1,37) < 1, p = .608$).

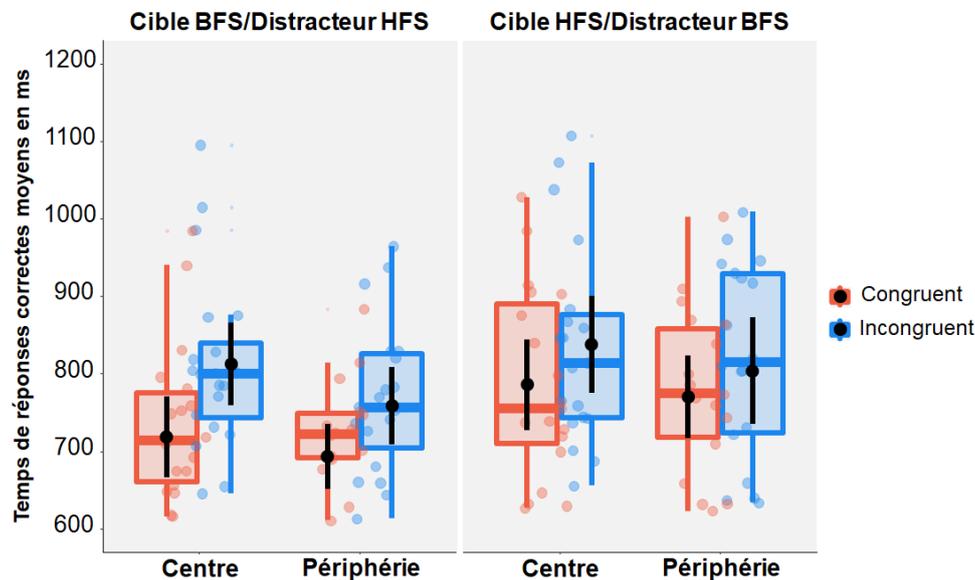


Figure 2.11 : Résultats des Sessions 2 et 3 de l'Expérience 2 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : **Temps de réponses correctes moyennes en ms** en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et de la Position spatiale des images hybrides dans le champ visuel (Centre ; Périphérie). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

2.2.3. Session 3 : Catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche

Dans cette partie, nous avons analysé uniquement les résultats de la Session 3 afin de comparer directement les effets obtenus entre le champ visuel droit (CVD) et le champ visuel gauche (CVG). Cette analyse nous permettait de répondre à notre dernière hypothèse sur les

différences interhémisphériques dans le processus d'intégration des fréquences spatiales². Ainsi, une ANOVA de type $2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm (Tableau 2.7) avec comme facteur intra-sujet la Congruence sémantique (Congruent ; Incongruent) et le Champ visuel/Hémisphère de présentation des images hybrides (CVD/HG ; CVG/HD) et comme facteur inter-sujet la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS).

Tableau 2.7 : Résultats de la Session 3 de l'Expérience 2. Pourcentages d'erreurs moyens, temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et du Champ visuel de présentation de stimuli (CVD ; CVG).

		CVD		CVG	
		%ERm	TRm	%ERm	TRm
Cible BFS/Distracteur HFS	Congruent	9,37 ± 2,46	694 ± 24	8,00 ± 2,41	691 ± 22
	Incongruent	33,13 ± 2,25	754 ± 29	30,38 ± 2,49	762 ± 28
Cible HFS/Distracteur BFS	Congruent	20,99 ± 2,53	773 ± 24	22,96 ± 2,47	767 ± 23
	Incongruent	45,53 ± 2,31	798 ± 30	45,39 ± 2,56	812 ± 29

L'ANOVA menée sur les %ERm et sur les TRm n'a pas révélé d'effet principal du Champ visuel de présentation des images hybrides (%ERm : $F(1,37) < 1$; TRm : $F(1,37) < 1$; Figure 2.12), ni d'interaction Champ visuel × Fréquence spatiale Cible/Distracteur (%ERm : $F(1,37) = 2,90$, $p = .097$; TRm : $F(1,37) < 1$, $p = .904$), ni d'interaction Champ visuel × Fréquence spatiale Cible/Distracteur × Congruence sémantique (%ERm : $F(1,37) < 1$, $p = .792$; TRm : $F(1,36) < 1$, $p = .621$).

² Comme dans l'Expérience 1, nous avons mené une analyse en incluant la variable contrôle Miroir (i.e. chaque image hybride était présentée une fois dans une version originale et une fois dans une version retournée en miroir) afin de nous assurer que les performances obtenues dans la condition Miroir soient équivalentes à celles obtenues dans la condition Originale quel que soit le champ visuel de présentation des stimuli (CVD ; CVG). L'ANOVA menée sur les %ERm et sur les TRm n'a révélé ni d'interaction Miroir × Champ visuel (%ERm : $F(1,37) = 3,04$, $p = .090$; TRm : $F(1,37) < 1$), ni d'interaction Miroir × Champ visuel × Fréquence spatiale Cible/Distracteur (%ERm : $F(1,37) < 1$; TRm : $F(1,37) < 1$), ni d'interaction Miroir × Champ visuel × Fréquence spatiale Cible/Distracteur × Congruence sémantique (%ERm : $F(1,37) < 1$; TRm : $F(1,36) < 1$). Cette variable n'a donc pas été considérée dans la suite des analyses. Dans l'Expérience 1, nous avons testé l'effet de la variable Sexe (Femme ; Homme) afin de nous assurer que cette variable n'influence pas les différences de performance entre les champs visuels de présentation des stimuli (CVD ; CVG). L'Expérience 2 a été menée auprès de 40 participants dont 33 étaient des femmes (soit plus de 80% de l'échantillon), ce qui ne nous permet pas ici de tester l'effet de la variable Sexe (Femme, Homme). En effet, notre recrutement se fait au sein du Département de Psychologie de l'Université Grenoble Alpes où les étudiants sont majoritairement des femmes. Toutefois, les résultats de l'Expérience 1 n'ayant pas révélé d'effet du sexe sur les autres variables manipulées, nous ne pensons pas que ce biais dans notre échantillon soit susceptible d'impacter nos résultats en termes de différences interhémisphériques.

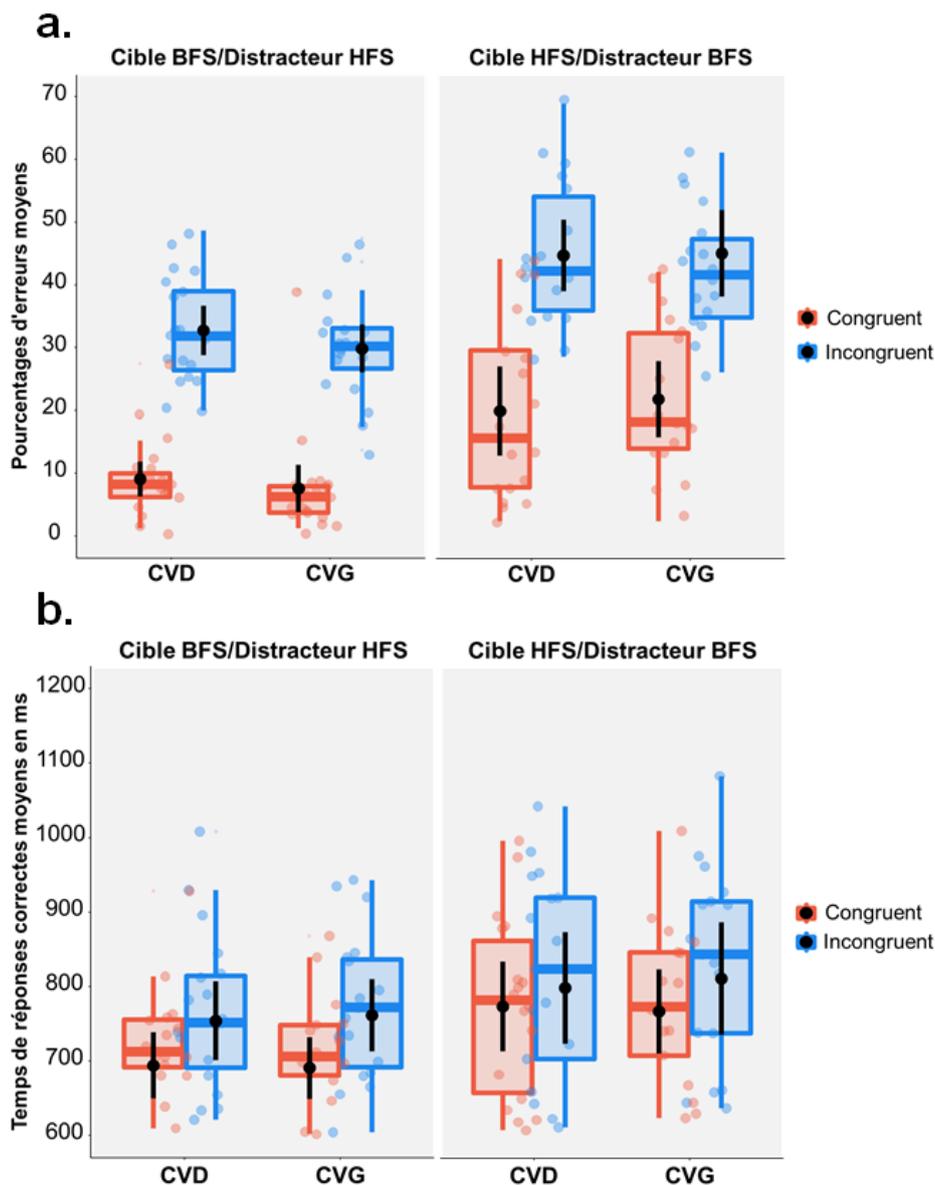


Figure 2.12 : Résultats de la Session 3 de l'**Expérience 2** (catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche) : **(a) Pourcentage d'erreurs moyens** et **(b) Temps de réponses correctes moyens en ms** en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent) et du Champ visuel de présentation (CVD ; CVG). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

2.3. Discussion

Nous avons mené cette deuxième expérience afin de comprendre pourquoi les résultats de l'**Expérience 1** ne répliquaient par l'ensemble des résultats de Kauffmann et al. (2017). En effet, dans leur expérience menée en vision centrale, Kauffmann et al. (2017) observaient un

effet d'interférence plus important des HFS sur les BFS que l'inverse. Dans notre **Expérience 1** au contraire, nos résultats ont montré qu'en vision centrale, l'effet d'interférence des BFS sur les HFS était aussi fort que celui des HFS sur les BFS. Étant donné que Kauffmann et al. (2017) utilisaient des images hybrides à la fois physiquement dissimilaires et physiquement similaires dans leur expérience, nous nous sommes demandés si la similarité physique d'une image hybride pouvait augmenter le poids de l'interférence des HFS sur les BFS en vision centrale. Pour cela, nous avons mené l'**Expérience 2** dans laquelle nous avons cette fois-ci utilisé uniquement des images hybrides physiquement similaires précédemment utilisées par Kauffmann et al. (2017). Pour rappel, une image hybride est physiquement similaire lorsque les deux scènes en BFS et HFS se superposent spatialement et que les spectres d'amplitude sont très proches.

- **Avantage des BFS sur les HFS lors de la catégorisation de scènes**

Comme dans l'**Expérience 1**, les résultats de la **Session 1** ont montré que les participants étaient capables de catégoriser très précisément les scènes filtrées en BFS et en HFS. A nouveau, nous n'avons pas observé de différences significatives entre la catégorisation des BFS et des HFS, mais il y avait à nouveau un effet plafond (pourcentages d'erreurs inférieurs à 4%). De la même façon, les résultats de la **Session 2** (catégorisation d'images hybrides en vision centrale) n'ont pas révélé de différences significatives entre la catégorisation des scènes cibles en BFS et celle des scènes cibles en HFS, Bien que nous ayons utilisé des stimuli plus ambigus (des images hybrides physiquement similaires). Pris ensemble, les résultats des Sessions 1 et 2 montrent une nouvelle fois que la catégorisation en vision centrale est aussi efficace lorsqu'elle est basée sur les BFS que sur les HFS.

Enfin, les résultats de la **Session 3** (catégorisation d'images hybrides en vision périphérique) ont montré que les participants étaient plus précis pour catégoriser les scènes cibles en BFS par rapport aux scènes cibles en HFS. Ce résultat réplique celui observé dans la Session 3 de l'**Expérience 1**. Comme nous l'avons mentionné précédemment, étant donné que les BFS sont majoritairement extraites de la vision périphérique, il n'est pas surprenant d'observer un avantage des BFS par rapport aux HFS en vision périphérique.

- **Effets d'interférence sémantique des BFS et des HFS lors de la catégorisation de scènes hybrides**

Les résultats des **Sessions 2 et 3** (catégorisation d'images hybrides en vision centrale et en vision périphérique) ont montré comme attendu que la catégorisation de la scène cible était améliorée lorsque les deux scènes d'une image hybride étaient sémantiquement congruentes qu'incongruentes. Cet effet d'interférence sémantique indique que l'information sémantique contenue dans la scène distractive était quand même traitée, et perturbait la catégorisation de la scène cible, quel que soit son contenu en fréquences spatiales et sa position spatiale dans le champ visuel. Ce résultat avait déjà été observé dans l'**Expérience 1**.

Par ailleurs, les résultats des **Sessions 2 et 3** ont révélé que cet effet d'interférence sémantique était modulé par la position spatiale des images hybrides lorsque la scène cible était

en HFS et la scène distractive en BFS. Conformément à notre hypothèse et aux résultats de l'**Expérience 1**, la scène distractive en BFS interférait plus fortement sur la catégorisation de la scène cible en HFS en vision périphérique qu'en vision centrale. Ce résultat suggère une nouvelle fois que le système visuel utiliserait très rapidement les BFS disponibles principalement en vision périphérique afin de générer des prédictions qui viendraient influencer le traitement des informations plus lentes en HFS. Autrement dit, lorsque la catégorisation d'une image hybride s'effectue en vision périphérique, le système visuel augmenterait le poids de l'interférence des prédictions issues de l'analyse en BFS (dans la tâche de catégorisation d'une scène cible en HFS) et réduirait le poids d'une analyse détaillée portée par les HFS.

Néanmoins, cette interprétation peut être remise en question. En effet, étant donné que les signaux de HFS sont majoritairement issus de la vision centrale, il est également possible que ce résultat traduise le fait que la perception des HFS se complexifie à mesure qu'on s'éloigne de la vision centrale et donc, que les participants perçoivent très mal la scène en HFS de l'image hybride présentées en vision périphérique. Cette interprétation semble toutefois peu probable dans notre **Expérience 2** du fait de l'interférence des HFS sur la catégorisation de scènes en BFS même en vision périphérique, qui suggère que les HFS sont perçues même en vision périphérique. De plus, nos stimuli (dont les bords internes et externes des stimuli étaient présentés respectivement à 2° et 8° d'excentricité rétinienne à partir du point de fixation central) étaient en réalité présentés relativement proche des limites de la vision centrale. En effet, comme nous l'avons vu dans le **Chapitre 1** de cette thèse, on distingue la vision centrale, comprenant la fovéa (au centre de la rétine) qui s'étend approximativement à 1° d'excentricité. La partie restante du champ visuel correspond à la parafovéa, qui s'étend jusqu'à environ 4-5° d'excentricité rétinienne et la vision périphérique qui couvre la surface restante du champ visuel. Une grande partie de nos images hybrides étant donc présentées en parafovéa, où les HFS peuvent être aisément perçues. Néanmoins, bien qu'une interprétation en termes de difficultés de perception des HFS dans les images hybrides périphériques soit peu probable, nous avons décidé de conduire une nouvelle expérience (**Expérience 3**) afin de tester explicitement la catégorisation des scènes filtrées en BFS et en HFS en vision périphérique.

Par ailleurs, sur la base des résultats de Kauffmann et al. (2017) obtenus en vision centrale uniquement, nous nous attendions à ce qu'une scène distractive en HFS interfère également sur la catégorisation d'une scène cible en BFS et à ce que cet effet devrait être plus grand que celui des BFS sur les HFS. De plus, étant donné la perte de résolution spatiale en vision périphérique, nous avons fait l'hypothèse que la scène distractive en HFS interfère moins fortement en vision périphérique qu'en vision centrale. Conformément à notre hypothèse, nos résultats ont montré que l'interférence de la scène distractive en HFS sur la catégorisation d'une scène en BFS était plus forte en vision centrale qu'en vision périphérique. Etant donné que les signaux en HFS proviennent principalement de la vision centrale, le système visuel pourrait augmenter le poids d'une analyse détaillée portée par les HFS dans cette partie du champ visuel, ce qui expliquerait que l'interférence des HFS sur la catégorisation des scènes en BFS soit plus importante en vision centrale qu'en vision périphérique. D'autre part, nos résultats ont montré que l'interférence de la scène distractive en HFS en vision centrale était plus forte que celle de la scène distractive en BFS en vision centrale. Ce dernier résultat, conforme à l'étude de Kauffmann et al. (2017), n'avait pas été observé dans l'**Expérience 1** où

nous avons présenté uniquement des images hybrides physiquement dissimilaires. Il semblerait donc que la similarité physique entre les deux scènes d'une image hybride influence les effets d'interférence sémantique. Cependant, nous n'avons pas pu tester directement la façon dont la similarité physique interagit avec les autres variables car nous avons manipulé les deux modalités de cette variable dans des expériences différentes (les **Expériences 1 et 2**). Nous aurions pu faire des analyses statistiques a posteriori mais nous nous serions alors mis volontairement dans des conditions peu favorables pour tester nos hypothèses. Ainsi, dans une troisième expérience (**Expérience 3**), nous avons manipulé explicitement à la fois la similarité et la dissimilarité physique entre les deux scènes d'une image hybride dans le même design expérimental.

- **Différences interhémisphériques**

Dans cette expérience, nous nous sommes à nouveau intéressés à la façon dont chaque hémisphère cérébral serait impliqué dans le processus d'intégration des fréquences spatiales. Les résultats de la **Session 3** n'ont montré aucune différence interhémisphérique significative, que ce soit concernant la catégorisation de la scène en BFS ou HFS que concernant les effets d'interférence.

Comment pouvons-nous expliquer ici l'absence de dominance de l'hémisphère droit lors de la catégorisation des scènes en BFS observée dans l'**Expérience 1** ? Une possible explication est que l'utilisation d'images hybrides physiquement similaires présentées en vision périphérique rendent la tâche trop complexe pour faire émerger des différences hémisphériques. Les taux d'erreurs étaient parfois très élevés en vision périphérique, surtout lorsque les images hybrides étaient sémantiquement incongruentes (taux d'erreur supérieurs à 30% dans la condition Cible BFS/Distracteur HFS et taux d'erreurs supérieurs à 40% dans la condition Cible HFS/Distracteur BFS). Dans ce cas, les participants pourraient mettre en place des mécanismes compensatoires afin de pallier la difficulté de la tâche en recrutant des régions supplémentaires dans l'autre hémisphère cérébral. Ce point sera détaillé dans la discussion générale de ce chapitre.

- **Conclusion**

Dans l'ensemble, les résultats de cette expérience suggèrent une nouvelle fois que les BFS peuvent être utilisées par le système visuel pour générer des prédictions et guider la catégorisation de scènes visuelles. Conformément à nos hypothèses, la génération de prédictions issue des BFS utilisées pour guider le traitement des HFS serait plus importante en vision périphérique qu'en vision centrale. Les HFS, quant à elles, ne pourraient être ignorées du processus de catégorisation en vision centrale, et en particulier en cas d'incertitude ou d'ambiguïté sur le stimulus visuel (cas des images hybrides). Ce processus dépendrait du conflit entre les informations physiques propres à chaque scène. En effet, nous avons observé que la dissimilarité physique (**Expérience 1**) et la similarité physique (**Expérience 2**) amenait à des résultats sensiblement différents. Plus précisément, l'interférence des HFS sur les BFS en vision

centrale était renforcée dans le cas où les deux scènes d'une image hybride étaient physiquement similaires que dissimilaires et donc quand les stimuli étaient plus ambigus.

Enfin, ces deux expériences nous ont permis de soulever plusieurs biais méthodologiques. Par exemple, nous demandions toujours aux participants de réaliser la tâche la catégorisation des scènes hybrides en vision centrale (**Session 2**) avant celle de catégorisation des scènes hybrides en vision périphérique (**Session 3**). Nous avons fait ce choix méthodologique afin de familiariser les participants aux images hybrides en vision centrale avant de les confronter à une catégorisation bien plus difficile en vision périphérique. Il est alors possible que des effets de familiarité aient biaisé nos comparaisons statistiques entre les résultats obtenus en vision centrale et ceux obtenus en vision périphérique. Par ailleurs, nous n'avons pas testé la catégorisation des scènes filtrées en BFS et en HFS en vision périphérique. C'est pour ces différentes raisons que nous avons mené une troisième expérience (**Expérience 3**) afin de tester dans un même design expérimental l'influence de la similarité physique des deux scènes d'une image hybride sur les effets d'interférence sémantique, mais aussi afin de pallier les biais méthodologiques rencontrés dans les **Expériences 1 et 2**.

3. Expérience 3 - Test de l'influence de la similarité physique (scènes hybrides physiquement similaire et dissimilaires)

L'objectif théorique de l'**Expérience 3** était le même que les deux premières expériences (i.e. tester l'effet d'interférence sémantique des BFS sur les HFS en vision périphérique et en vision centrale respectivement), mais cette nouvelle expérience avait pour but de tester directement l'influence de la similarité physique des deux scènes d'une image hybride sur les effets d'interférence sémantique, ainsi que de pallier les biais méthodologiques rencontrés dans les **Expériences 1 et 2**. La méthode expérimentale était identique à celle des deux précédentes expériences, mais nous avons utilisé cette fois-ci dans la même expérience des images hybrides physiquement similaires et physiquement dissimilaires.

Nous avons décidé de manipuler la variable similarité physique en intra-sujet afin de nous mettre dans des conditions méthodologiques similaires à celles de Kauffmann et al. (2017). Ainsi, dans cette **Expérience 3**, les images hybrides composées de scènes physiquement similaires et de scènes physiquement dissimilaires étaient présentées de façon aléatoire dans une même session expérimentale et pour un même participant. Par ailleurs, afin de limiter les effets d'ordre entre la passation en vision centrale vs. périphérique, les scènes hybrides étaient désormais présentées aléatoirement dans une même et seule session expérimentale en vision centrale et en vision périphérique.

Enfin, dans les **Expériences 1 et 2**, nous avons confirmé que l'interférence des BFS sur la catégorisation de scènes en HFS était plus importante en vision périphérique qu'en vision centrale. Nous avons interprété ce résultat dans le contexte du codage prédictif, à savoir que la vision périphérique serait particulièrement utile pour la génération de prédictions sur la base de BFS. Cependant, il est également possible que ce résultat traduise le fait que la perception des HFS se complexifie à mesure qu'on s'éloigne de la vision centrale et donc, que les participants

perçoivent très mal la composante en HFS de images hybrides présentées en vision périphérique. Nous rappelons ici que cette interprétation semble toutefois peu probable après avoir observé dans l'**Expérience 2** une interférence des HFS sur la catégorisation de scènes en BFS même en vision périphérique, suggérant que les HFS restent bien perçues et utilisées en vision périphérique. Néanmoins, nous avons décidé d'inclure dans cette expérience la catégorisation de scènes en HFS en vision périphérique. Nous avons donc présenté aléatoirement dans une même et seule session expérimentale des scènes filtrées en BFS et HFS en vision centrale et en vision périphérique.

Ainsi, dans cette **Expérience 3**, les participants réalisaient deux sessions expérimentales. Dans une première session, nous avons testé la catégorisation des scènes filtrées en BFS et HFS en fonction du champ visuel (central versus périphérique) afin de nous assurer que la perception des scènes en HFS reste possible en vision périphérique. Dans une deuxième session expérimentale, nous avons testé les effets d'interférence sémantique en fonction de la similarité physique et du champ visuel et, cette fois-ci, les images hybrides étaient présentées aléatoirement en vision centrale ou en vision périphérique dans la même session expérimentale.

Les hypothèses étaient identiques à celles des **Expériences 1 et 2**. Cependant, dans cette expérience, nous voulions tester directement l'influence de la similarité physique des deux scènes d'une image hybride sur les effets d'interférence. Ainsi, deux scènes sémantiquement congruentes pouvaient être soit physiquement similaires, soit physiquement dissimilaires et deux scènes sémantiquement incongruentes pouvaient être soit physiquement similaires, soit physiquement dissimilaires (Figure 2.13). Nous avons vu à travers les **Expériences 1 et 2** que la similarité physique entre les deux scènes d'une image hybride amenait à des résultats sensiblement différents. Par ailleurs, les résultats de Mu et Li (2013) en vision centrale ont montré que l'effet d'interférence d'une scène en BFS sur la catégorisation d'une scène en HFS était plus important (concernant uniquement les pourcentages d'erreurs moyens) lorsque les deux scènes étaient physiquement similaires que dissimilaires. Les résultats de Kauffmann et al. (2017) en vision centrale ont montré quant à eux un effet d'interférence des HFS sur la catégorisation des BFS plus important lorsque les deux scènes dans l'image hybride étaient physiquement similaires que dissimilaires. Sur la base de ces précédentes études, nous nous attendions à ce que les effets d'interférence sémantique soient globalement plus importants dans la condition physiquement similaire que dissimilaire (interaction Congruence sémantique \times Similarité physique). Pour aller plus loin, sur la base des résultats de Kauffmann et al. (2017), nous nous attendions à ce que l'effet d'interférence sémantique des HFS (scène distractive) observé en vision centrale soit plus important lorsque les deux scènes d'une image hybride sont physiquement similaires que dissimilaires (interaction Congruence sémantique \times Similarité physique en vision centrale pour la condition Cible BFS/Distracteur HFS). A l'inverse, on pouvait s'attendre à ce que l'effet d'interférence sémantique des BFS (scène distractive) observé en vision périphérique soit plus important lorsque les deux scènes d'une image hybride sont physiquement similaires que dissimilaires (interaction Congruence sémantique \times Similarité physique en vision périphérique pour la condition Cible HFS/Distracteur BFS).

3.1. Méthode

3.1.1. Participants

Quarante étudiants de l'Université Grenoble Alpes (35 femmes ; âge moyen : $20,30 \pm 2,27$) ont participé à cette expérience. Ils ont été répartis en deux groupes ($n = 20$) qui devaient respectivement ignorer les HFS ou les BFS de l'image hybride. La taille de l'échantillon de chaque groupe a été calculée comme dans l'**Expérience 1**. Nous avons obtenu une taille d'échantillon minimale de 20 participants par groupe expérimental, soit un total de 40 participants. Nous avons exclu des analyses un participant (groupe Cible HFS) dont les données différaient des autres participants (package *Outliers* sur Rstudio). Les analyses statistiques ont donc été conduites sur 39 participants (34 femmes ; âge moyen : $20,35 \pm 2,33$). Comme dans les précédentes expériences, tous les participants ont donné leur consentement libre et éclairé et l'expérience a été approuvée par un comité d'éthique (CERGA).

3.1.2. Stimuli et procédure

Dans cette expérience, nous avons cette fois utilisé les images hybrides physiquement similaires et dissimilaires (Figure [2.13](#)) issues de l'étude de Kauffmann et al. (2017). Les deux scènes pouvaient être soit sémantiquement congruentes, soit sémantiquement incongruentes. Au total, nous avons utilisé 160 images hybrides : 80 images hybrides sémantiquement congruentes (dont 40 images hybrides physiquement similaires et 40 images hybrides physiquement dissimilaires) et 80 images hybrides sémantiquement incongruentes (dont 40 images hybrides physiquement similaires et 40 images hybrides physiquement dissimilaires). Afin de réduire le temps d'expérimentation considérablement rallongé par l'ajout de nouvelles conditions expérimentales, nous avons décidé de ne pas présenter chaque image hybride dans une version retournée en miroir (i.e. retournée sur l'axe vertical) car aucun effet de la variable contrôle Miroir n'a été observé dans les **Expériences 1** et **2**. Ainsi, chaque image hybride était présentée une seule fois dans sa version originale. Dans la présente **Expérience 3**, nous avons utilisé les 160 scènes filtrées qui composent les images hybrides (80 scènes en BFS et 80 scènes en HFS) dans la Session 1 et les 160 images hybrides dans la Session 2.

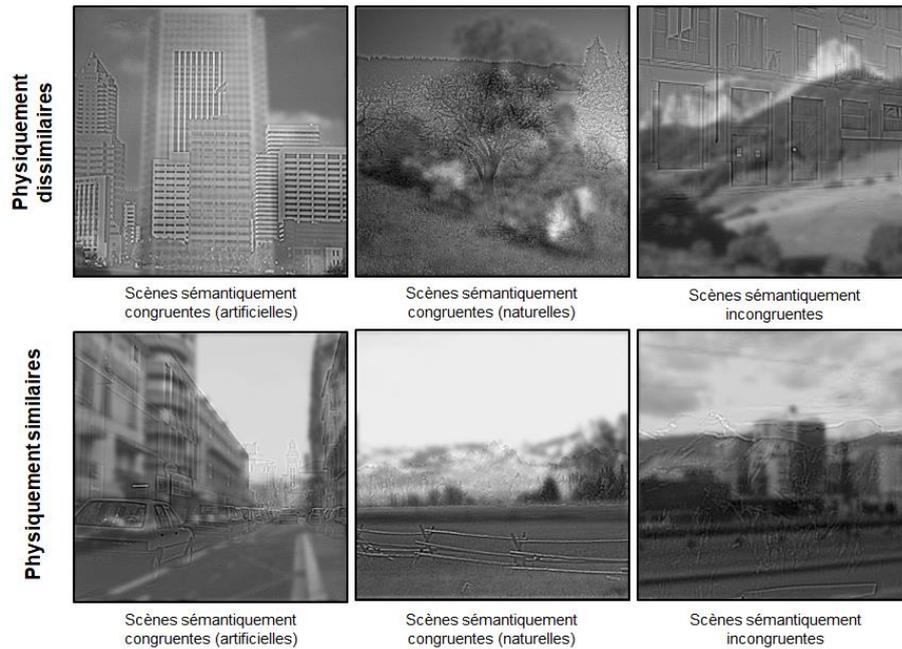


Figure 2.13 : Exemple d'images hybrides utilisées dans l'**Expérience 3**. Les images hybrides étaient créées en superposant une scène filtrée en BFS et une autre scène filtrée en HFS. Les deux scènes formant une image hybride pouvaient soit être sémantiquement congruentes (e.g., une scène naturelle en BFS superposée à une autre scène naturelle en HFS), soit être sémantiquement incongruentes (e.g., une scène naturelle en BFS superposée à une scène artificielle en HFS). De plus, les images hybrides sémantiquement congruentes et incongruentes pouvaient être constituées de deux scènes physiquement similaires ou de deux scènes physiquement dissimilaires.

Tous les participants réalisaient deux sessions expérimentales. Dans la **Session 1**, nous avons utilisé les scènes en BFS et HFS qui composaient les images hybrides. Nous les avons présentées individuellement cette fois-ci soit en vision centrale, soit en vision périphérique (dans le CVD, ou dans le CVG). Cette session expérimentale avait pour but de nous assurer que les scènes en HFS étaient correctement catégorisées en vision périphérique, mais aussi de comparer le traitement des BFS et HFS en vision centrale et en vision périphérique. Dans la **Session 2**, nous avons utilisé les images hybrides que nous avons présentées soit en vision centrale, soit en vision périphérique (dans le CVD ou dans le CVG). Les participants étaient répartis en deux groupes ($n = 20$ dans le premier groupe et $n = 19$ dans deuxième groupe). Le premier groupe devait catégoriser la scène en BFS tout en ignorant la scène en HFS (condition Cible BFS/Distracteur HFS), et le deuxième groupe devait catégoriser la scène en HFS tout en ignorant la scène en BFS (condition Cible HFS/Distracteur BFS). Le reste des conditions de passation expérimentale était identique aux deux précédentes expériences.

3.2. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale de chaque session, les pourcentages d'erreurs moyens (%ERm) et les temps de réponses correctes moyens en

millisecondes (TRm) ont été calculés. Les temps de réponses supérieurs ou inférieurs à trois fois l'intervalle interquartile ont été supprimés (soit 1,26% des résultats pour la Session 1 et 1,27% pour la Session 2). Le seuil de significativité a été fixé à 0,05. La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2). Afin de faciliter la lecture de cette partie, nous avons consigné les résultats relatifs à nos hypothèses dans le Tableau 2.1.

3.2.1. Session 1 : Catégorisation de scènes filtrées en BFS et en HFS

a. Catégorisation de scènes filtrées en BFS et en HFS en vision centrale et en vision périphérique

Dans cette analyse, nous avons comparé les effets obtenus en vision centrale et en vision périphérique. Ainsi, une ANOVA de type 2×2 a été réalisée sur les valeurs de %ERm et les TRm (Tableau 2.8) avec comme facteur intra-sujet la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et le Position spatiale de la scène dans le champ visuel (Centre ; Périphérie).

Tableau 2.8 : Résultats de la Session 1 de l'Expérience 3. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et de la Position de la scène dans le champ visuel (Centre ; Périphérie).

	Centre		Périphérie	
	%ERm	TRm	%ERm	TRm
BFS	4,39 ± 0,68	633 ± 14	6,25 ± 0,65	661 ± 14
HFS	4,17 ± 0,53	643 ± 13	5,80 ± 0,57	673 ± 14

Nos résultats indiquent tout d'abord que la catégorisation de scènes en HFS est possible en vision périphérique (taux d'erreurs inférieurs à 7%). L'ANOVA menée sur les %ERm et sur les TRm a révélé un effet principal de la Fréquence spatiale pour les TRm ($F(1,38) = 12,80$, $p < .001$, $\eta^2 = .25$; Figure 2.14b), mais pas pour les %ERm (BFS : $5,32 \pm 0,80\%$; HFS : $4,98 \pm 0,70\%$; $F(1,38) < 1$, $p < .400$). Les participants étaient plus rapides pour catégoriser des scènes en BFS (647 ± 19 ms) qu'en HFS (658 ± 18 ms). L'effet principal de la Position spatiale était significatif (%ERm : $F(1,38) = 16,13$, $p < .001$, $\eta^2 = .29$; TRm : $F(1,38) = 12,80$, $p < .001$, $\eta^2 = .25$). Les participants faisaient en moyenne moins d'erreurs et étaient plus rapides en vision centrale (%ERm : $4,28 \pm 0,76\%$; TRm : 638 ± 18 ms) qu'en vision périphérique (%ERm : $6,02 \pm 0,77\%$; TRm : 667 ± 19 ms). Par contre, ce facteur n'interagissait pas avec la Fréquence spatiale (%ERm : $F(1,38) < 1$; TRm : $F(1,38) < 1$).

Ces résultats indiquent donc que les scènes en BFS étaient catégorisées plus rapidement que les scènes en HFS quelle que soit la position de la scène dans le champ visuel.

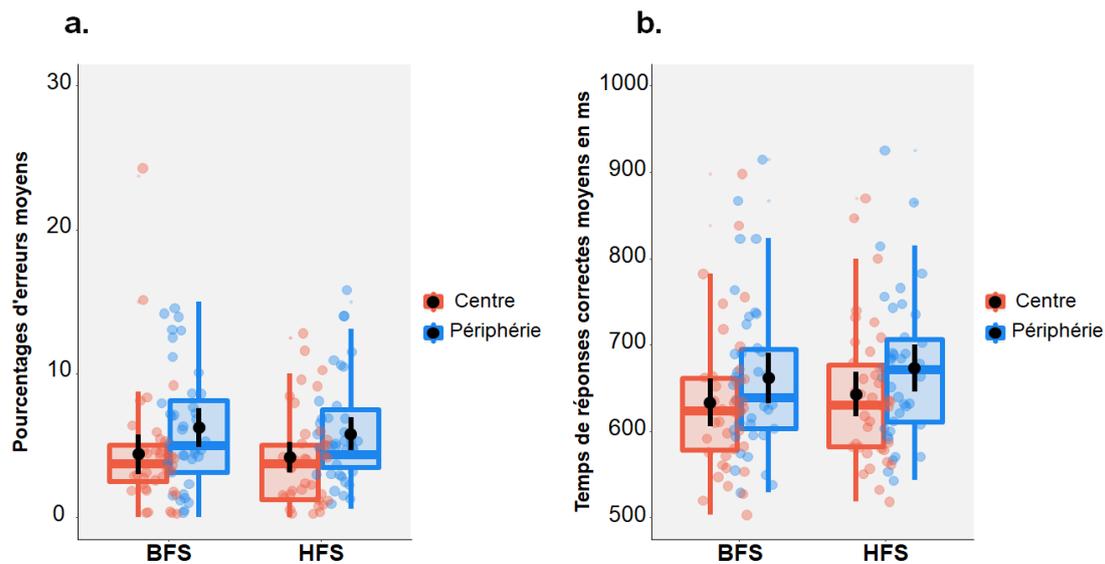


Figure 2.14 : Résultats de la Session 1 de l'**Expérience 3** (catégorisation de scènes filtrées en BFS et en HFS en vision centrale et en vision périphérique) : **(a) Pourcentages d'erreurs moyens** et **(b) temps de réponses correctes moyens en ms** en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et de la Position spatiale de la scène dans le champ visuel (Centre ; Périphérie). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

b. Catégorisation de scènes filtrées en BFS et en HFS dans le champ visuel droit et dans le champ visuel gauche

Dans cette analyse, nous avons comparé les effets obtenus dans le champ visuel droit (CVD) et dans le champ visuel gauche (CVG). Ainsi, une ANOVA de type 2×2 a été réalisée sur les valeurs de %ERm et les TRm (Tableau 2.9) avec comme facteur intra-sujet la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et le Champ visuel de présentation de la scène (CVD ; CVG).

Tableau 2.9 : Résultats de la Session 1 de l'**Expérience 3**. Pourcentages d'erreurs moyens, temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et du Champ visuel de présentation des stimuli (CVD ; CVG).

	CVD		CVG	
	%ERm	TRm	%ERm	TRm
BFS	6,60 ± 0,80	661 ± 15	5,90 ± 0,69	661 ± 14
HFS	5,96 ± 0,70	674 ± 14	5,64 ± 0,58	672 ± 14

L'ANOVA menée sur les %ERm et sur les TRm a montré un effet principal significatif de la Fréquence spatiale pour les TRm ($F(1,38) = 7,142, p < .011, \eta^2 = .16$; Figure 2.15), mais pas pour les %ERm ($F(1,38) < 1, p < .422$). Les participants étaient plus rapides pour catégoriser

des scènes en BFS (661 ± 20 ms) qu'en HFS (673 ± 19 ms). L'effet principal du Champ visuel de présentation n'était pas significatif ni pour les %ERm ($F(1,38) = 1,12, p < .296$), ni pour les TRm ($F(1,38) < 1, p < .810$), et ce facteur n'interagissait pas avec la Fréquence spatiale, ni pour les %ERm ($F(1,38) < 1, p < .671$), ni pour les TRm ($F(1,38) < 1, p < .788$).

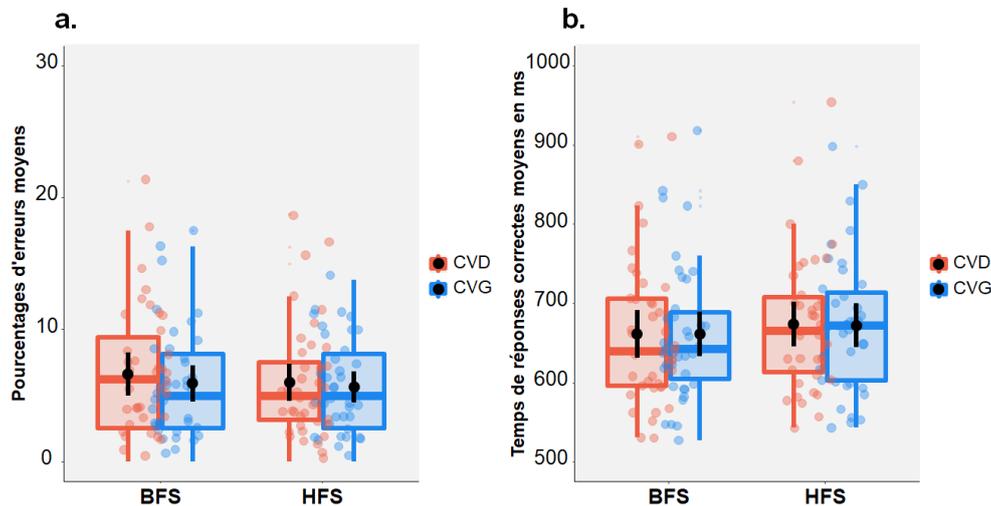


Figure 2.15 : Résultats de la Session 1 de l'**Expérience 3** (catégorisation de scènes filtrées en BFS et en HFS dans le champ visuel droit et dans le champ visuel gauche) : **(a) Pourcentages d'erreurs moyens** et **(b) temps de réponses correctes moyens en ms** en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et du Champ visuel de présentation (CVD ; CVG). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

3.2.2. Session 2 : Catégorisation d'images hybrides

a. Catégorisation d'images hybrides en vision centrale et en vision périphérique

Dans cette analyse, nous avons comparé les effets obtenus en vision centrale et en vision périphérique. Ainsi, une ANOVA de type $2 \times 2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm (Tableau 2.10) avec comme facteur intra-sujet la Congruence sémantique (Congruent ; Incongruent), la Position spatiale de l'image hybride dans le champ visuel (Centre ; Périphérie) et la Similarité physique (PHYSIM ; PHYDIS), et comme facteur inter-sujet la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS).

Tableau 2.10 : Résultats de la Session 2 de l'**Expérience 3**. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction du contenu en Fréquences spatiales de la cible et du distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent), de la Similarité physique (PHYSIM ; PHYDIS) et de la Position spatiale de l'image hybride dans le champ visuel (Centre ; Périphérie).

		Centre		Périphérie		
		%ERm	TRm	%ERm	TRm	
Cible BFS/Distracteur HFS	Congruent	PHYSIM	9,75 ± 2,06	692 ± 27	11,56 ± 1,91	702 ± 25
		PHYDIS	9,25 ± 1,52	687 ± 25	10,62 ± 1,71	697 ± 26
	Incongruent	PHYSIM	38,75 ± 2,4	788 ± 27	33,94 ± 2,22	751 ± 27
		PHYDIS	25,25 ± 3,09	750 ± 24	21,81 ± 2,62	729 ± 26
Cible HFS/Distracteur BFS	Congruent	PHYSIM	14,74 ± 2,11	843 ± 27	24,01 ± 1,96	857 ± 26
		PHYDIS	9,74 ± 1,56	789 ± 25	18,09 ± 1,75	845 ± 26
	Incongruent	PHYSIM	25,39 ± 2,47	849 ± 28	42,10 ± 2,27	887 ± 28
		PHYDIS	18,16 ± 3,18	822 ± 25	37,43 ± 2,69	869 ± 27

- Analyse des pourcentages d'erreurs moyens

Concernant tout d'abord notre hypothèse d'un avantage du traitement des BFS lors de la catégorisation de scène, l'ANOVA menée sur les %ERm n'a pas révélé d'effet principal de la Fréquence spatiale Cible/Distracteur (BFS : 20,11 ± 1,72% ; HFS : 23,71 ± 1,77% ; $F(1,37) = 2,11$, $p < .155$), mais ce facteur interagissait avec la Position spatiale des images hybrides ($F(1,37) = 129,00$, $p < .001$, $\eta^2 = .78$). Les comparaisons planifiées ont montré que, quel que soit la condition de congruence sémantique, les participants faisaient en moyenne moins d'erreurs pour catégoriser les scènes cibles en BFS que les scènes cibles en HFS lorsqu'elles étaient présentées en vision périphérique (BFS : 19,48 ± 3,48% ; HFS : 30,41 ± 3,57% ; $F(1,37) = 19,20$, $p < .001$), mais pas lorsqu'elles étaient présentées en vision centrale (BFS : 20,75 ± 3,65% ; HFS : 17,01 ± 3,74% ; $F(1,37) = 2,05$, $p = .160$).

Concernant ensuite nos hypothèses d'interférence sémantique, l'ANOVA a bien révélé un effet principal de la Congruence sémantique ($F(1,37) = 134,31$, $p < .001$, $\eta^2 = .78$; Figure 2.16). Les participants faisaient en moyenne plus d'erreurs pour catégoriser des images hybrides incongruentes (30,35 ± 3,38%) que congruentes (13,47 ± 2,23%). La Fréquence spatiale Cible/Distracteur n'interagissait pas avec la Congruence sémantique ($F(1,37) = 3,58$, $p = .066$), suggérant un effet d'interférence quel que soit le contenu en fréquences spatiales de la scène à ignorer. Cependant, comme attendu, nous avons observé que ces effets d'interférence étaient modulés par la position des images hybrides dans le champ visuel (interaction Congruence sémantique × Fréquence spatiale Cible/Distracteur × Position spatiale, $F(1,37) = 32,64$, $p < .001$, $\eta^2 = .47$). Les comparaisons planifiées ont révélé une interférence des BFS sur les HFS aussi bien en vision centrale (Congruent : 12,24 ± 2,35% ; Incongruent : 21,78 ± 3,74% ; $F(1,37) = 16,61$, $p < .001$) qu'en vision périphérique (Congruent : 21,05 ± 2,24% ; Incongruent : 39,67 ± 3,41% ; $F(1,37) = 70,43$, $p < .001$). L'interaction significative entre la Congruence sémantique et la Position spatiale pour la condition Cible HFS/Distracteur BFS ($F(1,37) = 24,16$, $p < .001$) suggère, comme pour les **Expériences 1** et **2**, que l'effet d'interférence sémantique des BFS sur

les HFS était plus fort en vision périphérique qu'en vision centrale. Par ailleurs, les comparaisons planifiées ont aussi révélé une interférence des HFS sur les BFS aussi bien en vision centrale (Congruent : $9,50 \pm 2,29\%$; Incongruent : $32,00 \pm 3,64\%$; $F(1,37) = 97,29$, $p < .001$) qu'en vision périphérique (Congruent : $11,09 \pm 2,40\%$; Incongruent : $27,88 \pm 3,33\%$; $F(1,37) = 59,60$, $p < .001$). L'interaction significative entre la Congruence sémantique et la Position spatiale pour la condition Cible BFS/Distracteur HFS ($F(1,37) = 9,87$, $p = .003$) suggère cette fois-ci que l'effet d'interférence sémantique des HFS sur les BFS était plus fort en vision centrale qu'en vision périphérique.

En vision centrale, l'interaction Congruence sémantique \times Fréquence spatiale Cible/Distracteur était significative ($F(1,37) = 15,73$, $p < .001$). Plus précisément, lorsque les images hybrides étaient présentées en vision centrale, l'effet d'interférence sémantique était significatif à la fois pour la condition cible en BFS/Distracteur en HFS ($F(1,37) = 97,29$, $p < .001$) et pour la condition cible en HFS/Distracteur en BFS ($F(1,37) = 16,61$, $p < .001$). Toutefois, l'effet d'interférence sémantique était plus fort lorsque le distracteur était en HFS (pour les scènes cibles en BFS) qu'en BFS (pour les scènes cibles en HFS). En revanche, en vision périphérique, l'interaction Congruence sémantique \times Fréquence spatiale Cible/Distracteur n'était pas significative ($F(1,37) < 1$, $p = .538$), indiquant que l'interférence de la scène distractive en BFS n'était pas plus forte que l'interférence de la scène en HFS en vision périphérique.

Les résultats concernant l'effet de la similarité physique ont d'abord montré une interaction significative entre la Congruence sémantique et la Similarité physique ($F(1,37) = 22,45$, $p < .001$). Les comparaisons planifiées ont montré un effet d'interférence sémantique à la fois pour la condition PHYSIM ($F(1,37) = 216,12$, $p < .001$) et pour la condition PHYDIS ($F(1,37) = 57,69$, $p < .001$). Comme attendu, cette différence était plus grande pour la condition PHYSIM (Congruent : $15,01 \pm 1,87\%$; Incongruent : $35,05 \pm 2,20\%$) que pour la condition PHYDIS (Congruent : $11,93 \pm 1,43\%$; Incongruent : $25,66 \pm 2,74\%$). Par ailleurs, l'interaction double Congruence sémantique \times Fréquence spatiale Cible/Distracteur \times Similarité physique était significative ($F(1,37) = 19,07$, $p < .001$). Spécifiquement, l'interaction entre la Congruence sémantique et la Similarité physique était significative pour la condition Cible BFS/Distracteur HFS ($F(1,37) = 42,54$, $p < .001$), mais pas pour la condition Cible HFS/Distracteur BFS ($F(1,37) < 1$, $p = .797$). Les comparaisons planifiées ont montré que lorsque la scène à ignorer était en HFS (catégorisation de la scène cible en BFS), l'effet d'interférence sémantique était significatif à la fois pour la condition physiquement similaire ($F(1,37) = 182,38$, $p < .001$) et pour la condition physiquement dissimilaire ($F(1,37) = 28,99$, $p < .001$). Néanmoins, l'effet d'interférence était plus fort lorsque les deux scènes d'une images hybrides étaient physiquement similaires (Congruent : $10,65 \pm 2,60\%$; Incongruent : $36,34 \pm 3,08\%$) que dissimilaires (Congruent : $9,94 \pm 2,00\%$; Incongruent : $23,53 \pm 3,83\%$). En d'autres termes, indépendamment du champ visuel de présentation des stimuli, l'interférence des HFS sur la catégorisation des BFS était plus importante lorsque les deux scènes de l'image hybride étaient physiquement similaires que dissimilaires. Par contre, lorsque la scène à ignorer était en BFS (catégorisation de la scène cible était en HFS), l'interférence des BFS sur les HFS était aussi importante lorsque les deux scènes d'une image hybride étaient physiquement similaires que physiquement dissimilaires.

Enfin, l'interaction Congruence sémantique \times Position spatiale \times Similarité physique n'était pas significative, ni pour la condition Cible HFS/Distracteur BFS ($F(1,37) = 1,33, p = .256$), ni pour la condition Cible BFS/Distracteur HFS ($F(1,37) < 1, p = .542$).

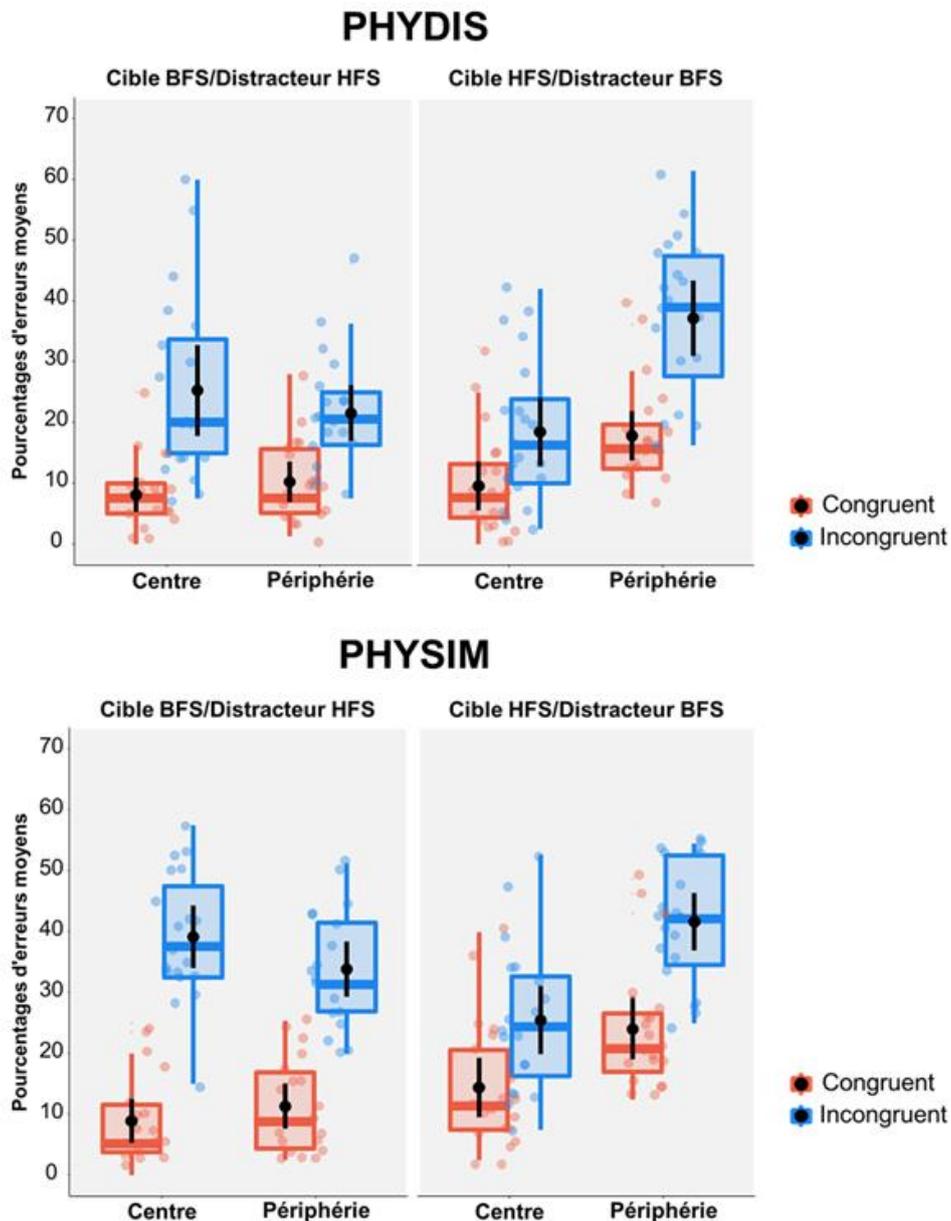


Figure 2.16 : Résultats des Sessions 2 de l'Expérience 3 (catégorisation d'images hybrides en vision centrale et en vision périphérique) : **Pourcentage d'erreurs moyens** en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Position spatiale des images hybrides dans le champ visuel (Centre ; Périphérie), de la Congruence sémantique (Congruent ; Incongruent) et de la Similarité physique (PHYDIS ; PHYSIM). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

- **Analyse des temps de réponses correctes**

Concernant tout d'abord notre hypothèse d'un avantage du traitement des BFS, l'ANOVA menée sur les TRm a révélé un effet principal de la Fréquence spatiale Cible/Distracteur ($F(1,37) = 11,93, p < .001$), à savoir que les participants étaient plus rapides pour catégoriser les scènes cibles en BFS (724 ± 24 ms) que pour catégoriser les scènes cibles en HFS (845 ± 25 ms). La Fréquence spatiale Cible/Distracteur interagissait avec la Position spatiale des images hybrides dans le champ visuel ($F(1,37) = 12,33, p = .001, \eta^2 = .25$). Les comparaisons planifiées ont montré que, quel que soit la condition de congruence sémantique, les participants étaient plus rapides pour catégoriser les scènes cibles en BFS que les scènes cibles en HFS à la fois lorsqu'elles étaient présentées en vision périphérique ($F(1,37) = 16,01, p < .001$) et en vision centrale ($F(1,37) = 7,61, p = .009$). Toutefois, cette différence était plus grande en vision périphérique (BFS : 720 ± 51 ms ; HFS : 865 ± 52 ms) qu'en vision centrale (BFS : 729 ± 49 ms ; HFS : 826 ± 50 ms). Ce résultat indique un avantage des BFS par rapport aux HFS à la fois en vision périphérique et en vision centrale.

Concernant ensuite nos hypothèses d'interférence sémantique, l'effet principal de la Congruence sémantique était significatif ($F(1,37) = 102,385, p < .001, \eta^2 = .73$; Figure 2.17). Les participants étaient plus rapides pour catégoriser des images hybrides congruentes (764 ± 35 ms) qu'incongruentes (806 ± 35 ms). Ce facteur interagissait avec la Fréquence spatiale Cible/Distracteur ($F(1,37) = 19,63, p < .001, \eta^2 = .35$). Les comparaisons planifiées ont montré que l'effet d'interférence était significatif à la fois pour la condition Cible BFS/Distracteur HFS (Congruent : 695 ± 49 ms ; Incongruent : 754 ± 49 ms ; $F(1,37) = 108,63, p < .001$) et pour la condition Cible HFS/Distracteur BFS (Congruent : 833 ± 50 ms ; Incongruent : 857 ± 50 ; $F(1,37) = 15,77, p < .001$). Toutefois, l'effet d'interférence était plus grand dans la condition Cible BFS/Distracteur HFS que dans la condition cible HFS/Distracteur BFS. Comme attendu, ces effets d'interférence étaient modulés par la position des images hybrides dans le champ visuel (interaction Congruence sémantique \times Fréquence spatiale Cible/Distracteur \times Position spatiale, $F(1,37) = 10,20, p = .003, \eta^2 = .22$). Les comparaisons planifiées ont révélé une interférence des BFS sur les HFS aussi bien en vision centrale (Congruent : 816 ± 36 ms ; Incongruent : 835 ± 36 ms ; $F(1,37) = 4,89, p = .033$) qu'en vision périphérique (Congruent : 851 ± 36 ms ; Incongruent : 878 ± 38 ms ; $F(1,37) = 15,44, p < .001$). L'interaction non significative entre la Congruence sémantique et la Position spatiale pour la condition Cible HFS/Distracteur BFS ($F(1,37) < 1, p = .454$) suggère que l'effet d'interférence sémantique des BFS sur les HFS était aussi fort en vision périphérique qu'en vision centrale. Par ailleurs, les comparaisons planifiées ont révélé une interférence des HFS sur les BFS aussi bien en vision centrale (Congruent : 690 ± 35 ms ; Incongruent : 769 ± 35 ms ; $F(1,37) = 86,42, p < .001$) qu'en vision périphérique (Congruent : 700 ± 35 ms ; Incongruent : 740 ± 37 ms ; $F(1,37) = 35,05, p < .001$). L'interaction significative entre la Congruence sémantique et le Champ visuel pour la condition Cible BFS/Distracteur HFS ($F(1,37) = 7,72, p = .008$) suggère que l'effet d'interférence sémantique des HFS sur les BFS était plus fort en vision centrale qu'en vision périphérique. Enfin, nous avons étudié les effets d'interférence séparément dans chaque partie du champ visuel. En vision centrale, l'interaction Congruence sémantique \times Fréquence spatiale Cible/Distracteur était significative ($F(1,37) = 24,06, p < .001$). Plus précisément, lorsque les

images hybrides étaient présentées en vision centrale, l'effet d'interférence sémantique était significatif à la fois pour la condition cible en BFS/Distracteur en HFS ($F(1,37) = 86,42, p < .001$) et pour la condition cible en HFS/Distracteur en BFS ($F(1,37) = 4,89, p = .033$). Toutefois, l'effet d'interférence sémantique était plus fort lorsque le distracteur était en HFS qu'en BFS. En vision périphérique, l'interaction Congruence sémantique \times Fréquences spatiales n'était pas significative ($F(1,37) = 1,74, p = .196$), indiquant que l'interférence de la scène distractrice en BFS n'était pas plus forte que l'interférence de la scène en HFS.

Les résultats concernant l'effet de la similarité physique n'ont pas montré d'interaction significative Congruence sémantique \times Similarité physique ($F(1,37) < 1, p = .334$). Par contre, l'interaction Congruence sémantique \times Fréquence spatiale Cible/Distracteur \times Similarité physique était significative ($F(1,37) = 5,34, p = .026$). Spécifiquement, l'interaction entre la Congruence sémantique et la Similarité physique était significative pour la condition Cible BFS/Distracteur HFS ($F(1,37) = 5,55, p = .024$), mais pas pour la condition Cible HFS/Distracteur BFS ($F(1,37) < 1$). Lorsque la scène à ignorer était en HFS (catégorisation de la scène cible en BFS), l'effet d'interférence sémantique était significatif à la fois pour la condition physiquement similaire ($F(1,37) = 102,57, p < .001$) et pour la condition physiquement dissimilaire ($F(1,37) = 30,71, p < .001$). Cet effet d'interférence sémantique était plus fort lorsque les deux scènes d'une images hybrides étaient physiquement similaires (Congruent : $10,65 \pm 2,60\%$; Incongruent : $36,34 \pm 3,08\%$) que dissimilaires (Congruent : $9,94 \pm 2,00\%$; Incongruent : $23,53 \pm 3,83\%$). Par contre, lorsque la scène à ignorer était en BFS (catégorisation de la scène cible était en HFS), l'interférence des BFS sur les HFS était aussi importante lorsque les deux scènes d'une image hybride étaient physiquement similaires que physiquement dissimilaires. Ces résultats répliquent ceux obtenus pour les %ERm.

Enfin, l'interaction Congruence sémantique \times Position spatiale \times Similarité physique n'était pas significative, ni pour la condition Cible HFS/Distracteur BFS ($F(1,37) = 2,81, p = .102$), ni pour la condition Cible BFS/Distracteur HFS ($F(1,37) < 1, p = .404$).

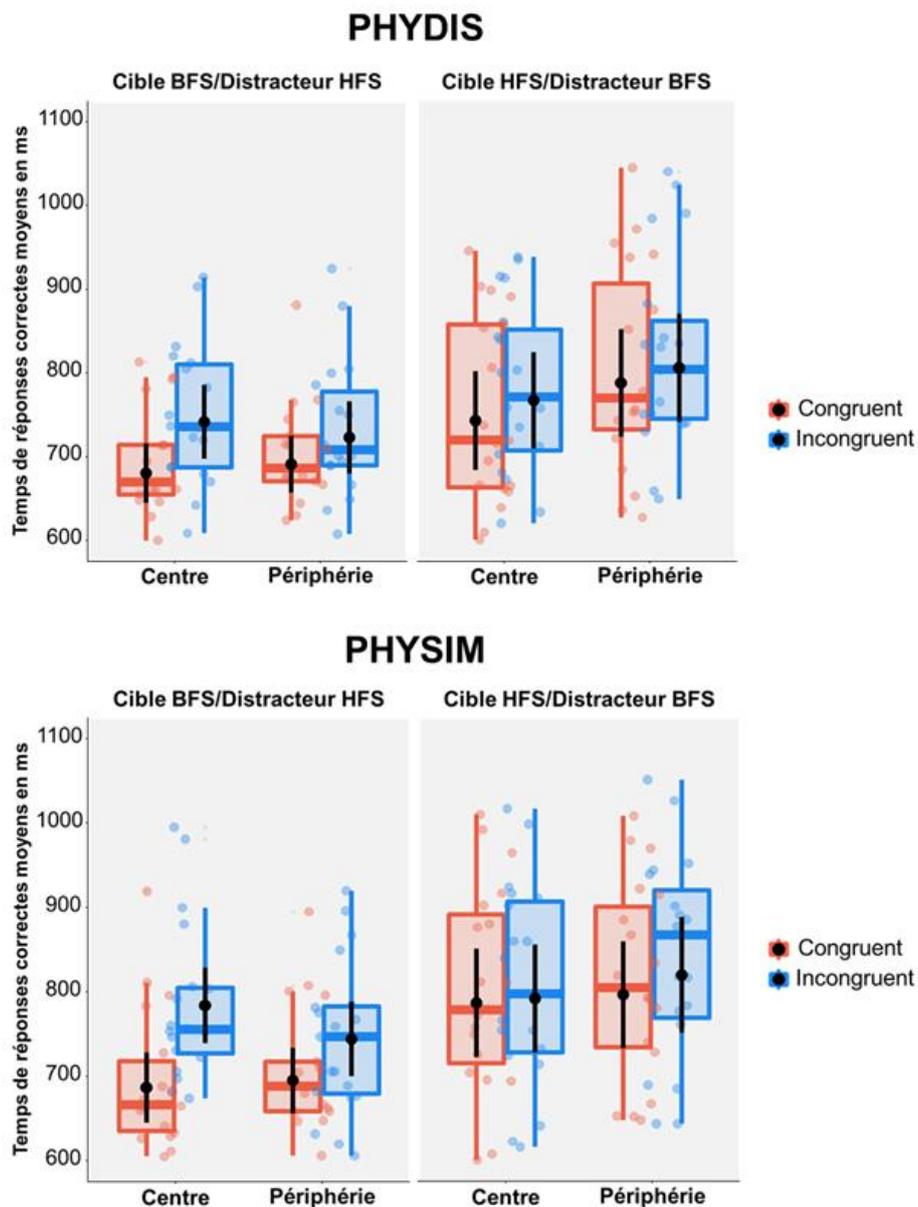


Figure 2.17 : Résultats des Sessions 2 de l'[Expérience 3](#) (catégorisation d'images hybrides en vision centrale et en vision périphérique) : **Temps de réponses correctes moyennes en ms** en fonction de la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent), du Champ visuel de présentation (Centre ; Périphérie) et de la Similarité physique (PHYDIS ; PHYSIM). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

b. Catégorisation d'images hybrides dans le champ visuel droit et dans le champ visuel gauche

Dans cette partie, nous avons comparé directement les effets obtenus entre le champ visuel droit (CVD) et le champ visuel gauche (CVG). Cette analyse nous permettait de répondre

à notre dernière hypothèse sur les différences interhémisphériques dans le processus d'intégration des fréquences spatiales et le traitement visuel prédictif³. Ainsi, une ANOVA de type $2 \times 2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm (Tableau 2.11) avec comme facteur intra-sujet la Congruence sémantique (Congruent ; Incongruent), la Similarité physique (PHYSIM ; PHYDIS), et le Champ visuel de présentation des stimuli/hémisphère (CVD/HG ; CVG/HD) et comme facteur inter-sujet la Fréquence spatiale Cible/Distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS).

Tableau 2.11 : Résultats de la Session 2 de l'Expérience 3. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction du contenu en Fréquences spatiales de la cible et du distracteur (Cible BFS/Distracteur HFS ; Cible HFS/Distracteur BFS), de la Congruence sémantique (Congruent ; Incongruent), de la Similarité physique (PHYSIM ; PHYDIS) et du Champ visuel de présentation de stimuli (CVD ; CVG).

			CVD		CVG	
			%ERm	TRm	%ERm	TRm
Cible BFS/Distracteur HFS	Congruent	PHYSIM	10,38 ± 2,09	703 ± 27	12,75 ± 2,13	701 ± 26
		PHYDIS	10,25 ± 2,01	698 ± 29	11,00 ± 1,76	697 ± 25
	Incongruent	PHYSIM	35,13 ± 2,64	753 ± 30	32,75 ± 2,42	749 ± 26
		PHYDIS	24,00 ± 2,77	729 ± 27	19,63 ± 2,82	728 ± 28
Cible HFS/Distracteur BFS	Congruent	PHYSIM	24,34 ± 2,15	868 ± 28	23,68 ± 2,18	846 ± 27
		PHYDIS	18,42 ± 2,06	843 ± 29	17,76 ± 1,81	846 ± 25
	Incongruent	PHYSIM	42,24 ± 2,70	884 ± 31	41,97 ± 2,49	890 ± 26
		PHYDIS	40,00 ± 2,85	853 ± 28	34,87 ± 2,89	885 ± 28

L'ANOVA menée sur les %ERm et sur les TRm n'a ni révélé d'effet principal du Champ visuel (%ERm : $F(1,37) = 2,64, p = .112$; TRm : $F(1,37) < 1, p = .831$), ni d'interaction Champ visuel × Fréquence spatiale Cible/Distracteur (%ERm : $F(1,37) < 1, p = .630$; TRm : $F(1,37) = 1,74, p = .196$), ni d'interaction Champ visuel × Fréquence spatiale Cible/Distracteur × Congruence sémantique (%ERm : $F(1,37) < 1, p = .401$; TR : $F(1,37) = 2,71, p = .108$), ni d'interaction Champ visuel × Fréquence spatiale Cible/Distracteur × Congruence sémantique × Similarité physique (%ERm : $F(1,37) = 1,38, p = .248$; TRm : $F(1,36) < 1, p = .954$).

³ Dans l'Expérience 1, nous avons testé l'effet de la variable Sexe (Femme ; Homme) afin de nous assurer que cette variable n'influence pas les différences de performance entre les champs visuels de présentation des stimuli (CVD ; CVG). L'Expérience 3 a été menée auprès de 40 participants dont 35 étaient des femmes (soit plus de 85% de l'échantillon). Comme pour l'Expérience 2, ceci ne nous a pas permis de tester l'effet de la variable Sexe (Femme, Homme). Toutefois, les résultats de l'Expérience 1 n'ayant pas révélé d'effet du sexe sur les autres variables manipulées (i.e. champ visuel de présentation des stimuli, fréquences spatiales de la scène cible et la congruence sémantique), nous ne pensons pas que ce biais dans notre échantillon soit susceptible d'impacter nos résultats en termes de différences interhémisphériques.

4. Discussion générale

Les **Expériences 1, 2 et 3** avaient pour objectif d'étudier comment l'intégration des fréquences spatiales évolue entre la vision centrale et la vision périphérique et plus précisément, comment les informations visuelles contenues dans les BFS et les HFS s'influencent mutuellement en vision centrale et en vision périphérique lors de la catégorisation de scènes. Nous avons pour cela mené trois expériences dans lesquelles nous avons présenté aux participants des images hybrides précédemment utilisées par Kauffmann et al. (2017). Nous avons manipulé la congruence sémantique entre les deux scènes des images hybrides et nous les avons présentées soit dans le champ visuel central, soit dans le champ visuel périphérique (aléatoirement dans le champ visuel droit ou gauche).

Dans l'**Expérience 1**, nous avons utilisé des images hybrides physiquement dissimilaires. Dans l'**Expérience 2**, nous avons utilisé des images hybrides physiquement similaires. Dans l'**Expérience 3**, nous avons testé directement l'effet de la similarité physique sur les effets d'interférence sémantique en incluant à la fois des images hybrides physiquement similaires et dissimilaires dans le même design expérimental. Dans chaque expérience, les participants étaient répartis dans deux conditions expérimentales dans lesquelles ils devaient catégoriser la scène cible en BFS ou en HFS, tout en ignorant la scène distractive en HFS ou en BFS, respectivement. Ce paradigme expérimental nous a permis d'étudier comment une bande de fréquence spatiale qui doit être ignorée, influence l'autre bande de fréquence spatiale qui doit être traitée, en fonction de la région du champ visuel dans laquelle l'information visuelle est présentée (central vs. périphérique).

- **Avantage des BFS sur les HFS lors de la catégorisation de scènes**

Dans chaque expérience, les participants réalisaient une première session expérimentale (Session 1) dans laquelle ils devaient catégoriser des scènes filtrées en BFS et en HFS présentées séparément, en vision centrale dans les **Expériences 1 et 2** et en vision centrale et en vision périphérique dans l'**Expérience 3**. Cette session expérimentale avait pour objectif d'une part, de nous assurer que l'information sémantique contenue dans les scènes filtrées en BFS et en HFS utilisées dans les images hybrides était correctement catégorisée en dehors des effets d'interférence sémantique et d'autre part, de comparer le traitement des BFS et des HFS. Les résultats de ces trois expériences ont montré que les participants étaient capables de catégoriser correctement les scènes filtrées en BFS et en HFS en vision centrale (pourcentages d'erreurs inférieurs à 5%), mais aussi en vision périphérique (pourcentages d'erreurs inférieurs à 7%). Par ailleurs, sur la base de précédents travaux qui ont montré une préférence temporelle de l'analyse des BFS sur l'analyse des HFS (e.g., Breitmeyer, 1975 ; Shulman & Wilson, 1987 ; Badcock et al., 1990) et une prédominance d'une analyse *coarse-to-fine* des fréquences spatiales lors de la catégorisation de scènes (e.g., Kauffmann, Chauvin, Guyader et Peyrin, 2015 ; Musel & al., 2012 ; Schyns & Oliva, 1994), nous avons fait l'hypothèse de meilleures performances de catégorisation d'une scène en BFS qu'elles soient présentées individuellement ou dans des images hybrides. Les résultats des **Expériences 1 et 2** n'ont pas montré de différences significatives entre la catégorisation des scènes en BFS et des scènes HFS en vision

centrale. Comme nous l'avons souligné dans les **Expériences 1 et 2**, les pourcentages d'erreurs extrêmement bas dans la Session 1 (inférieurs à 4%) témoignent d'un effet plafond qui pourrait augmenter la difficulté à détecter statistiquement l'effet attendu. Néanmoins, dans l'**Expérience 3** où nous avons présenté les scènes filtrées ou les images hybrides aléatoirement en vision centrale et vision périphérique dans une même session expérimentale, nous avons cette fois-ci observé que les scènes en BFS étaient catégorisées plus rapidement que les scènes en HFS. Il semblerait que cette contrainte attentionnelle additionnelle soit propice à la mise en évidence des effets attendus.

En confrontant les résultats de ces trois expériences, nous ne pouvons confirmer que la catégorisation des scènes en BFS est systématiquement plus rapide que celle des HFS. Par contre, lorsque le stimulus visuel superpose BFS et HFS (cas des images hybrides), la partie en BFS amène systématiquement à de meilleures performances de catégorisation que la partie en HFS. On peut toutefois nous reprocher que la fréquence de coupure utilisée pour filtrer les scènes en BFS dans ces expériences (4 cpd, soit 24 cpi) était relativement élevée par rapport à celles utilisées dans les travaux précédents (généralement autour de 2 cpd ; Kauffmann, Bourgin, Guyader et Peyrin, 2015 ; Mu et Li, 2013 ; Schyns et Oliva, 1994). Les scènes en BFS incluaient donc une assez grande partie du spectre de fréquences spatiales de la scène, y compris des basses fréquences aux fréquences intermédiaires et contenaient donc des caractéristiques plus diagnostiques pour la catégorisation des scènes et des objets qui se situent entre 0 et 4 cpd (Caplette et al., 2014 ; McCotter et al., 2005). Par conséquent, notre choix méthodologique était particulièrement favorable pour la catégorisation des scènes en BFS au sein des images hybrides. En outre, le contraste de luminance n'était pas égalisé entre les scènes en BFS et les scènes en HFS dans ces expériences. Etant donné que le contraste de luminance des scènes naturelles diminue lorsque la fréquence spatiale augmente (Field, 1987), le contraste des scènes en BFS était plus élevé que celui des scènes en HFS (RMS moyen \pm écart-type des scènes naturelles en BFS : 0.24 ± 0.06 ; scènes artificielles en BFS : 0.24 ± 0.05 ; scènes naturelles en HFS : 0.06 ± 0.02 ; scènes artificielles en HFS : 0.07 ± 0.02 ; ces valeurs représentent l'écart-type de la luminance pour une échelle allant de 0 à 1). Dans ce contexte, des travaux ont montré que les temps de réponse diminuaient à mesure que le contraste augmentait (Harwerth & Levi, 1978). De nouveau, notre choix méthodologique était particulièrement favorable pour la catégorisation des scènes en BFS au sein des images hybrides.

- ***Effets d'interférence sémantique des BFS et des HFS lors de la catégorisation de scènes hybrides***

Les résultats obtenus lors de la catégorisation d'images hybrides en vision centrale et en vision périphérique ont montré dans les trois expériences que les performances de catégorisation de la scène cible étaient meilleures lorsque les deux scènes d'une image hybride étaient sémantiquement congruentes qu'incongruentes. Cet effet d'interférence sémantique indique que l'information sémantique contenue dans la scène distractive était quand même traitée et qu'elle perturbait la catégorisation de la scène cible, quel que soit son contenu en fréquences spatiales et sa position centrale ou périphérique dans le champ visuel. Toutefois, nos résultats montrent clairement que ces deux derniers facteurs expérimentaux modulent l'effet

d'interférence sémantique. En accord avec nos hypothèses, la scène distractive en BFS interférait plus fortement sur la catégorisation de la scène cible en HFS en vision périphérique qu'en vision centrale. Ce résultat suggère que le système visuel utiliserait très rapidement les BFS disponibles en vision centrale et en vision périphérique afin de générer des prédictions qui seraient validées ou non par les informations plus lentes en HFS. Mais ce que nous apprend nos expériences, c'est que lorsque la catégorisation d'une image hybride s'effectue en vision périphérique, le système visuel augmenterait le poids de l'interférence des prédictions issues de l'analyse en BFS et réduirait le poids d'une analyse détaillée portée par les HFS.

Alternativement, il est possible que l'augmentation de l'interférence des BFS sur les HFS en vision périphérique s'explique tout simplement par une baisse d'acuité visuelle et de la perception de la scène en HFS à mesure qu'on s'éloigne de la vision centrale. Cette interprétation alternative est tout d'abord remise en question par la mise en évidence dans les **Expériences 2 et 3** d'une interférence des HFS sur les BFS même en vision périphérique, suggérant l'utilisation des HFS et donc leur perception. Par ailleurs, dans l'**Expérience 3**, nous avons inclus une condition expérimentale où les participants devaient catégoriser des scènes en HFS en vision périphérique. Les résultats ont montré que les participants parvenaient à catégoriser aussi précisément les HFS que les BFS en vision périphérique (taux d'erreurs HFS : 5,80% ; taux d'erreurs BFS : 6,25% ; absence de différence statistique). Ainsi, il nous semble légitime de conclure qu'en s'éloignant de la vision centrale, le système visuel augmenterait le poids de l'interférence des prédictions issues de l'analyse en BFS pour guider le traitement des HFS, plutôt que de conclure que les HFS ne sont pas perçues.

Nos expériences montrent également que la catégorisation des scènes en BFS, bien plus rapide que celle des HFS dans l'**Expérience 3**, subit toutefois une forte interférence des HFS. En vision centrale, cet effet dépassait même l'interférence des BFS sur les HFS. Concrètement, les HFS congruentes avec les BFS amènent aux temps les plus rapides de la catégorisation des scènes, tandis que les HFS incongruentes sont celles qui ralentissent le plus fortement cette catégorisation. Ces résultats corroborent les résultats de Kauffmann et al. (2017). Comme ces auteurs, nous pensons que cet effet serait plus particulièrement imputable à la complexité des stimuli visuels utilisés dans nos expériences. Les images hybrides sont des stimuli visuels extrêmement complexes à traiter, d'autant plus lorsque les deux scènes d'une image hybride sont physiquement similaires. La perception de ces stimuli pourrait favoriser une stratégie d'analyse minutieuse de l'information visuelle (basée sur les HFS) au dépend d'une simple analyse grossière (basée sur les BFS). Dans la condition expérimentale où nous demandions explicitement aux participants de focaliser leur attention sur la scène en BFS de l'image hybride, la complexité visuelle de ces stimuli pourrait rendre impossible, voir même contre-productif d'ignorer le contenu sémantique en HFS pour donner une réponse sur la scène en BFS qui reste visuellement très peu informative d'un point de vue sémantique. Cette incapacité à ignorer le contenu sémantique en HFS pourrait alors tout particulièrement gêner le traitement du contenu sémantique en BFS lorsque les deux scènes appartiennent à des catégories sémantiques différentes. En conclusion, l'utilisation d'images hybrides pourrait avoir induit un biais dans les résultats de Kauffmann et al. (2017) et les nôtres en introduisant des stratégies bien particulières de la part des participants. Enfin, même si les HFS sont susceptibles d'interférer sur les BFS en vision périphérique, les résultats des **Expériences 2 et 3** montrent

que cette interférence demeure plus importante en vision centrale. Les HFS provenant principalement de la vision centrale, il semble logique que le système visuel augmente le poids d'une analyse détaillée portée par les HFS dans cette partie du champ visuel en particulier.

- ***Effet de la similarité physique entre les deux scènes des images hybrides***

Dans l'**Expérience 3**, nous avons manipulé la similarité physique entre les deux scènes qui composent une image hybride. Ce choix a été motivé par les résultats hétérogènes obtenus dans les **Expériences 1** et **2** dans lesquelles nous avons utilisé des images hybrides seulement physiquement dissimilaires pour la première et seulement physiquement similaires pour la seconde. Contrairement aux résultats de Mu et Li (2013), notre effet d'interférence des BFS sur la catégorisation d'une scène en HFS n'était pas intensifié par la similarité physique entre les deux scènes, que ce soit en vision centrale ou vision périphérique. Cette absence de résultats n'est pas surprenante au regard des travaux de Kauffmann et al. (2017) qui n'avaient pas non plus répliqué cet effet. Par contre, nous avons observé que l'effet d'interférence des HFS sur la catégorisation d'une scène en BFS était plus important lorsque les deux scènes étaient physiquement similaires. Ces mêmes résultats ont été observés par Kauffmann et al. (2017) et posent à nouveau la question du rôle particulier des HFS lors de la catégorisation d'images hybrides. Nous avons discuté plus haut de l'influence irrésistible des informations en HFS lors de la catégorisation d'images hybrides. En cas d'incertitude ou d'ambiguïté sur le stimulus visuel, comme c'est le cas avec des images hybrides, le traitement des détails contenus dans les HFS pourraient être plus pertinents pour démêler les interprétations et pour effectuer une tâche de catégorisation de scènes. Ainsi, lorsque les deux scènes sont sémantiquement incongruentes, mais que la ressemblance physique entre les deux scènes amène à penser qu'elles devraient donc appartenir à la même catégorie, cela aurait comme effet d'augmenter le conflit et l'effet d'interférence sémantique. On peut d'ailleurs observer sur la Figure [2.17](#) qu'une scène en HFS sémantiquement incongruente mais physiquement similaire augmente considérablement le taux d'erreurs de catégorisation de la scène en BFS (de 20%) par rapport à lorsqu'elle est physiquement dissimilaire.

- ***Différences interhémisphériques***

Dans ces trois expériences, nous nous sommes également intéressés à la façon dont chaque hémisphère cérébral serait impliqué dans le processus d'intégration des fréquences spatiales. Malheureusement, les résultats des trois expériences n'ont pas montré l'interaction double attendue entre la congruence sémantique, le champ visuel de présentation des stimuli et le contenu en fréquences spatiales de la scène cible. Nos résultats ont uniquement montré dans l'**Expérience 1** que l'avantage du traitement des BFS sur le traitement des HFS lors de la catégorisation des images hybrides était plus important lorsque les images étaient présentées dans le champ visuel gauche que droit et donc, initialement traitées par le cortex visuel de l'hémisphère droit. Ce résultat suggère que l'hémisphère droit serait plus performant que l'hémisphère gauche dans le traitement des BFS par rapport au traitement des HFS et soutient les précédentes études qui ont montré une supériorité de l'hémisphère droit dans le traitement

des BFS (Peyrin et al., 2003, 2004 ; Kauffmann, Chauvin, Pichat et Peyrin, 2015). Cependant, dans l'**Expérience 3**, nous avons directement testé la catégorisation des scènes filtrées en BFS et en HFS en vision périphérique, dans le champ visuel droit ou gauche, et les résultats n'ont pourtant pas mis en évidence de différences de traitement entre les champs visuels de présentation qui reflèterait des différences de traitement interhémisphérique. Au contraire, nous avons observé une précedence temporelle du traitement des BFS dans les deux hémichamps visuels qui est susceptible d'avoir masqué les différences interhémisphériques.

A nouveau, nous insistons sur le fait qu'il est difficile d'observer des différences interhémisphériques sur la seule base de données comportementales, comme c'est le cas dans nos expériences. En effet, les différences interhémisphériques correspondent à une réalité neuroanatomique, difficile à identifier à travers des données comportementales issues de la catégorisation d'images présentées en champs visuels divisés. En ce sens, l'absence de différences statistiques dans nos données comportementales ne signifie pas nécessairement qu'au niveau neuroanatomique, les deux hémisphères soient engagés de la même façon dans le processus d'intégration des fréquences spatiales. A titre d'exemple, Kauffmann et al. (2015) ont présenté des images hybrides en vision centrale et la technique de l'IRMf leur a permis de montrer que l'interférence sémantique des BFS sur la catégorisation d'une scène en HFS impliquait principalement des échanges entre le gyrus frontal inférieur et le gyrus fusiforme de l'hémisphère droit. Nous pouvons suggérer que c'est davantage notre paradigme expérimental qui ne contraindrait pas suffisamment les différences interhémisphériques plutôt que de conclure sur une absence de résultats. De nouvelles expériences menées en neuroimagerie pourraient permettre d'observer directement les différences d'activations entre les deux hémisphères.

- **Conclusion et limites**

En résumé, les résultats de ces trois expériences montrent que les informations en BFS et en HFS s'influencent mutuellement lors de la catégorisation de scènes. Ils montrent également que l'intégration et l'influence respective des BFS et des HFS lors de la catégorisation de scènes dépendraient des propriétés physiques du stimulus visuel et du champ visuel dans lequel le stimulus est catégorisé.

Malheureusement, notre paradigme ne permet pas de rendre compte des mécanismes sous-jacents aux effets d'interférence. Est-ce que l'interférence des BFS sur les HFS résulte bien d'un rôle prédictif des BFS ou est-ce que cet effet d'interférence résulterait tout simplement d'une intégration ascendante entre les informations issues des différentes bandes de fréquences spatiales (sans nécessairement faire de prédictions) ? En nous appuyant sur les précédents travaux de Kauffmann, Bourgin et al. (2015) et Kauffmann et al. (2017), nous défendons la thèse que les BFS seraient être utilisées par le système visuel pour catégoriser très rapidement une scène visuelle, mais aussi pour générer des prédictions qui guideraient l'analyse des HFS. Dans notre paradigme de catégorisation d'images hybrides, nous proposons que l'effet d'interférence des BFS sur les HFS résulterait de mécanismes cognitifs différant de ceux impliqués dans l'effet d'interférence des HFS sur les BFS.

- La tâche de catégorisation rapide de la scène en BFS rendrait compte de la stratégie d'analyse *coarse-to-fine* d'une scène et de l'**intégration ascendante des informations détaillées issues des HFS**. En ce sens, nos résultats montrent effectivement que la catégorisation d'une scène en BFS tend systématiquement à être plus rapide que celle des HFS. Mais il est également intéressant d'observer qu'à l'image de l'analyse *coarse-to-fine*, l'intégration d'informations en HFS congruentes amènent aux meilleures performances de catégorisation (en termes de précision et de temps de réponse). Dans ce contexte, l'interférence des HFS incongruentes sur la catégorisation d'une scène en BFS serait la conséquence de l'intégration d'une information ascendante erronée. Nous avons observé que cette interférence était d'autant plus grande que les deux scènes se ressemblaient physiquement. Nos résultats suggèrent ainsi que le contenu physique de l'information ascendante jouerait un rôle fondamental dans le processus d'intégration *coarse-to-fine*.
- La tâche de catégorisation plus lente de la scène en HFS permettrait de mettre en évidence les **mécanismes prédictifs descendants rapidement issus des BFS**. On peut voir en effet que même si l'effet d'interférence des BFS sur la catégorisation d'une scène en HFS en vision centrale demeure faible, cet effet d'interférence intervient dans un contexte où les participants mettent globalement bien plus de temps pour catégoriser une scène en HFS, qu'elle soit congruente ou incongruente. L'interférence des BFS sur la catégorisation d'une scène en HFS serait alors la conséquence d'une erreur de prédictions.

Par rapport à l'étude de Kauffmann et al. (2017), nos expériences permettent de souligner l'importance de considérer les différences de traitement existant entre la vision centrale et la vision périphérique dans l'étude des effets d'interférence sémantique. Conformément à nos hypothèses, la génération de prédictions issue des BFS utilisées pour guider le traitement des HFS serait plus importante en vision périphérique qu'en vision centrale. Au quotidien où les informations disponibles en vision périphérique sont floues, il ne semble pas surprenant que le système visuel se soit doté d'un système de gestion des prédictions suffisamment performant pour guider la détection de détails pertinents en vision périphérique et ainsi les mouvements des yeux qui déplaceraient la vision centrale sur ces détails.

Cependant, comme nous l'avons souligné au-dessus, l'utilisation d'images hybrides peut refléter des limites :

- (1) Nous pouvons nous interroger sur le caractère particulièrement ambigu des images hybrides qui pourrait augmenter le poids d'une analyse détaillée portée par les HFS lors de la réalisation de la tâche visuelle. Qui plus est, la superposition spatiale des informations en BFS et en HFS dans les images hybrides est non écologique. Dans la vie réelle, il est rare que les informations en BFS et HFS qui se superposent dans une partie du champ visuel proviennent de stimuli différents.
- (2) L'excentricité des bords internes et externes des stimuli latéralisés était fixée respectivement à 2° et 8° d'excentricité rétinienne à partir du point de fixation central. Dans le préambule de cette thèse, nous avons précisé que la vision centrale comprenait la fovéa (1° d'excentricité rétinienne) et la parafovéa (4-5° d'excentricité rétinienne), la

vision périphérique correspondant à la surface restante. Nos stimuli n'étaient donc pas présentés exclusivement en vision périphérique, mais couvraient une bonne partie de la vision parafovéale. Nous avons fait le choix de présenter nos stimuli à seulement 2° d'excentricité sur la base des précédents travaux Carole Peyrin et collaborateurs sur la spécialisation hémisphérique du traitement des fréquences spatiales (Peyrin et al., 2003 ; Peyrin, Chokron et al., 2006 ; Peyrin, Mermillod et al., 2006). Pour ces auteurs, 2° est une excentricité rétinienne suffisante pour adresser l'information visuelle à un seul hémisphère, mais cette valeur garantit aussi une bonne visibilité des stimuli en HFS même lorsqu'ils sont excentrés. En effet, nous avons vu dans le **Chapitre 1** que l'extraction des HFS n'est possible que dans la rétine centrale, alors que les BFS sont principalement extraites au niveau de la rétine périphérique. La fonction de sensibilité au contraste (FSC) en fonction de l'excentricité rétinienne (Rovamo, Virsu & Näsänen, 1978) indique que pour des stimuli excentrés à 1,5° et 7,5° d'angle visuel comme c'est approximativement le cas pour les bords interne (2°) et externe (8°) de nos stimuli, la plus haute fréquence spatiale visible est comprise entre 32 cpd et 8 cpd, respectivement. Nos scènes en HFS conservant les fréquences spatiales supérieures à 6 cpd, elles restaient visibles en choisissant ces excentricités dans nos expériences. Malheureusement, ces choix d'excentricité rétinienne ne nous ont pas permis de tester exclusivement la vision périphérique. De façon générale, cette seconde limite pose la question de la manipulation des stimuli en HFS lorsqu'ils sont présentés en vision périphérique. L'extraction des HFS en vision périphérique semble contre-nature avec les propriétés du système visuel. Il semble donc non écologique d'utiliser des tâches visuelles où les participants doivent traiter des HFS en vision périphérique (ou parafovéale dans le cas présent).

Dans les **Chapitres 3 et 4** de cette thèse, nous nous sommes donc plus particulièrement intéressés à l'intégration des BFS et des HFS sans manipuler les fréquences spatiales de nos stimuli, mais en considérant leur disponibilité dans le champ visuel. Ceci nous a nécessairement amené à nous intéresser aux traitements réalisés en vision centrale (plutôt en HFS) et en vision périphérique (plutôt en BFS) et leur contribution respective lors de la catégorisation de scènes dans le **Chapitre 3**, avant d'aborder leurs interactions dans le **Chapitre 4**.

Chapitre 3 - Etude de la contribution respective de la vision centrale et de la vision périphérique lors de la catégorisation de scènes

Cette partie expérimentale a pour objectif d'étudier la catégorisation de scènes en vision centrale et en vision périphérique et de quelle façon ces différentes régions du champ visuel contribuent respectivement à la reconnaissance rapide de scènes.

Comme nous l'avons vu dans la partie théorique de cette thèse, la partie centrale de la rétine coderait principalement les HFS, tandis que la partie plus périphérique coderait plutôt les BFS. Nous avons également vu plusieurs modèles qui démontrent le rôle fondamental des fréquences spatiales pour la reconnaissance visuelle (Bar, 2003 ; Bar, 2007 ; Kauffmann et al., 2014 ; Roux-Sibilon, 2020 ; Schyns & Oliva, 1994). Selon ces modèles, l'analyse visuelle commence par l'extraction parallèle de différentes caractéristiques élémentaires d'un stimulus visuelle à différentes fréquences spatiales et suit une séquence d'analyse par défaut *coarse-to-fine*. Les BFS sont extraites en premier, ce qui permet une analyse grossière de la scène, avant l'analyse d'informations détaillées contenues dans les HFS. Nous avons vu que plusieurs études conduites par l'équipe de Carole Peyrin au Laboratoire de Psychologie et NeuroCognition ont montré que l'analyse *coarse-to-fine* était avantageuse pour catégoriser rapidement des scènes (Musel et al., 2012 ; Musel et al., 2014 ; Kauffmann, Chauvin, Guyader & Peyrin, 2015 ; Kauffmann, Chauvin, Pichat & Peyrin, 2015). Pour rappel, les chercheurs présentaient des séquences dynamiques composées de six versions filtrées de la même scène, assemblées des BFS aux HFS afin d'imiter un traitement *coarse-to-fine*, ou des HSF aux BFS afin d'imiter un traitement *fine-to-coarse*. Les participants devaient catégoriser ces séquences comme étant des scènes d'intérieur ou d'extérieur. Les résultats ont montré qu'ils étaient plus rapides pour catégoriser les séquences *coarse-to-fine* que les séquences *fine-to-coarse*, confirmant ainsi que le traitement des informations en BFS avec celui des informations en HFS est plus avantageux pour la catégorisation rapide des scènes.

Étant donné que les signaux en BFS proviendraient principalement du champ visuel périphérique, tandis que les signaux en HFS proviendraient principalement du champ visuel central, ces études passées nous amènent à nous interroger sur la contribution relative de la vision périphérique par rapport à la vision centrale pour la catégorisation des scènes. En accord avec une séquence *coarse-to-fine* du traitement des fréquences spatiales, est-ce que la catégorisation des scènes suit également un traitement allant de la vision périphérique à la vision centrale ? La précedence temporelle des BFS et leur prédominance en vision périphérique nous amène alors à nous interroger sur le rôle de la vision périphérique pour catégoriser rapidement

une scène, mais aussi sur son rôle dans le processus de génération de prédictions. Plus globalement, est-ce que la vision périphérique est utilisée pour catégoriser rapidement une scène visuelle ? Cette question est d'autant plus pertinente que malgré la faible résolution spatiale de la vision périphérique, elle resterait très efficace pour catégoriser des scènes. Par exemple, Boucart et al. (2013) ont directement étudié la catégorisation de scène en vision périphérique. Les auteurs présentaient à des participants des paires de grandes scènes (20° d'angle visuel), l'une à droite et l'autre à gauche d'un point de fixation central. L'excentricité des scènes variait entre 10° et 70° d'angle visuel. Les participants devaient décider de quel côté apparaissait la scène qui appartenait à une catégorie prédéfinie. Les résultats ont montré que les performances de catégorisation étaient bien au-dessus du hasard lors de la présentation des scènes au niveau périphérique, et ce même lorsqu'elles étaient présentées à 70° d'excentricité rétinienne. Ces résultats suggèrent que malgré la faible résolution spatiale de la vision périphérique, elle reste très efficace pour catégoriser des scènes.

Plus important, plusieurs études psychophysiques menées par Lester Loschky et ses collaborateurs ont mis en évidence l'avantage de la vision périphérique sur la vision centrale lors de la catégorisation des scènes (Larson et al., 2014 ; Larson & Loschky, 2009 ; Loschky et al., 2019). Les travaux de Larson et Loschky (2009) figurent comme l'étude pionnière en la matière. Les auteurs utilisaient un paradigme Fenêtre-Scotome (*Window-Scotoma*) dans lequel les participants pouvaient voir soit la partie centrale de la scène dans la condition Fenêtre, soit la partie périphérique de la scène en cachant son centre dans la condition Scotome. Dans une condition contrôle, la scène était présentée entièrement. La taille de la fenêtre et du scotome variait suivant quatre rayons différents (1° , 5° , $10,8^\circ$ et $13,6^\circ$ d'angle visuel). La condition 5° de rayon correspondait à la présence ou l'absence effective de vision centrale (vision fovéale et parafovéale) et dans la condition $0,8^\circ$ d'excentricité, la quantité (surface) d'information visible était la même dans la condition Fenêtre et la condition Scotome (Figure 3.1). Pour chaque essai, la scène était suivie d'un mot congruent ou incongruent avec la catégorie de la scène qui restait affiché jusqu'à la réponse du participant. Les participants devaient choisir si le mot présenté (e.g., plage, montagne, maison, rue) correspondait ou non à la catégorie de la scène. Les résultats ont montré que les performances des participants étaient meilleures dans la condition Scotome à 5° (vision périphérique) que dans la condition Fenêtre à 5° (vision centrale). De plus, les performances dans cette condition Scotome ne différaient pas de celles obtenues en condition contrôle lorsque la scène était présentée entièrement. La vision périphérique - de faible résolution spatiale - serait plus utile que la vision centrale - de haute résolution spatiale - pour catégoriser rapidement une scène. Bien que ce résultat ait été observé à plusieurs reprises par cette équipe de recherche, Larson et al. (2014) font toutefois remarquer que la contribution relative de la vision centrale et de la vision périphérique varie avec le temps. En manipulant le temps de présentation des stimuli Fenêtre-Scotome, ces auteurs ont observé qu'il y aurait en fait une supériorité de la vision centrale pendant les 100 premières millisecondes de traitement et que la supériorité de la vision périphérique apparaissait par la suite. Pour interpréter ces résultats, les auteurs ont proposé l'hypothèse *zoom-out* selon laquelle l'attention est d'abord focalisée en vision centrale, puis se développe rapidement en vision périphérique.

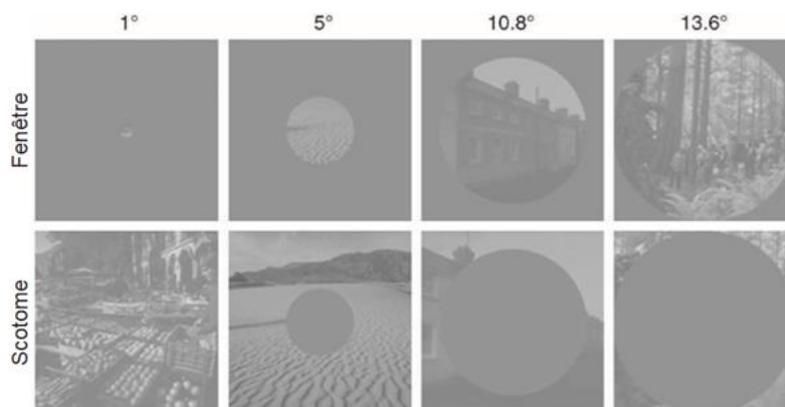


Figure 3.1 : Exemple de stimuli utilisés dans Larson et Loschky (2009) en condition "Fenêtre ou Scotome" selon différents rayons exprimés en degré d'angle visuel.

L'importance de la vision périphérique pour la catégorisation de scènes a également été observée chez des patients atteints de troubles ophtalmologiques (Tran et al., 2010 ; Musel et al., 2011 ; Peyrin et al., 2017). Par exemple, les études menées auprès de patients atteints de DMLA, une maladie rétinienne qui affecte particulièrement la macula et qui se traduit donc par une perte progressive de la vision centrale, permettent d'une part d'étudier les déficits de ces patients en vision centrale, mais aussi leurs capacités résiduelles en vision périphérique. Tran et al. (2010) ont par exemple montré que des patients atteints de DMLA étaient capables de catégoriser avec précision des scènes de 30° de diamètre (taux de réussite supérieur à 75%). Musel et al. (2011), puis Peyrin et al. (2017) ont étudié la catégorisation de grandes scènes filtrées en BFS et HFS auprès de patients DMLA. Alors que la catégorisation des scènes en HFS était affectée par l'atteinte de la rétine centrale, les capacités des patients pour catégoriser les scènes en BFS étaient totalement préservées. Ces études suggèrent que l'information visuelle de faible résolution spatiale disponible en vision périphérique suffit à catégoriser efficacement des scènes. D'autres études menées auprès de patients atteints d'un glaucome ont quant à elles permis d'étudier directement l'impact d'une perte de la vision périphérique sur la reconnaissance visuelle (Lenoble et al., 2016 ; Roux-Sibilon et al., 2018). Cette pathologie, causée par une destruction progressive des cellules ganglionnaires et du nerf optique, affecte principalement la rétine périphérique au début de la maladie. Roux-Sibilon et al. (2018) ont, par exemple, montré que des patients atteints de glaucome faisaient plus d'erreurs que des participants sains pour catégoriser des scènes présentées dans leur champ visuel central pourtant considéré comme intact. Nous reviendrons sur l'interprétation de ces résultats dans le **Chapitre 5** de ce manuscrit qui porte sur des patients atteints de glaucome, mais cette étude suggère que l'intégrité de la rétine périphérique est essentielle à une catégorisation efficace des scènes visuelles. Ainsi les études menées auprès de patients ophtalmiques démontrent elles aussi l'importance de la vision périphérique pour la catégorisation de scènes.

Prises dans leur ensemble, les études qui portent sur la stratégie d'analyse *coarse-to-fine* des fréquences spatiales et celles qui comparent la catégorisation de scènes en vision centrale et en vision périphérique sont cohérentes avec l'idée que l'analyse visuelle de la scène suit une séquence de traitement *coarse/peripheral-to-fine/central*. Le premier objectif de ce

chapitre expérimental était de tester directement cette hypothèse à l'aide de séquences dynamiques imposant soit une analyse *peripheral-to-central*, soit une analyse *central-to-peripheral* des informations visuelles (Figure 3.2). Ces séquences, inspirées des séquences *coarse-to-fine* et *fine-to-coarse* utilisées par Kauffmann, Chauvin et al. (2015), ont été construits à partir de cinq stimuli d'une même scène, chaque stimulus révélant différentes parties du champ visuel à travers des anneaux circulaires de différentes excentricités et assemblés de la périphérie au centre (séquence *peripheral-to-central*) ou du centre à la périphérie (séquence *central-to-peripheral*).

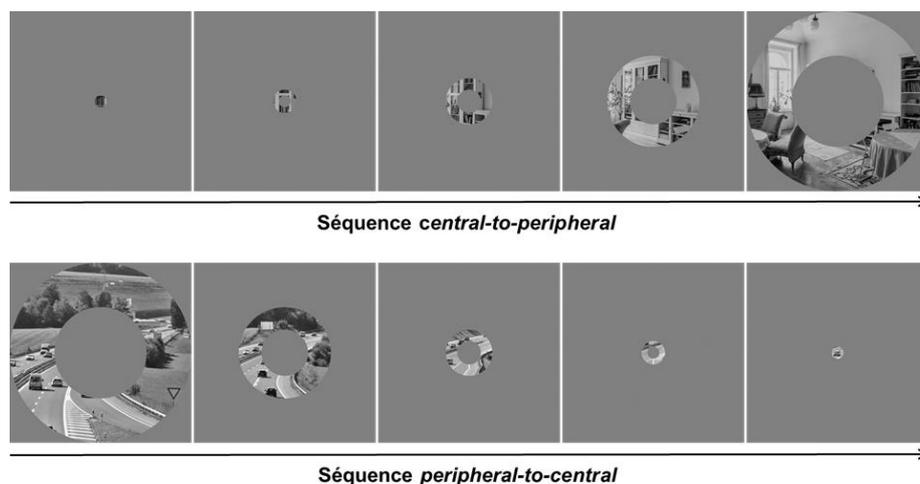


Figure 3.2 : Représentation schématique des séquences *central-to-peripheral* et *peripheral-to-central*.

Il est important de préciser que dans les expériences de ce chapitre, nous avons choisi de ne pas filtrer le contenu en fréquence spatiales de nos stimuli d'une part en raison des limites rapportées en fin de **Chapitre 2** (i.e. que la manipulation des HFS en vision périphérique nous semble contre-nature avec les propriétés du système visuel) et, d'autre part, nous considérons que la vision périphérique agit comme un filtre passe-bas naturel (i.e. l'information en HFS disparaît progressivement en vision périphérique tandis que l'information en BFS est préservée).

Dans l'**Expérience 4** de cette thèse, nous avons testé la catégorisation de chaque anneau composant les scènes dynamiques de la même manière que dans l'expérience de Larson et Loschky (2009). Dans l'**Expérience 5**, nous avons testé la catégorisation des séquences *central-to-peripheral* et *peripheral-to-central*. Pour ce faire, le disque central et les anneaux ont été assemblés afin de créer des séquences dynamiques.

Les **Expériences 4 et 5** de cette thèse ont été soumises à publication dans la revue *Vision Research*. Les experts (dont Lester Loschky) ont soulevé de nombreux biais méthodologiques que nous avons considérés en réalisant trois nouvelles expériences (**Expériences 6, 7 et 8**). Ce sont ces dernières qui ont été publiées dans la revue *Vision Research*.

I. Expériences 4 et 5 - Importance de la vision périphérique par rapport à la vision centrale

L'objectif de l'**Expérience 4** de cette thèse était d'évaluer l'importance de la vision périphérique par rapport à la vision centrale lors de la catégorisation de scènes. Nous avons utilisé de grandes photographies de scènes de la vie de tous les jours. Ces photographies étaient suffisamment grandes (30° d'angle visuel) pour couvrir à la fois le champ visuel central et le champ visuel périphérique. Ces photographies ont été prises soit en extérieur (e.g., ville, autoroute, montagne, plage), soit en intérieur (e.g., cuisine, salle de bain, salon, chambre, bureau). Nous avons présenté aux participants différentes parties de ces scènes, de différentes excentricités réiniennes (Figure 3.3), pendant un temps suffisamment court pour empêcher les mouvements oculaires (100 ms). Plus précisément, nous leur présentions soit la partie centrale de la scène, soit des parties plus périphériques de la scène révélée à travers quatre anneaux de différentes excentricités réiniennes. Les participants devaient catégoriser chaque stimulus comme étant soit une scène d'intérieur, soit une scène d'extérieur. Sur la base de l'étude de Larson et Loschky (2009), nous nous attendions à ce que les performances soient meilleures pour les stimuli périphériques que les stimuli centraux.

Dans cette expérience, nous avons également manipulé le contenu disponible dans la partie centrale de la scène. En effet, il est possible que l'avantage de la vision périphérique observé par Larson et Loschky (2009) lors de la catégorisation des scènes résulte d'une absence d'information pertinente / diagnostique dans la partie centrale de la scène. Par exemple, comme on peut le voir dans la Figure 3.3a, la partie centrale d'une scène de désert ne comporte aucun objet. Notre objectif était donc d'évaluer si la catégorisation d'une scène en vision centrale repose en fait sur la présence d'un contenu sémantique pertinent, mais aussi si la présence d'un tel contenu suffit à supprimer expérimentalement l'avantage de la vision périphérique observé par Larson et Loschky (2009). Dans notre expérience, la partie centrale des scènes contenait donc soit un objet sémantiquement lié à la catégorie scène (par exemple, des ustensiles de cuisine pour une scène d'intérieur, ou une maison pour une scène d'extérieur), soit aucun objet sémantiquement pertinent (par exemple, une partie d'un mur pour une scène d'intérieur, une partie de la mer pour une scène d'extérieur). De plus, nous avons construit nos stimuli de manière à ce que la surface des informations visuelles révélées par chaque anneau prenne en compte le facteur de magnification corticale. En effet, comme nous l'avons vu au début du chapitre théorique, les projections des informations réiniennes vers le cortex visuel primaire (V1) sont déformées. Un grand nombre de cellules dans V1 sont dédiées au traitement des informations visuelles provenant de la rétine centrale tandis qu'un plus petit nombre de cellules sont dédiées aux informations provenant de la vision périphérique (Engel et al., 1994 ; Engel, Glover, & Wandell, 1997 ; Daniel & Whitteridge, 1961). La portion fovéale de la rétine, qui ne représente que 1% de la surface totale de la rétine, est représentée sur 50% dans V1. Par conséquent, les informations visuelles provenant du champ visuel central sont sur-représentées au niveau cortical relativement aux informations visuelles provenant du champ visuel périphérique. Ce facteur de magnification corticale explique d'ailleurs en partie la diminution de l'acuité visuelle avec l'excentricité réinienne (Duncan & Boynton, 2003). Afin de considérer

ce facteur de magnification dans notre expérience, nous avons augmenté la largeur et la surface de nos stimuli en fonction de leur excentricité rétinienne (Figure 3.2). Nous avons bien conscience qu'en respectant les propriétés naturelles du système visuel, le stimulus le plus périphérique contenait plus d'informations visuelles potentiellement plus utiles pour la catégorisation de scènes que le stimulus le plus central. Ce biais inhérent à la perception visuelle sera discuté plus loin. Mais il faut toutefois remarquer que même si les informations sémantiques disponibles en vision périphérique étaient plus nombreuses, elles demeuraient de plus faible résolution spatiale.

L'objectif de l'**Expérience 5** était de tester l'hypothèse d'une analyse *coarse/peripheral-to-fine/central* lors de la catégorisation de scènes. Nous avons créé des séquences dynamiques composées des stimuli utilisés dans l'**Expérience 4**. De cette façon, la scène entière était révélée successivement à travers les stimuli, soit en allant de la périphérie vers le centre de la scène (séquence *peripheral-to-central*), soit en allant du centre vers la périphérie de la scène (séquence *central-to-peripheral* ; Figure 3.2). Les stimuli étant issus de l'**Expérience 4**, nous avons pu également considérer dans cette expérience l'effet de la présence d'un contenu sémantique pertinent au centre des scènes et contrôler le facteur de magnification corticale. Les participants devaient catégoriser chaque séquence comme étant soit une scène d'intérieur, soit une scène d'extérieur. Nous nous attendions à de meilleures performances de catégorisation pour les séquences *peripheral-to-central* que *central-to-peripheral*. Dans cette expérience, nous avons également manipulé le temps de présentation des séquences. Notre hypothèse était qu'un traitement *peripheral-to-central* serait d'autant plus privilégié que le système visuel dispose de peu de temps pour traiter l'information. En effet, les informations en BFS disponibles en vision périphérique pourraient être davantage prises en compte lors de présentations courtes, tandis que l'analyse plus lente des informations en HFS disponibles en vision centrale pourrait prendre l'avantage lors de présentations plus longues. Ainsi, les séquences étaient présentées soit pendant 500 ms (chaque stimulus dans la séquence étant présenté pendant 100 ms, comme dans l'**Expérience 4**), soit 100 ms (chaque stimulus était présenté pendant 20 ms de façon à ce que la séquence totale ait la même durée de présentation qu'un seul stimulus de l'**Expérience 4**).

1. **Expérience 4 - Catégorisation de larges scènes et de parties de scènes de différentes excentricités rétiniennes**

1.1. Méthode

1.1.1. Participants

Trente étudiants de l'Université Grenoble Alpes (22 femmes ; âge moyen : $20,53 \pm 1,94$) ont participé à cette expérience. Aucune donnée de référence n'était disponible pour estimer la taille de l'échantillon, nous nous sommes donc basés sur un échantillon conservateur de 30 participants par rapport à la taille d'échantillon minimale de 20 participants dans nos précédentes **Expériences 1, 2 et 3**. Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Tous les participants ont donné leur consentement libre et éclairé avant de participer à cette expérience qui a été menée en conformité avec le code d'éthique de l'*American Psychological Association* (APA) et approuvée par un comité et approuvée par un comité d'éthique (CERGA, Comité d'éthique de la recherche Grenoble Alpes, IRB00010290-2018-10-16-50). A la fin de l'expérience, chaque participant recevait un bon de 0,25 point par quart d'heure d'expérience à valoir lors des examens universitaires.

1.1.2. Stimuli

Les stimuli utilisés ont été construits à partir de 50 photographies en couleur de scènes issues du site web Pixabay (<https://pixabay.com/fr/>), un site de partage de photographies sous licence CC0 (Creative Commons Zero ; i.e. placées dans le domaine public, sans aucune restriction d'usage). La moitié des photographies représentait des scènes d'intérieur (e.g., salon, cuisine, salle de bain) et l'autre moitié représentait des scènes d'extérieur (e.g., paysage urbain, montagne, plage). Puisque nous nous plaçons dans le contexte expérimental de la catégorisation rapide des scènes, le choix de ces catégories a été motivé par le fait que la catégorisation au niveau superordonné précède la catégorisation au niveau de base lors de la reconnaissance visuelle (voir par exemple Rousselet, Joubert & Fabre-Thorpe, 2005 ; Joubert, Rousselet, Fize & Fabre-Thorpe, 2007 ; Loschky & Larson, 2010 ; Kadar & Ben-Shahar, 2012).

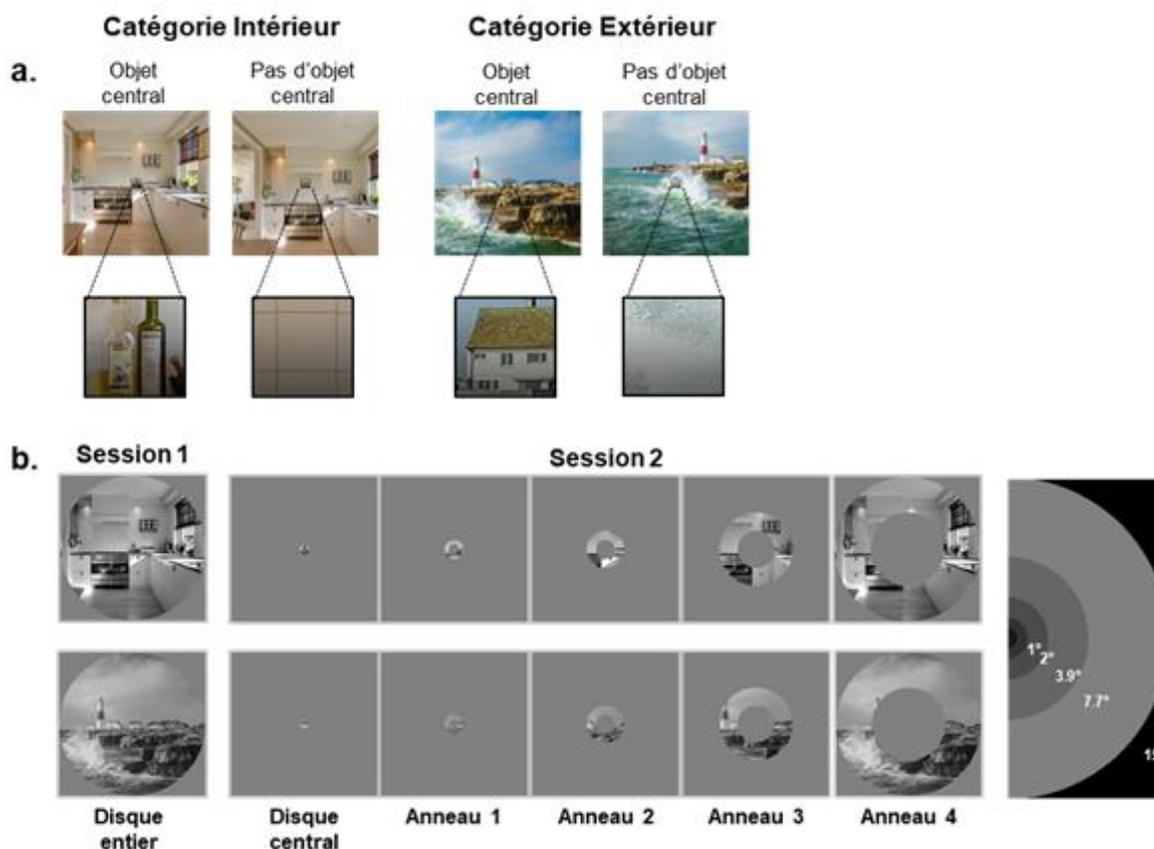


Figure 3.3 : (a) Exemple de scènes d'intérieur et d'extérieur de taille 1500×1500 pixels présentées avec ou sans objet sémantiquement relié à la catégorie de la scène à l'intérieur d'un carré central de taille 100×100 pixels. (b) Exemple de stimuli utilisés dans la Session 1 (Disque entier) et dans la Session 2 (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4) de l'Expérience 4. L'image de droite illustre les différentes excentricités (en degrés d'angle visuel) des bords intérieurs et extérieurs des stimuli qui augmentent selon une progression logarithmique.

La manipulation des images a été réalisée à l'aide de MATLAB (Mathworks Inc., Sherborn, MA, USA). Nous avons d'abord découpé chaque photographie de façon à construire deux images carrées de 1500×1500 pixels (Figure 3.3a). Pour la première image, le découpage de la scène était centré sur un objet tenant dans un carré de 100×100 pixels. Nous avons veillé à ce que l'objet soit représentatif de la catégorie de la scène⁴. Par exemple, pour les scènes d'intérieur, l'objet pouvait être un appareil électroménager, de la vaisselle, des ustensiles de cuisine, un livre, une horloge, une lampe, une radio, une bougie, un vase ou une paire de lunettes. Pour les scènes d'extérieur, l'objet pouvait être une voiture, un bateau, une maison, un

⁴Afin de vérifier la valeur sémantique de l'objet contenu dans le disque central, nous avons mené une expérience pilote. Nous avons pour cela montré à un groupe de dix étudiants tous les stimuli du disque central contenant un objet sémantiquement relié à la catégorie de la scène (soit 100 stimuli). Ils devaient catégoriser les stimuli (Intérieur vs. Extérieur) et ensuite nommer les objets (e.g., une lampe, un vase, un bateau ou une maison). La catégorie de la scène a été correctement catégorisée (100% de réponses correctes) et les objets ont été nommés correctement (97% de réponses correctes) confirmant ainsi que nos stimuli étaient centrés sur un objet sémantiquement relié à la catégorie de la scène.

building, un arbre, une plante, un réverbère, une montagne ou un pont. Pour la deuxième image, le découpage de la scène était centré sur carré de la scène de 100×100 pixels ne contenant pas d'objet. Toutes les images ont ensuite été converties en 256 niveaux de gris en faisant la moyenne des valeurs rouge, verte et bleue de chaque pixel. Nous avons également égalisé la luminance et le contraste des stimuli. Tous les stimuli ont été normalisés de façon à obtenir une luminance moyenne de 0,5 et un contraste RMS (Root Mean Square) de 0,2 pour des valeurs d'intensité des pixels comprises entre 0 et 1 (ceci correspond à une luminance moyenne de 128 et un contraste RMS de 51 sur une échelle de 256 niveaux de gris).

Nous avons ainsi créé 100 scènes (50 scènes d'intérieur et 50 scènes d'extérieur) de taille 1500×1500 pixels. Nous avons fixé une taille angulaire des scènes à $\sim 30^\circ$ d'angle visuel de façon à ce qu'elles couvrent à la fois la vision centrale et la vision périphérique. Pour cela, les participants étaient installés à une distance de 70 cm d'un écran DELL ULTRASHARP (60 Hz) de 30 pouces ($61,4 \times 40$ cm), avec un affichage de résolution 2560×1600 pixels. Pour chaque scène, nous avons ensuite construit six stimuli qui révélaient différentes parties de la scène (Figure 3.3b). La largeur des stimuli augmentait de façon logarithmique en fonction de l'excentricité rétinienne, selon un facteur f obtenu comme suit : $f = (\text{rayon maximum}/\text{rayon minimum})^{(1/\text{nombre d'anneaux}-1)}$. Nous avons privilégié cette progression logarithmique afin de considérer le facteur de magnification corticale (Daniel & Whitteridge, 1961). La surface de l'image révélée par les stimuli augmentait donc avec leur excentricité. Le premier stimulus (ou Disque entier) révélait la scène à travers un large disque d'un rayon de 15° d'angle visuel (i.e. 750 pixels). Le second stimulus (ou Disque central) révélait la scène à travers un petit disque d'un rayon de seulement 1° d'angle visuel (i.e. 49 pixels). Les quatre derniers stimuli révélaient la scène à travers des anneaux de différentes excentricités rétiniennes. Pour le premier anneau (Anneau 1), adjacent au Disque central, les bords interne et externe étaient respectivement fixés à 1° d'angle visuel (49 pixels) et $1,98^\circ$ d'angle visuel (97 pixels). Pour le second anneau (Anneau 2), adjacent au premier anneau, les bords interne et externe étaient respectivement fixés à $1,98^\circ$ d'angle visuel (97 pixels) et $3,91^\circ$ d'angle visuel (192 pixels). Pour le troisième anneau (Anneau 3), adjacent au second anneau, les bords interne et externe étaient respectivement fixés à $3,91^\circ$ d'angle visuel (192 pixels) et $7,74^\circ$ d'angle visuel (379 pixels). Enfin, pour le quatrième anneau (Anneau 4), adjacent au troisième anneau, les bords interne et externe étaient respectivement fixés à $7,74^\circ$ d'angle visuel (379 pixels) et 15° d'angle visuel (750 pixels). Le premier stimulus, ou Disque entier, correspondait ainsi à la somme ou superposition de tous les autres stimuli. Tous les stimuli étaient présentés sur un fond gris de luminance moyenne de 0,5 pour des valeurs d'intensité des pixels comprises entre 0 et 1.

1.1.3. Procédure

L'expérience a été réalisée dans un box expérimental du Laboratoire de Psychologie et NeuroCognition (LPNC) de l'Université de Grenoble Alpes. Les participants étaient installés dans l'obscurité face à un écran d'ordinateur. Ils posaient leur menton sur une mentonnière installée sur l'axe sagittal médian du participant, à 70 cm de l'écran afin de respecter la taille angulaire de 30° d'angle visuel des stimuli. L'expérience était programmée grâce au logiciel E-

prime 2.0 (E-prime Psychology Software Tools Inc., Pittsburgh, USA) et elle était présentée sur l'écran DELL ULTRASHARP.

Tous les participants réalisaient deux sessions expérimentales. Dans la première session expérimentale (Session 1), les participants devaient catégoriser les Disques entiers. Cette session expérimentale avait pour but de nous assurer que les scènes entières étaient catégorisées correctement, mais aussi de tester si la seule présence d'un objet au centre de la scène améliorerait les performances de catégorisation de scènes visuelles. Dans la deuxième session expérimentale (Session 2), les participants devaient catégoriser une partie de la scène. Nous avons pour cela présenté aléatoirement la portion centrale de la scène (Disque central), ou les portions périphériques révélées à travers les anneaux de différentes excentricités (Anneau 1, Anneau 2, Anneau 3 et Anneau 4).

Pour chaque session expérimentale, un essai se déroulait comme suit (Figure 3.4) : un point de fixation central (noir sur un fond gris de luminance 0,5) était présenté pendant 500 ms et suivi de la présentation du stimulus (Session 1 : Disque entier ; Session 2 : Disque central ou anneaux périphériques) pendant 100 ms et d'un écran gris (luminance : 0,5) de 1900 ms au cours duquel les participants pouvaient répondre. Les participants devaient catégoriser la scène comme étant soit une scène d'intérieur, soit une scène d'extérieur par l'intermédiaire des touches 1 et 2 du clavier d'ordinateur placé sur l'axe sagittal médian des participants. La moitié des participants appuyait sur la touche 1 avec l'index de la main droite pour la catégorie Intérieur et sur la touche 2 avec le majeur de la main droite pour la catégorie Extérieur. Les touches de réponse étaient contrebalancées pour l'autre moitié des participants. Les participants avaient pour consigne de répondre le plus correctement et le plus rapidement possible. La Session 1 comprenait 100 essais (25 scènes d'intérieur et 25 scènes d'extérieur avec un objet au centre, 25 scènes d'intérieur et 25 scènes d'extérieur sans objet au centre) et durait environ 4 minutes. La Session 2 comprenait 500 essais (pour chaque type de stimulus Disque Central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4, il y avait 25 scènes d'intérieur et 25 scènes d'extérieur avec un objet au centre, 25 scènes d'intérieur et 25 scènes d'extérieur sans objet au centre) et durait environ 20 minutes, entrecoupés de pauses tous les 100 essais.

Pour chaque essai, nous avons enregistré l'exactitude de la réponse et le temps de réponse (en millisecondes). Avant de commencer chaque session expérimentale, les participants réalisaient quelques essais d'entraînement afin de se familiariser avec la tâche et les stimuli (25 essais pour la Session 1, 25 essais pour la Session 2).

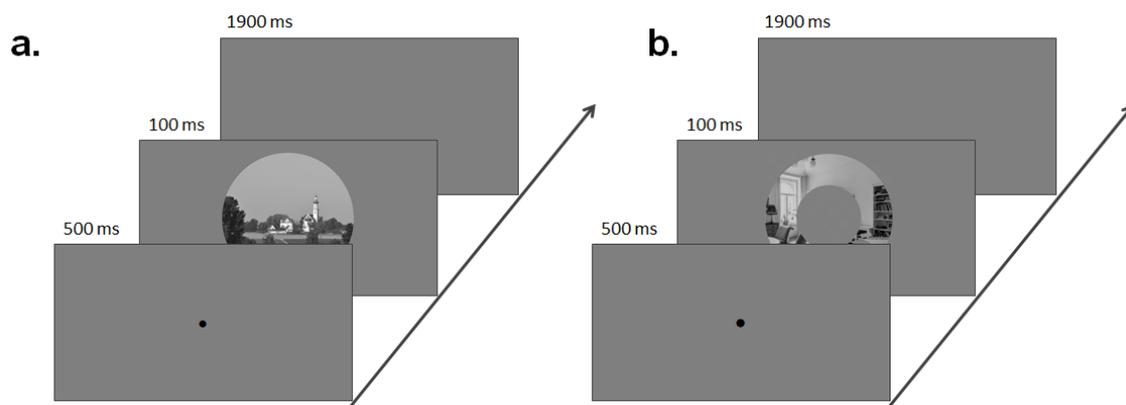


Figure 3.4 : Schéma de la procédure expérimentale de l'**Expérience 4**. Les stimuli étaient présentés pendant 100 ms. Les participants devaient catégoriser **(a)** les Disques entier dans la Session 1 et **(b)** les stimuli de différentes excentricités (Disque central, Anneau 1, Anneau 2, Anneau 3, Anneau 4) dans la Session 2.

1.2. Hypothèses

Nous avons formulé nos hypothèses sur la base des précédents travaux de Larson et Loschky (2009). Pour rappel, dans cette étude, les participants étaient plus performants pour catégoriser des scènes lorsque seule l'information périphérique de la scène était disponible (i.e. lorsque la partie centrale de la scène était cachée par un scotome circulaire), que lorsque seule la partie centrale de la scène était disponible. Sur la base de ces résultats, nous nous attendions à ce que les performances de catégorisation soient meilleures pour les stimuli périphériques (i.e. à ce que les pourcentages d'erreurs et les temps de réponses diminuent à mesure que l'excentricité des stimuli périphériques augmentent). Néanmoins, dans l'expérience de Larson et Loschky (2009), il est possible que les stimuli qui ne présentaient que la portion centrale d'une scène ne contenaient pas suffisamment d'information sémantique pertinente pour permettre une catégorisation efficace, ce qui pourrait expliquer les performances moindres pour la catégorisation de stimuli centraux relativement aux stimuli périphériques. C'est pourquoi, dans notre expérience, nous avons décidé de manipuler le contenu sémantique de la partie centrale des scènes en présentant ou non un objet sémantiquement relié à la catégorie de la scène. Nous nous attendions à ce que la présence d'un objet central influence les différences de performances entre la catégorisation de l'information présente dans le Disque central et celle de l'information présente dans les Anneaux périphériques. En d'autres termes, nous nous attendions à une interaction entre l'excentricité de l'information visuelle disponible (centrale vs. périphérique) et la présence d'un objet en vision centrale.

Par ailleurs, si la présence d'un objet pertinent en vision centrale améliore à lui seul la catégorisation de la scène, cette variable devrait également influencer les performances de catégorisation lorsque les scènes sont vues dans leur intégralité. Nous avons donc également testé cette nouvelle hypothèse à travers la Session 1 où les participants catégorisaient seulement les scènes entières (ou Disques entiers). Nous nous attendions à de meilleures performances

pour la catégorisation des scènes entières qui contenaient un objet en vision centrale que pour la catégorisation des scènes entières qui n'en contenaient pas.

1.3. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale de chaque session, les pourcentages d'erreurs moyens (%ERm) et les temps de réponses correctes moyens en millisecondes (TRm) ont été calculés. Afin de réduire l'effet des valeurs extrêmes sur les résultats, les temps de réponses atypiques pour chaque participant dans chaque condition ont été supprimés à l'aide de la méthode des boîtes à moustaches. Cette méthode permet de mettre en évidence pour chaque participant et dans chaque condition les données supérieures ou inférieures à 3 fois l'intervalle interquartile. A l'aide de cette méthode, nous avons seulement supprimé 1,86% des temps de réponses pour la Session 1 et 3,30% des temps de réponse pour la Session 2. L'analyse des données a été effectuée à l'aide du logiciel Statistica 13.3. Le seuil de significativité a été fixé à 0,05. La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2).

1.3.1. Session 1 : Catégorisation des Disques entiers

Pour l'analyse de la Session 1, une analyse de la variance à mesure répétées (ANOVA) de type 2×2 a été réalisée sur les valeurs de %ERm et les TRm avec comme facteur intra-sujet la Présence d'un objet central (Objet central présent ; Objet central absent) et la Catégorie de la scène (Intérieur ; Extérieur). Les pourcentages d'erreurs moyens (%ERm) et leurs erreurs-types ainsi que les temps de réponses moyens (TRm) et leurs erreurs-types sont présentés dans le Tableau 3.1.

Tableau 3.1 : Résultats de la Session 1 de l'Expérience 4. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Présence d'un objet central (Objet central présent ; Objet central absent) et de la Catégorie de la scène (Intérieur ; Extérieur).

	Objet central présent		Objet central absent	
	%ERm	TRm	%ERm	TRm
Intérieur	3,20 ± 0,59	622 ± 14	3,07 ± 0,87	622 ± 12
Extérieur	3,60 ± 0,93	623 ± 13	3,73 ± 0,92	636 ± 15

L'analyse de la variance (ANOVA) menée sur les %ERm et sur les TRm n'a révélé aucun résultat significatif. Quelle que soit la condition expérimentale, les taux d'erreurs étaient bas (inférieurs à 4% ; Figure 3.5). L'effet principal de la Présence d'un objet central n'était pas significatif, ni pour les %ERm, (Objet central présent : $3,40 \pm 0,90\%$; Objet central absent :

$3,40 \pm 0,96\%$; $F(1,29) < 1$), ni pour les TRm, (Objet central présent : 624 ± 18 ms ; Objet central absent : 631 ± 18 ms ; $F(1,29) = 3,04$, $p = .09$, $\eta^2 = .09$). De plus, la Catégorie de la scène n'interagissait pas avec la Présence d'un objet central, ni pour les %ERM, ($F(1,29) < 1$, $p = .847$, $\eta^2 = .09$), ni pour les TRm, ($F(1,29) < 1$, $p = .460$, $\eta^2 = .02$). Ces résultats suggèrent que la seule présence d'un objet sémantiquement relié à la catégorie de la scène en vision centrale n'améliore pas les performances de catégorisation des scènes, bien que cette interprétation puisse être limitée par un potentiel effet plafond.

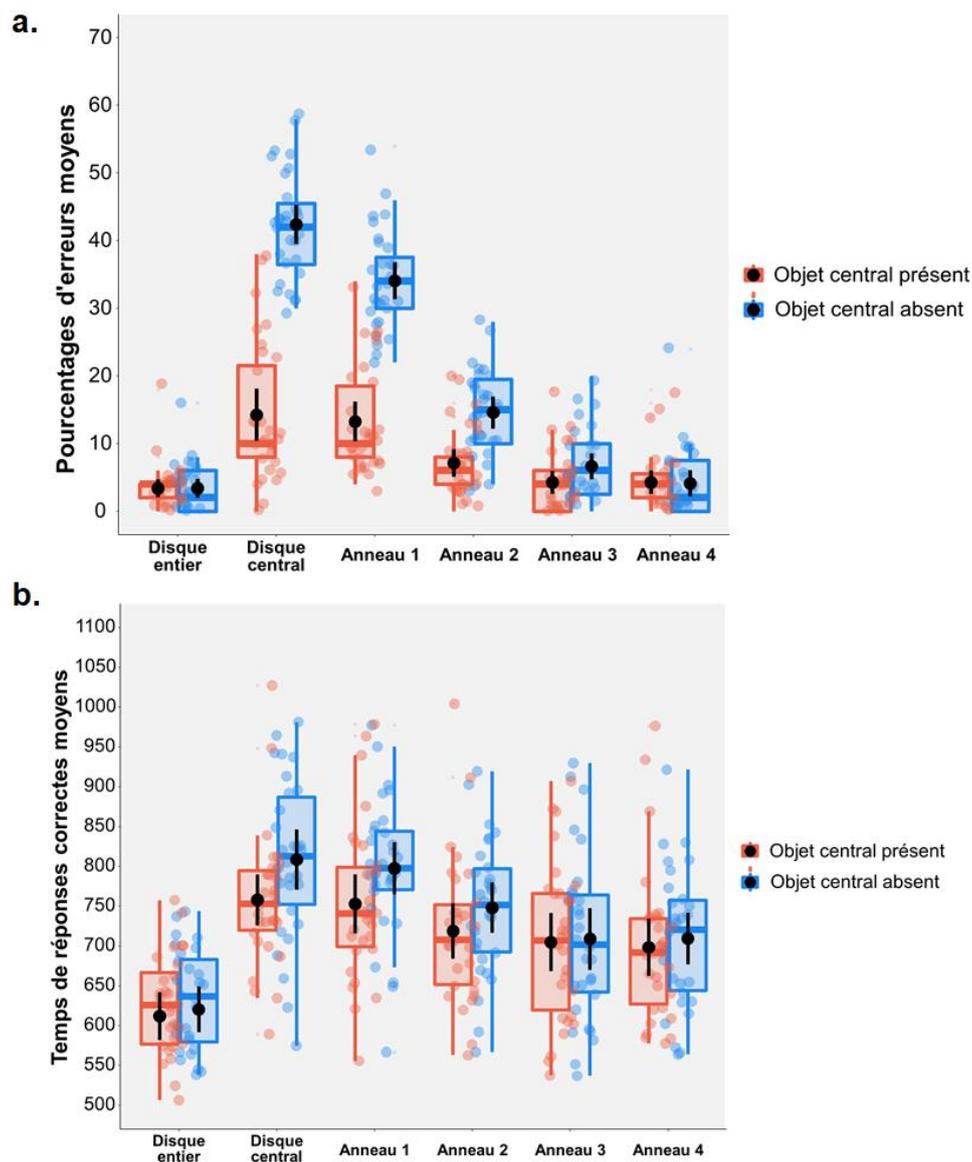


Figure 3.5 : Résultats de la Session 1 (catégorisation du Disque entier) et de la Session 2 (catégorisation du Disque central, de l'Anneau 1, de l'Anneau 2, de l'Anneau 3 et de l'Anneau 4) de l'Expérience 4 : (a) Pourcentages d'erreurs moyens et (b) Temps de réponses correctes moyens en ms en fonction de l'Excentricité des stimuli (Disque entier ; Disque central ; Anneau 1 ; Anneau 2 ; Anneau 3 ; Anneau 4) et de la Présence d'un objet central (Objet central présent ; Objet central absent). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

1.3.2. Session 2 : Catégorisation des stimuli de différentes excentricités rétinienne (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4)

Pour la Session 2, une analyse de la variance à mesure répétées (ANOVA) de type $5 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm avec comme facteur intra-sujet l'Excentricité (Disque central ; Anneau 1 ; Anneau 2 ; Anneau 3 ; Anneau 4), la Présence d'un objet central (Objet central présent ; Objet central absent) et la Catégorie de la scène (Intérieur ; Extérieur). Les pourcentages d'erreurs moyens (%ERm) et leurs erreurs-types ainsi que les temps de réponses moyens (TRm) et leurs erreurs-types sont présentés dans le Tableau 3.2.

Tableau 3.2 : Résultats de la Session 2 de l'Expérience 4. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de l'Excentricité (Disque central ; Anneau 1 ; Anneau 2 ; Anneau 3 ; Anneau 4), de la Présence d'un objet central (Objet central présent ; Objet central absent) et de la Catégorie de la scène (Intérieur ; Extérieur).

		Objet central présent		Objet central absent	
		%ERm	TRm	%ERm	TRm
Intérieur	Disque central	14,13 ± 2,41	733 ± 14	53,33 ± 3,37	813 ± 13
	Anneau 1	18,00 ± 1,96	731 ± 16	45,07 ± 3,17	807 ± 18
	Anneau 2	8,00 ± 1,96	709 ± 16	16,80 ± 1,62	736 ± 15
	Anneau 3	4,80 ± 1,24	683 ± 16	6,13 ± 1,21	679 ± 16
	Anneau 4	4,80 ± 1,08	671 ± 15	4,27 ± 1,10	678 ± 19
Extérieur	Disque central	14,40 ± 2,31	732 ± 14	31,47 ± 3,33	743 ± 13
	Anneau 1	8,53 ± 1,29	723 ± 16	23,07 ± 2,35	740 ± 16
	Anneau 2	6,27 ± 1,21	690 ± 15	12,40 ± 1,47	708 ± 14
	Anneau 3	3,73 ± 0,89	677 ± 16	7,20 ± 1,14	683 ± 17
	Anneau 4	3,73 ± 0,98	682 ± 17	4,00 ± 1,07	696 ± 17

- Analyse des pourcentages d'erreurs moyens

L'analyse de la variance (ANOVA) menée sur les %ERm a révélé un effet principal de l'Excentricité ($F(4,116) = 287,92, p < .001, \eta^2 = .91$). Un test de contraste de linéarité a été effectué entre les différentes excentricités (i.e. Disque central, Anneau 1, Anneau 2, Anneau 3, Anneau 4). Cette analyse a révélé une diminution linéaire des taux d'erreurs à mesure que l'excentricité augmentait ($F(1,29) = 520,70, p < .001$; Disque central : $28,33 \pm 2,73\%$; Anneau 1 : $23,67 \pm 2,58\%$; Anneau 2 : $10,87 \pm 1,89\%$; Anneau 3 : $5,47 \pm 1,60\%$; Anneau 4 : $4,20 \pm 1,66\%$; Figure 3.5a). L'effet principal de la Présence d'un objet central était également significatif ($F(1,29) = 359,44, p < .001, \eta^2 = .93$). Les participants faisaient en moyenne moins d'erreurs pour catégoriser une scène qui contenait un objet en vision centrale ($8,64 \pm 3,22\%$) par rapport à une scène que n'en contenait pas ($20,37 \pm 2,81\%$). La Présence d'un objet central interagissait avec l'Excentricité ($F(4,116) = 122,67, p < .001, \eta^2 = .81$). Les comparaisons planifiées ont montré une différence significative entre le Disque central et l'Anneau 4, à la fois quand il n'y avait pas d'objet central ($F(1,29) = 693,18, p < .001$) et quand il y en avait un

($F(1,29) = 39.98, p < .001$). Toutefois, l'interaction suggère que cette différence était plus grande lorsqu'il n'y avait pas d'objet central (Disque central : $42,40 \pm 1,2 \%$; Anneau 4 : $4,13 \pm 1,33 \%$), que lorsqu'il y en avait un (Disque central : $14,27 \pm 2,67 \%$; Anneau 4 : $4,27 \pm 1,36$; $F(1,29) = 215,22, p < .001$). De plus, les comparaisons planifiées ont montré que les participants faisaient moins d'erreurs pour catégoriser le Disque central qui contenait un objet ($14,27 \pm 2,67\%$; $F(1,29) = 217,95, p < .001$) par rapport au Disque central qui n'en contenait pas ($42,40 \pm 2,00\%$). L'ensemble de ces résultats suggèrent que malgré la présence d'un objet central sémantiquement relié à la catégorie de la scène qui facilite les performances de catégorisation, les participants faisaient toujours moins d'erreurs lorsque l'information était disponible en vision périphérique qu'en vision centrale.

Par ailleurs, l'effet principal de la Catégorie était significatif ($F(1,29) = 24,34, p < .001, \eta^2 = .46$). Les participants faisaient moins d'erreurs pour catégoriser des scènes d'extérieur ($11,48 \pm 3,29\%$), que des scènes d'intérieur ($17,53 \pm 3,62\%$). L'analyse montrait une interaction significative entre la Catégorie et l'Excentricité ($F(4,116) = 13,12, p < .001, \eta^2 = .31$) et ces deux facteurs interagissaient avec la Présence d'un objet central ($F(4,116) = 10,37, p < .001, \eta^2 = .26$). Spécifiquement, l'interaction entre les deux extrêmes excentricités (Disque central et Anneau 4) et la Présence d'un objet central était significative, à la fois pour les scènes d'intérieur ($F(1,29) = 138,49, p < .001$) et pour les scènes d'extérieur ($F(1,29) = 22,68, p < .001$). Pour les scènes d'intérieur, les participants faisaient moins d'erreurs pour catégoriser l'Anneau 4 que le Disque central, à la fois lorsqu'il y avait un objet central (Anneau 4 : $4,80 \pm 1,06\%$; Disque central : $14,13 \pm 2,39\%$; $F(1,29) = 18,45, p < .001$) et pas d'objet central (Anneau 4 : $4,27 \pm 1,11\%$; Disque central : $53,33 \pm 3,31\%$; $F(1,29) = 178,38, p < .001$). Cependant, cette différence était plus grande lorsqu'il n'y avait pas d'objet central que lorsqu'il y en avait un. Pour les scènes d'extérieur, les participants faisaient moins d'erreurs pour catégoriser l'Anneau 4 que le Disque central à la fois lorsqu'il y avait un objet central (Anneau 4 : $3,73 \pm 0,98\%$; Disque central : $14,40 \pm 2,30\%$; $F(1,29) = 20,85, p < .001$) et pas d'objet central (Anneau 4 : $4,00 \pm 1,05\%$; Disque central : $31,47 \pm 3,27\%$; $F(1,29) = 84,84, p < .001$). A nouveau, l'interaction suggère que cette différence était plus grande lorsqu'il n'y avait pas d'objet central que lorsqu'il y avait un objet central. Ainsi, la présence d'un objet central facilitait la catégorisation, mais les participants demeuraient toujours meilleurs lorsque l'information était disponible en vision périphérique qu'en vision centrale pour les deux types de catégorie.

Nous avons ensuite testé comment la Catégorie interagissait avec la Présence d'un objet central indépendamment pour chacune des deux extrêmes excentricités (Disque central et Anneau 4). Spécifiquement, l'interaction entre la Présence d'un objet central et la Catégorie était significative pour le Disque central ($F(1,29) = 13,19, p < .001$), mais pas pour l'Anneau 4 ($F(1,29) < 1$). Les participants faisaient plus d'erreurs pour catégoriser un Disque central appartenant à la catégorie Intérieur qu'un Disque central appartenant à la catégorie Extérieur lorsqu'il n'y avait pas d'objet central (Intérieur : $53,33 \pm 3,31\%$; Extérieur : $31,47 \pm 3,27\%$; $F(1,29) = 13,50, p < .001$), mais pas lorsqu'il y avait un objet central. Dans ce cas, les performances de catégorisation du Disque central ne différaient pas entre les scènes d'intérieur et les scènes d'extérieur (Extérieur : $14,40 \pm 2,30\%$; Intérieur : $14,13 \pm 2,39\%$; $F(1,29) < 1, p$

= .925). Ces résultats suggèrent que l'absence d'un objet central désavantagerait plus la catégorisation des scènes d'intérieur que la catégorisation des scènes d'extérieur.

Enfin, nous avons comparé les performances entre chaque condition de la Session 2 (i.e. les différentes excentricités et la présence d'un objet central) aux performances obtenues dans la Session 1 (Disque entier). Plus précisément, lorsqu'il n'y avait pas d'objet central, ces comparaisons montraient que seuls les taux d'erreurs de l'Anneau 4 n'étaient pas significativement différents de ceux observés pour le Disque entier ($t(29) = 0,85, p = .401$). Les performances des autres excentricités étaient toutes significativement différentes de celles observées pour le Disque Entier (Disque central : $t(29) = 25,00, p < .001$; Anneau 1 : $t(29) = 19,07, p < .001$; Anneau 2 : $t(29) = 9,92, p < .001$; Anneau 3 : $t(29) = 3,21, p = .003$). Lorsqu'il y avait un objet central, les performances de catégorisation des Anneaux 3 et 4 n'étaient pas significativement différentes de celles observées pour le Disque entier (Anneau 3 : $t(29) = 1,11, p = .277$; Anneau 4 : $t(29) = 1,03, p < .314$). Les performances des autres excentricités étaient significativement différentes de celles observées pour le Disque entier (Disque central : $t(29) = 5,79, p < .001$; Anneau 1 : $t(29) = 6,91, p < .001$; Anneau 2 : $t(29) = 3,33, p = .002$). Ces résultats suggèrent que la catégorisation de scènes n'est pas défavorisée lorsque seule l'information visuelle périphérique est disponible.

- **Analyse des temps de réponses correctes moyens**

L'analyse de la variance (ANOVA) menée sur les TRm a révélé un effet principal de l'Excentricité ($F(4,116) = 62,95, p < .001, \eta^2 = .68$). Similairement à l'analyse des %ERm, le test de contraste de linéarité effectué entre les différentes excentricités a révélé une diminution linéaire des temps de réponses à mesure que l'excentricité augmentait ($F(1,29) = 102,81, p < .001$; Disque central : 758 ± 27 ms ; Anneau 1 : 755 ± 30 ms ; Anneau 2 : 715 ± 29 ms ; Anneau 3 : 686 ± 32 ms ; Anneau 4 : 686 ± 31 ms ; Figure 3.5b). L'effet principal de la Présence d'un objet central était également significatif ($F(1,29) = 32,12, p < .001, \eta^2 = .53$). Les participants étaient plus rapides pour catégoriser une scène qui contenait un objet en vision centrale (708 ± 47 ms) par rapport à une scène qui n'en contenait pas (732 ± 45 ms). La Présence d'un objet central interagissait de nouveau avec l'Excentricité ($F(4,116) = 8,04, p < .001, \eta^2 = .22$). En considérant uniquement le Disque central, les participants catégorisaient les scènes avec un objet central plus rapidement (736 ± 19 ms ; $F(1,29) = 22,25, p < .001$) que les scènes qui ne contenaient pas d'objet central (781 ± 22 ms). En considérant cette fois-ci uniquement l'Anneau 4, les participants catégorisaient les scènes avec un objet central aussi rapidement que les scènes sans objet central ($F(1,29) = 2,35, p = .136$). Les comparaisons planifiées ont également montré une différence significative entre le Disque central et l'Anneau 4, à la fois lorsqu'il n'y avait pas d'objet central ($F(1,29) = 58,25, p < .001$) et lorsqu'il y en avait un ($F(1,29) = 55,12, p < .001$). Toutefois, l'interaction suggère que la différence entre le Disque central et l'Anneau 4 était plus grande lorsqu'il n'y avait pas d'objet central (Disque central : 781 ± 22 ms ; Anneau 4 : 691 ± 21 ms), que lorsqu'il y en avait un (Disque central : 736 ± 19 ms ; Anneau 4 : 682 ± 23 ms ; $F(1,29) = 17,19, p < .001$). Ces résultats suggèrent, comme pour les %ERm, que malgré la présence d'un objet sémantiquement relié à la catégorie de la scène en vision centrale qui

facilite la catégorisation, les participants demeuraient plus rapides lorsque l'information était disponible en vision périphérique qu'en vision centrale.

Par ailleurs, l'effet principal de la Catégorie était à nouveau significatif ($F(1,29) = 17,71, p < .001, \eta^2 = .38$). Les participants étaient plus rapides lorsqu'ils devaient catégoriser des scènes d'extérieur (711 ± 46 ms), que lorsqu'ils devaient catégoriser des scènes d'intérieur (729 ± 47 ms). Encore une fois, la Catégorie interagissait avec l'Excentricité ($F(4,116) = 11,22, p < .001, \eta^2 = .28$) et ces deux facteurs interagissaient avec la Présence d'un objet central ($F(4,116) = 11,16, p < .001, \eta^2 = .28$). Spécifiquement, l'interaction entre les deux extrêmes excentricités (Disque central et Anneau 4) et la Présence d'un objet central était significative pour la catégorie Intérieur ($F(1,29) = 29,66, p < .001$), mais pas pour la catégorie Extérieur ($F(1,29) < 1$). Pour les scènes d'Intérieur, les participants étaient plus rapides pour catégoriser l'Anneau 4 que le Disque central à la fois lorsqu'il y avait un objet central (Anneau 4 : 678 ± 16 ms ; Disque central : 738 ± 15 ms ; $F(1,29) = 35,40, p < .001$) que lorsqu'il n'y en avait pas (Anneau 4 : 682 ± 14 ms ; Disque central : 817 ± 19 ms ; $F(1,29) = 68,61, p < .001$). Toutefois, l'interaction suggère que cette différence était plus grande lorsqu'il n'y avait pas d'objet central que lorsqu'il y en avait un. Ainsi, la présence d'un objet central facilitait spécifiquement la catégorisation des scènes d'intérieur, mais les participants étaient toujours plus rapides lorsque l'information était disponible en vision périphérique qu'en vision centrale pour les deux types de catégorie.

Comme pour l'analyse des %ERm, nous avons testé comment la Catégorie interagissait avec la Présence d'un objet central pour chacune des deux extrêmes excentricités (Disque central et Anneau 4). L'interaction entre la Présence de l'objet central et la Catégorie était significative pour le Disque central ($F(1,29) = 24,37, p < .001$), mais pas pour l'Anneau 4 ($F(1,29) = 1,45, p = .239$). Les participants étaient plus lents pour catégoriser les scènes d'intérieur que les scènes d'extérieur lorsqu'il n'y avait pas d'objet central (Intérieur : 817 ± 19 ms ; Extérieur : 745 ± 13 ms ; $F(1,29) = 34,27, p < .001$). Toutefois, lorsqu'il y avait un objet central, les temps de réponses ne différaient pas entre les scènes d'intérieur et les scènes d'extérieur (Extérieur : 734 ± 14 ms ; Intérieur : 738 ± 15 ms ; $F(1,29) < 1$). Dans le sens de l'analyse des %ERm, ces résultats suggèrent que l'absence d'un objet central désavantagerait plus la catégorisation des scènes d'intérieur que la catégorisation des scènes d'extérieur.

Pour finir, nous avons comparé les TRm entre chaque condition (i.e. les différentes excentricités et la présence d'un objet central) avec les performances de la Session 1 (Disque entier). Ces comparaisons montraient que les temps de réponses de toutes les excentricités étaient significativement différents de ceux obtenus de la Session 1, à la fois lorsqu'il n'y avait pas d'objet central (Disque central : $t(29) = 9,53, p < .001$; Anneau 1 : $t(29) = 9,79, p < .001$; Anneau 2 : $t(29) = 8,22, p < .001$; Anneau 3 : $t(29) = 4,30, p = .003$; Anneau 4 : $t(29) = 5,11, p < .001$) que lorsqu'il y en avait un (Disque central : $t(29) = 9,69, p < .001$; Anneau 1 : $t(29) = 8,59, p < .001$; Anneau 2 : $t(29) = 6,77, p = .001$; Anneau 3 : $t(29) = 4,93, p < .001$; Anneau 4 : $t(29) = 4,72, p < .001$). Ainsi, différemment des %ERm, les temps de réponses pour catégoriser des scènes étaient systématiquement plus longs lorsque seulement une portion de la scène était présentée.

1.4. Discussion

Les résultats de l'**Expérience 4** ont montré que les performances de catégorisation des participants s'amélioraient (i.e. que les pourcentages d'erreurs moyens et les temps de réponses correctes moyens diminuaient) à mesure que l'information périphérique était découverte. Nous avons également observé que les performances de catégorisation du stimulus le plus périphérique (Anneau 4 de la Session 2) ne différaient pas de celles obtenues lors de la catégorisation de la scène entière (Disque entier de la Session 1). Ces résultats sont cohérents avec les travaux précédents menés par Larson et Loschky (2009) et indiquent que la vision périphérique jouerait un rôle plus important que la vision centrale pour catégoriser rapidement une scène.

Nous avons spécifiquement testé si la supériorité de la vision périphérique sur la vision centrale observée par Larson et Loschky (2009) pouvait s'expliquer par l'absence d'information pertinente en vision centrale. Nous avons donc utilisé des scènes qui contenaient soit un objet sémantiquement relié à la catégorie de la scène en vision centrale, soit aucun objet. Les résultats de la Session 1 (catégorisation des Disques entiers) n'ont pas montré d'effet de la présence d'un objet central sur les performances de catégorisation des participants. Ainsi, la seule présence d'un objet sémantiquement pertinent centre de la scène ne permettrait pas d'améliorer les performances de catégorisation des scènes. Cela pourrait s'expliquer par le fait que des objets présents en vision centrale, même pertinents, seraient peu considérées pour catégoriser rapidement une scène visuelle. Dans ce cas, le système visuel se baserait en priorité sur les informations en provenance de la périphérie qui seraient suffisantes pour catégoriser des scènes d'intérieur ou d'extérieur. Cette hypothèse est cohérente avec les résultats obtenus par Larson et Loschky (2009) montrant que les performances de catégorisation sont aussi bonnes lorsque la scène est entièrement visible que et lorsque seule l'information périphérique est disponible. Nous pouvons toutefois noter que les taux d'erreurs étaient extrêmement bas dans notre expérience (inférieurs à 4%). Cet effet plafond peut amener à penser que la tâche de catégorisation était trop facile dans la Session 1, augmentant alors la difficulté à détecter statistiquement un effet significatif. En effet, dans la Session 2, les taux d'erreurs obtenus dans la condition de catégorisation du Disque central était globalement significativement plus importants que ceux obtenus dans la condition de catégorisation de la scène entière (Disque entier de la Session 1) et cette fois, dans ces conditions plus difficiles, la présence d'un objet sémantiquement relié à la catégorie de la scène améliorait significativement les performances de catégorisation du Disque central.

Plus important, dans la Session 2, nous avons observé que la différence des performances de catégorisation entre le Disque central et l'anneau le plus périphérique (Anneau 4) était plus faible lorsqu'il y avait un objet central que lorsqu'il n'y avait pas d'objet central. Mais malgré l'amélioration des performances lorsqu'un objet sémantiquement relié était présent en vision centrale, les performances observées pour l'Anneau 4 demeuraient meilleures que celles observées pour le Disque central. Ces résultats suggèrent donc que la catégorisation d'une scène en vision centrale se base en partie sur la présence d'un contenu visuel spécifique, tel qu'un objet sémantiquement pertinent, pour catégoriser une scène visuelle et ils confortent

également les résultats de Larson et Loschky (2009) en démontrant la supériorité de la vision périphérique sur la vision centrale indépendamment de leurs contenus informatifs respectifs.

Dans l'ensemble, ces résultats vont dans le sens d'une plus grande efficacité et utilité de la vision périphérique par rapport à la vision centrale lors de la catégorisation de scènes visuelles (en tant que scènes d'intérieur et d'extérieur), et cela malgré la faible acuité visuelle inhérente à la vision périphérique. La catégorisation de scènes ne nécessiterait pas une analyse fine et détaillée des différents éléments qui la compose. Une information de basse résolution spatiale (ou BFS) extraite en vision périphérique serait suffisante à la catégorisation de scènes (Boucart et al., 2013 ; Musel et al., 2011 ; Peyrin et al., 2017).

Cette expérience nous a donc permis de confirmer que le traitement de l'information de basse résolution spatiale (BFS) disponible en vision périphérique serait privilégié lors de la catégorisation rapide de scènes. Dans la continuité de ce travail, notre problématique de recherche était plus particulièrement de déterminer si l'analyse rapide des informations disponibles en vision périphérique pourrait permettre une première catégorisation qui serait validée ou non par l'analyse des informations disponibles en vision centrale (principalement une information détaillée en HFS). En d'autres termes, est-ce que la catégorisation de scènes suivrait un traitement *peripheral-to-central* ? Nous avons testé cette hypothèse dans l'**Expérience 5** de cette thèse. Pour cela, nous avons construit des scènes dynamiques inspirées des précédents travaux Carole Peyrin et collaborateurs (Musel et al., 2012 ; Musel et al., 2014 ; Kauffmann, Chauvin, Pichat & Peyrin, 2015 ; Kauffmann, Chauvin, Guyader & Peyrin, 2015). Nous avons imposé un traitement *peripheral-to-central* ou *central-to-peripheral* en assemblant les stimuli utilisés dans la Session 2 de l'**Expérience 4** en allant du stimulus le plus périphérique au stimulus central ou en allant du stimulus central au stimulus le plus périphérique, respectivement. En utilisant ces stimuli, nous avons pu à nouveau considérer l'influence de la présence d'un contenu sémantique pertinent en vision centrale, mais aussi contrôler le facteur de magnification corticale. Dans cette expérience, nous avons également manipulé le temps de présentation des séquences afin de tester l'hypothèse que le traitement *peripheral-to-central* serait d'autant plus privilégié que le système visuel dispose de peu de temps pour traiter l'information. Les séquences étaient présentées soit pendant 500 ms (chaque stimulus dans la séquence étant présenté pendant 100 ms, comme dans l'**Expérience 4**), soit 100 ms (chaque stimulus était présenté pendant 20 ms de façon à ce que la séquence totale ait la même durée de présentation qu'un seul stimulus de l'**Expérience 4**).

2. **Expérience 5 - Catégorisation de larges scènes dynamiques en expansion ou en contraction**

2.1. Méthode

2.1.1. Participants

Vingt-huit étudiants de l'Université Grenoble Alpes (24 femmes ; âge moyen : $19,66 \pm 1,15$) ont participé à cette expérience. Ils ont été répartis en deux groupes ($n = 14$) qui devaient respectivement catégoriser les séquences présentées pendant 100 ms ou pendant 500 ms. Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Tous les participants ont donné leur consentement libre et éclairé avant de participer à cette expérience qui a été menée en conformité avec le code d'éthique de l'*American Psychological Association* (APA) et approuvée par un comité et approuvée par un comité d'éthique (CERGA, Comité d'éthique de la recherche Grenoble Alpes, IRB00010290-2018-10-16-50). A la fin de l'expérience, chaque participant recevait un bon de 0,25 point par quart d'heure d'expérience à valoir lors des examens universitaires.

2.1.2. Stimuli et procédure

Les stimuli qui composaient les séquences dynamiques dans cette expérience étaient les mêmes que ceux utilisés dans la Session 2 de l'**Expérience 4** (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4). Ainsi, nous avons utilisé 100 stimuli (50 scènes d'intérieur et 50 scènes d'extérieur). A partir du Disque central et des quatre anneaux de chaque scène, nous avons créé deux types de séquences (Figure 3.6) : séquences *peripheral-to-central* et séquences *central-to-peripheral*. Pour les séquences *peripheral-to-central*, les cinq stimuli étaient assemblés en allant de l'information périphérique à l'information centrale (dans l'ordre : l'Anneau 4, l'Anneau 3, l'Anneau 2, l'Anneau 1, puis le Disque central). Pour les séquences *central-to-peripheral*, les cinq stimuli étaient assemblés en allant de l'information centrale à l'information périphérique (dans l'ordre : le Disque central, puis l'Anneau 1, l'Anneau 2, l'Anneau 3 et l'Anneau 4). L'expérience a été programmée avec le logiciel E-Prime 2.0 (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) et les stimuli étaient présentés sur le même écran de 30 pouces que dans l'**Expérience 4**. Afin de respecter la taille angulaire des stimuli à 30° d'angle visuel, les participants étaient installés à une distance de 70 cm de l'écran, le menton positionné sur une mentonnière.

L'expérience a été réalisée dans un box expérimental du Laboratoire de Psychologie et NeuroCognition (LPNC) de l'Université de Grenoble Alpes, dans les mêmes conditions expérimentales que l'**Expérience 4**. Les participants étaient répartis dans deux groupes ($n = 14$ dans chaque groupe) qui devaient respectivement catégoriser les séquences présentées pendant 100 ms (chaque stimulus étant présenté pendant 20 ms de façon à ce que la durée totale de la séquence soit égale au temps de présentation d'un seul stimulus de l'**Expérience 4**) ou pendant

500 ms (chaque stimulus étant présenté pendant 100 ms, comme chaque stimulus de l'**Expérience 4**). Les séquences *peripheral-to-central* et *central-to-peripheral* étaient présentées aléatoirement. Chaque essai commençait par un point de fixation central (noir sur un fond gris de luminance 0,5) présenté pendant 500 ms, suivi d'une séquence de cinq stimuli (séquence *peripheral-to-central* ou *central-to-peripheral*) présentée pendant 500 ms ou 100 ms selon le groupe expérimental, immédiatement suivi d'un masque présenté pendant 20 ms, puis d'un écran gris présenté pendant 1900 ms au cours duquel les participants pouvaient répondre. Le masque représentait un bruit dont le spectre d'amplitude décroît avec la fréquence spatiale (bruit distribué en 1/F). Il était présenté à travers un large disque d'un rayon de 15° d'angle visuel (750 pixels) afin d'empêcher la persistance rétinienne du dernier stimulus de la séquence. Quel que soit le groupe expérimental, les participants devaient catégoriser la séquence comme étant soit une scène d'intérieur, soit une scène d'extérieur par l'intermédiaire des touches 1 et 2 du clavier d'ordinateur placé sur l'axe sagittal médian des participants. La moitié des participants de chaque groupe appuyait sur la touche 1 avec l'index de la main droite pour la catégorie Intérieur et sur la touche 2 avec le majeur de la main droite pour la catégorie Extérieur. Les touches de réponse étaient contrebalancées pour l'autre moitié des participants. Les participants avaient pour consigne de répondre le plus correctement et rapidement possible, dès l'apparition du premier stimulus de la séquence. L'expérience comprenait 200 essais (pour chaque type de séquence *peripheral-to-central* et *central-to-peripheral* : 25 scènes d'intérieur et 25 scènes d'extérieur avec un objet au centre, 25 scènes d'intérieur et 25 scènes d'extérieur sans objet au centre). L'expérience durait environ 10 minutes pour les séquences présentées pendant 500 ms et durait environ 8 minutes pour les séquences présentées pendant 100 ms, avec une pause au bout de 100 essais. Pour chaque essai, l'exactitude de la réponse et le temps de réponse (en millisecondes) étaient enregistrés. Les participants étaient entraînés au préalable avec des stimuli construits à partir de photographies non utilisées lors de l'expérience (20 essais).

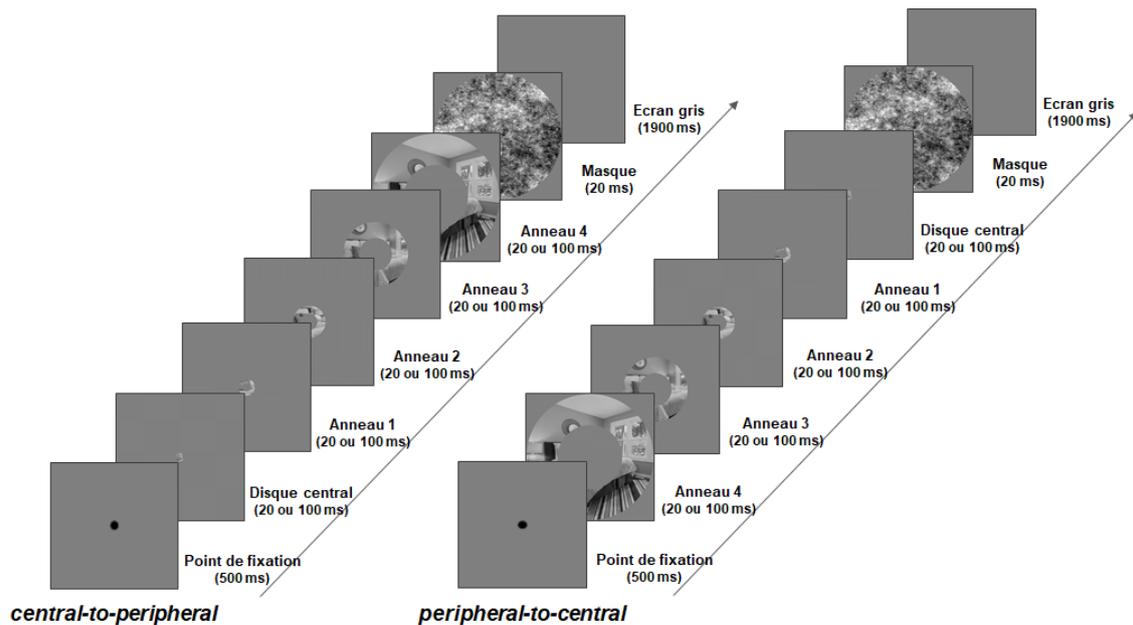


Figure 3.6 : Schéma de la procédure expérimentale de l'**Expérience 5**. Le Disque central, l'Anneau 1, l'Anneau 2, l'Anneau 3 et l'Anneau 4 sont assemblés du centre vers la périphérie (séquences *central-to-peripheral*) ou de la périphérie vers le centre (séquences *peripheral-to-central*). Les séquences étaient présentées pendant 100 ms (chaque stimulus était présenté pendant 20 ms) ou 500 ms (chaque stimulus était présenté pendant 100 ms). Les participants devaient catégoriser les séquences le plus rapidement et le plus précisément possible, dès l'apparition du premier stimulus de la séquence.

2.2. Hypothèses

Nous nous attendions à ce que les séquences *peripheral-to-central* soient catégorisées plus rapidement et plus précisément que les séquences *central-to-peripheral* dans les deux conditions de temps de présentation. Par ailleurs, en partant de l'hypothèse qu'une stratégie d'analyse *coarse-to-fine* prédomine lors de la catégorisation de scènes, nous nous attendions à ce que l'avantage des séquences *peripheral-to-central* soit plus prononcé pour les séquences présentées pendant 100 ms que pendant 500 ms. En effet, les informations en BFS disponibles en vision périphérique pourraient être davantage prises en compte lors de présentations courtes, tandis que l'analyse plus lente des informations en HFS disponibles en vision centrale pourrait prendre l'avantage lors de présentations plus longues.

Enfin, sur la base des résultats de l'**Expérience 4** indiquant que la présence d'un objet central facilite la catégorisation de scènes, nous nous attendions à ce que la présence d'un objet central facilite la catégorisation des séquences *central-to-peripheral* par rapport aux séquences *peripheral-to-central*. En d'autres termes, nous nous attendions à une interaction entre la séquence (*peripheral-to-central* vs *central-to-peripheral*) et la présence d'un objet central (objet central présent vs. objet central absent).

2.3. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale, les pourcentages d'erreurs moyens (%ERm) et les temps de réponses correctes moyens en millisecondes (TRm) ont été calculés. Afin de réduire l'effet des valeurs extrêmes sur les résultats, les temps de réponses supérieurs ou inférieurs à trois fois l'intervalle interquartile ont été supprimés. A l'aide de cette méthode, nous avons supprimé seulement 1,41% des temps de réponse. L'analyse des données a été effectuée à l'aide du logiciel Statistica 13.3 et du logiciel RStudio. Le seuil de significativité a été fixé à 0,05 pour les effets principaux et les interactions et une correction de Bonferroni a été appliquée afin de corriger les comparaisons multiples par paires ($0.05/2 = 0.02$ pour les 2 tests par paire effectués sur les %ERm ; $0.05/16 = 0.003$ pour les 16 tests par paire effectués sur les TRm). La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2).

Une analyse de la variance à mesure répétées (ANOVA) de type $2 \times 2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm avec comme facteur intra-sujet la Séquence (*peripheral-to-central* ; *central-to-peripheral*), la Présence d'un objet central (Objet central présent ; Objet central absent) et la Catégorie de la scène (Intérieur ; Extérieur) et comme facteur inter-sujet le Temps de présentation des séquences (100 ms ; 500 ms). Les pourcentages d'erreurs moyens (%ERm) et leurs erreurs-types ainsi que les temps de réponses moyens (TRm) et leurs erreurs-types sont présentés dans le Tableau 3.3.

Tableau 3.3 : Résultats de l'Expérience 5. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Séquence (*peripheral-to-central* ; *central-to-peripheral*), de la Présence d'un objet central (Objet central présent ; Objet central absent), du Temps de présentation (100 ms ; 500 ms) et de la Catégorie de la scène (Intérieur ; Extérieur).

			Objet central présent		Objet central absent	
			%ERm	TRm	%ERm	TRm
<i>peripheral-to-central</i>	100 ms	Intérieur	4,67 ± 2,82	745 ± 17	5,33 ± 2,49	722 ± 21
		Extérieur	4,33 ± 2,38	768 ± 23	5,67 ± 2,06	744 ± 22
	500 ms	Intérieur	10,29 ± 2,61	563 ± 16	11,71 ± 2,31	561 ± 20
		Extérieur	11,71 ± 2,20	564 ± 21	5,71 ± 1,91	568 ± 20
<i>central-to-peripheral</i>	100 ms	Intérieur	9,00 ± 3,23	810 ± 25	10,00 ± 4,44	820 ± 18
		Extérieur	6,33 ± 2,93	808 ± 22	8,00 ± 2,80	825 ± 21
	500 ms	Intérieur	10,57 ± 2,99	656 ± 23	13,71 ± 4,11	772 ± 16
		Extérieur	8,57 ± 2,71	671 ± 20	6,86 ± 2,59	743 ± 19

- Analyse des pourcentages d'erreurs moyens

L'analyse de la variance (ANOVA) menée sur les %ERm a révélé, conformément à notre hypothèse, un effet principal de la Séquence ($F(1,24) = 6,68$, $p = .016$, $\eta^2 = .20$). Les participants faisaient moins d'erreurs pour catégoriser des séquences *peripheral-to-central* ($7,75 \pm 2,65$ %) que des séquences *central-to-peripheral* ($9,71 \pm 3,60$ % ; Figure 3.7a). L'effet

principal du Temps de présentation n'était pas significatif ($F(1,24) < 1$), mais ce facteur interagissait avec la Séquence ($F(1,24) = 6,20, p = .019, \eta^2 = .19$). Les comparaisons planifiées ont montré que les participants faisaient en moyenne moins d'erreurs pour catégoriser les séquences *peripheral-to-central* que les séquences *central-to-peripheral* uniquement lorsqu'elles étaient présentées pendant 100 ms (*peripheral-to-central* : $5,64 \pm 3,75\%$; *central-to-peripheral* : $9,50 \pm 5,09\%$; $F(1,24) = 12,87, p < .001$). Par contre, cette différence n'était pas significative lorsque les séquences étaient présentées pendant 500 ms (*peripheral-to-central* : $9,85 \pm 3,75\%$; *central-to-peripheral* : $9,93 \pm 5,09\%$; $F(1,24) < 1$). Ainsi, lorsque les séquences étaient présentées rapidement (100 ms), les performances des participants étaient meilleures quand l'information visuelle était d'abord disponible en vision périphérique qu'en vision centrale.

L'effet principal de la Présence d'un objet central n'était pas significatif ($F(1,24) < 1$). L'interaction entre la Séquence et la Présence d'un objet central n'était pas significative ($F(1,24) = 1,75, p = .198, \eta^2 = .06$) et ces deux facteurs n'interagissaient pas avec le Temps de présentation ($F(1,24) < 1$). Par ailleurs, l'analyse n'a révélé ni d'effet principal de la Catégorie ($F(1,24) = 3,41, p = .076, \eta^2 = .12$), ni d'interaction entre la Catégorie et les autres facteurs manipulés (Séquence \times Catégorie : $F(1,24) = 2,31, p = .141, \eta^2 = .08$; Séquence \times Temps de présentation \times Catégorie : $F(1,24) < 1$; Séquence \times Présence d'un objet central \times Catégorie : $F(1,24) < 1$; Séquence \times Présence d'un objet central \times Temps de présentation \times Catégorie : $F(1,24) < 1$).

- **Analyse des temps de réponses correctes moyens**

L'analyse de la variance (ANOVA) menée sur les TRm a révélé un effet principal significatif du type de Séquence ($F(1,24) = 904,03, p < .001, \eta^2 = .97$). Les participants étaient en moyenne plus rapides pour catégoriser des séquences *peripheral-to-central* (648 ± 26 ms) par rapport aux séquences *central-to-peripheral* (756 ± 26 ms ; Figure 3.7b). Nous avons également observé un effet principal significatif du Temps de présentation ($F(1,24) = 24,85, p < .001, \eta^2 = .49$). Les participants catégorisaient plus rapidement les séquences présentées pendant 500 ms (637 ± 18 ms) que les séquences présentées pendant 100 ms (766 ± 18 ms). Le Temps de présentation interagissait de nouveau avec la Séquence ($F(1,24) = 113,51, p < .001, \eta^2 = .81$). Les comparaisons planifiées ont montré que les participants catégorisaient les séquences *peripheral-to-central* plus rapidement que les séquences *central-to-peripheral* à la fois lorsqu'elles étaient présentées pendant 100 ms (*peripheral-to-central* : 731 ± 36 ms ; *central-to-peripheral* : 801 ± 37 ms ; $F(1,24) = 188,43, p < .001$) et pendant 500 ms (*peripheral-to-central* : 564 ± 36 ms ; *central-to-peripheral* : 710 ± 37 ms ; $F(1,24) = 829,11, p < .001$). Plus précisément, la différence entre les deux types de séquences était plus grande lorsqu'elles étaient présentées pendant 500 ms que pendant 100 ms.

Par ailleurs, l'effet principal de la Présence d'un objet central était significatif ($F(1,24) = 33,00, p < .001, \eta^2 = .56$). Les participants catégorisaient plus rapidement les séquences qui contenaient un objet central (690 ± 28 ms) que les séquences qui n'en contenaient pas (713 ± 25 ms). La présence d'un objet central interagissait également avec la Séquence ($F(1,24) =$

64,15, $p < .001$, $\eta^2 = .71$). Les comparaisons planifiées ont montré que la différence entre les séquences *peripheral-to-central* et les séquences *central-to-peripheral* était significative, à la fois lorsqu'il y avait un objet central (*peripheral-to-central* : 653 ± 18 ms ; *central-to-peripheral* : 727 ± 22 ms, $F(1,24) = 179,61$, $p < .001$) et lorsqu'il n'y en avait pas (*peripheral-to-central* : 642 ± 19 ms ; *central-to-peripheral* : 784 ± 16 ms, $F(1,24) = 655,67$, $p < .001$). Toutefois, la différence entre les deux types de séquences était plus grande lorsqu'il n'y avait pas d'objet central que lorsqu'il y en avait un.

Enfin, l'analyse a révélé une interaction significative entre la Séquence, la Présence d'un objet central et le Temps de présentation ($F(1,24) = 8,96$, $p = .006$, $\eta^2 = .26$). Nous avons testé l'interaction entre la Séquence et la Présence d'un objet central pour chaque Temps de présentation. Cette interaction était significative à la fois pour un temps de présentation de 500 ms ($F(1,24) = 60,53$, $p < .001$) et pour un temps de présentation de 100 ms ($F(1,24) = 12,58$, $p = .001$). Spécifiquement, pour une présentation de 500 ms, les participants catégorisaient plus rapidement les séquences *peripheral-to-central* que les séquences *central-to-peripheral* à la fois lorsqu'il y avait un objet central (*peripheral-to-central* : 563 ± 26 ms ; *central-to-peripheral* : 663 ± 30 ms ; $F(1,24) = 162,59$, $p < .001$) et lorsqu'il n'y en avait pas (*peripheral-to-central* : 565 ± 27 ms ; *central-to-peripheral* : 757 ± 23 ms ; $F(1,24) = 605,43$, $p < .001$). De la même façon, pour une présentation de 100 ms, les participants catégorisaient plus rapidement les séquences *peripheral-to-central* que les séquences *central-to-peripheral* à la fois lorsqu'il y avait un objet central (*peripheral-to-central* : 743 ± 26 ms ; *central-to-peripheral* : 791 ± 30 ms ; $F(1,24) = 38,46$, $p < .001$) et lorsqu'il n'y en avait pas (*peripheral-to-central* : 720 ± 27 ms ; *central-to-peripheral* : 811 ± 23 ms ; $F(1,24) = 134,72$, $p < .001$). Toutefois, pour chaque condition de temps de présentation, cette différence était plus grande lorsqu'il n'y avait pas d'objet central que lorsqu'il y avait un objet central. Ainsi, l'avantage des séquences *peripheral-to-central* était réduit lorsqu'il y avait un objet central et ce, pour chaque condition de temps de présentation. Concernant spécifiquement les séquences *central-to-peripheral*, les participants étaient plus rapides pour catégoriser les séquences qui contenaient un objet central, par rapport aux séquences qui n'en contenaient pas lorsque les séquences étaient présentées pendant 500 ms ($F(1,24) = 120,53$, $p < .001$), mais pas pendant 100 ms ($F(1,24) = 5,00$, $p = .034$). Pour les séquences *peripheral-to-central*, il n'y avait pas d'effet de la Présence d'un objet central, ni pour les séquences présentées pendant 500 ms ($F(1,24) < 1$), ni pour celles présentées pendant 100 ms ($F(1,24) = 8,71$, $p = .007$). Ainsi, la présence d'un objet central améliorait les performances de catégorisation seulement pour les séquences *central-to-peripheral* présentées pendant 500 ms. A noter toutefois que malgré la présence d'un objet sémantiquement relié à la catégorie de la scène, les participants étaient toujours plus rapides lorsque l'information visuelle était disponible d'abord en vision périphérique (*peripheral-to-central*) qu'en vision centrale (*central-to-peripheral*).

Enfin, l'effet principal de la Catégorie n'était pas significatif ($F(1,24) < 1$), mais l'interaction entre la Catégorie et la Séquence était significative ($F(1,24) = 4,47$, $p = .044$, $\eta^2 = .15$). Les comparaisons planifiées montraient que la différence entre les séquences *central-to-peripheral* et les séquences *peripheral-to-central* était significative à la fois pour la catégorie Intérieur (*central-to-peripheral* : 757 ± 20 ms ; *peripheral-to-central* : 641 ± 18 ms ; $F(1,24) = 472,74$, $p < .001$) et pour la catégorie Extérieur (*central-to-peripheral* : 755 ± 20 ms ;

peripheral-to-central : 655 ± 20 ms ; $F(1,24) = 379,01$, $p < .001$). L'interaction suggère que cet avantage des séquences *peripheral-to-central* était plus important pour la catégorie Intérieur que pour la catégorie Extérieur. Par contre, la Catégorie n'interagissait pas avec les autres facteurs manipulés (Séquence \times Catégorie : $F(1,24) = 3,77$, $p = .064$, $\eta^2 = .14$; Séquence \times Temps de présentation \times Catégorie : $F(1,24) < 1$; Séquence \times Présence d'un objet central \times Catégorie : $F(1,24) = 1,90$, $p = .181$, $\eta^2 = .07$; Séquence \times Présence d'un objet central \times Temps de présentation \times Catégorie : $F(1,24) = 3,73$, $p = .065$, $\eta^2 = .13$).

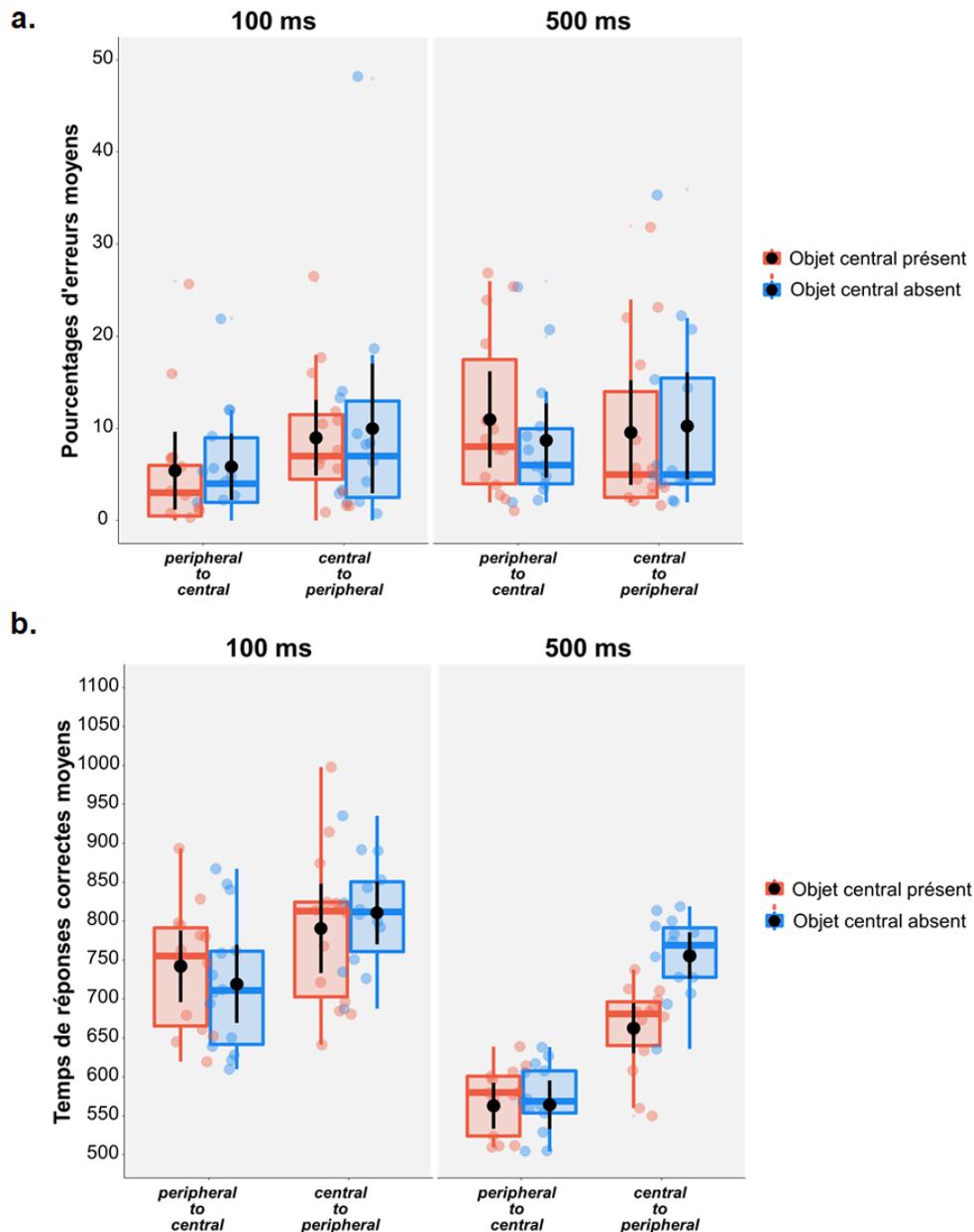


Figure 3.7 : Résultats de l'Expérience 5 : (a) Pourcentages d'erreurs moyens et (b) Temps de réponses correctes moyens en ms en fonction de la Séquence (*peripheral-to-central* ; *central-to-peripheral*), de la Présence d'un objet central (Objet central présent ; Objet central absent) et du Temps de présentation des séquences (100 ms ; 500 ms). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

3. Discussion générale

Les **Expériences 4 et 5** avaient pour objectif d'étudier l'avantage de la vision périphérique sur la vision centrale lors de la catégorisation rapide de scènes. Étant donné que la résolution spatiale diminue avec l'excentricité rétinienne, les informations visuelles périphériques seraient principalement encodées en BFS alors que les informations visuelles centrales seraient principalement encodées en HFS. Sur la base de l'hypothèse de la prédominance d'un traitement *coarse-to-fine* des fréquences spatiales, nous nous sommes alors demandés si l'analyse de scènes suivrait également un traitement *coarse/peripheral-to-central/fine*. Pour cela, nous avons mené deux expériences où nous avons utilisé de grandes photographies de scènes à partir desquelles nous avons construit un disque central et quatre anneaux circulaires révélant différentes régions des scènes, de différentes excentricités rétiniennes. Dans une première expérience (**Expérience 4**), les participants devaient catégoriser les scènes entières (Disque entier) et les stimuli individuels (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4) comme étant des scènes d'intérieur ou d'extérieur. Dans une deuxième expérience (**Expérience 5**), les participants devaient catégoriser des séquences dynamiques composées du disque central et des quatre anneaux. Les stimuli étaient assemblés soit de la vision périphérique à la vision centrale (séquences *peripheral-to-central*), soit de la vision centrale à la vision périphérique (séquences *central-to-peripheral*). Dans les deux expériences, nous avons manipulé le contenu sémantique disponible dans la partie centrale de la scène. Ainsi, le Disque central contenait soit un objet sémantiquement relié à la catégorie de la scène, soit aucun objet. Ces deux expériences ont révélé deux résultats principaux : (1) les performances de catégorisation s'amélioraient à mesure que l'information visuelle périphérique de la scène était dévoilée (**Expérience 4**), (2) les performances de catégorisation étaient meilleures pour la catégorisation des séquences *peripheral-to-central* que pour la catégorisation des séquences *central-to-peripheral* (**Expérience 5**).

Ces deux expériences ont également montré que la présence d'un objet central sémantiquement relié à la catégorie de la scène améliorait significativement les performances de catégorisation du Disque central et des séquences *central-to-peripheral* présentées pendant 500 ms. Dans l'**Expérience 4**, nous avons également observé que l'absence d'un objet au centre de la scène était davantage préjudiciable à la catégorisation des scènes d'intérieur que celle des scènes d'extérieur. Une étude computationnelle de Quattoni et Torralba (2009) suggère que les scènes d'extérieur sont généralement bien caractérisées par des propriétés spatiales globales. Les scènes d'intérieur peuvent également être caractérisées par les propriétés spatiales globales (e.g., une scène avec un couloir), mais la plupart du temps, elles sont mieux caractérisées par les objets qu'elles contiennent (e.g., un livre dans une bibliothèque, un canapé dans un salon). Ainsi, l'absence d'un objet pertinent dans la vision centrale pourrait inciter les participants à considérer par défaut une scène d'intérieur comme une étant une scène d'extérieur. Cependant, même si la catégorisation des scènes pouvait être facilitée par la présence d'un objet en son centre, nous avons systématiquement observé un avantage de la vision périphérique. De plus, dans l'**Expérience 4**, les performances obtenues en vision périphérique (et en particulier celles de la catégorisation de l'Anneau 4 de la Session 2) ne différaient pas des performances obtenues lors de la catégorisation des scènes entières (Disque entier de la Session 1), suggérant que les

seules informations disponibles en vision périphérique pourraient suffire à la catégorisation des scènes. Par conséquent, ces deux expériences soutiennent la plus grande efficacité et plus grande utilité de la vision périphérique par rapport à la vision centrale lors de la catégorisation rapide des scènes (Larson & Loschky, 2009), malgré la faible acuité visuelle inhérente à la vision périphérique et la valeur informative de la vision centrale.

Il faut remarquer que nous avons construit nos stimuli de manière à ce que la surface des informations visuelles révélées par chaque anneau prenne en compte le facteur de magnification corticale (Engel et al., 1994 ; Engel, Glover & Wandell, 1997 ; Daniel & Whitteridge, 1961 ; Virsu, Näsänen & Osmoviita, 1987 ; Virsu & Rovamo, 1979). La largeur et la surface des stimuli de nos deux expériences augmentaient avec l'excentricité et l'anneau le plus périphérique (Anneau 4) contenait plus d'informations visuelles potentiellement plus utiles pour la catégorisation que l'anneau central. Dans ce contexte, Geuzebroek et van den Berg (2018) ont montré que la vision périphérique aurait justement un avantage sur la vision centrale lors de la catégorisation des scènes précisément parce qu'elle couvre une plus grande surface de champ visuel et qu'elle permet ainsi de révéler plus de contenu de la scène. Dans leur expérience, les auteurs présentaient des scènes à l'aide d'un paradigme Fenêtre-Scotome. Les scènes étaient présentées selon cinq conditions expérimentales (Figure 3.8). Dans une condition de visualisation centrale C1, une fenêtre d'un rayon de 5° était placée sur la scène et les participants ne voyaient que le centre de la scène. Dans une condition de visualisation centrale C2, la scène entière était réduite dans une fenêtre d'un rayon de 5°. Ainsi, les stimuli C1 et C2 activaient la même surface de V1, mais le stimulus C2 contenait plus d'informations visuelles que le stimulus C1. Dans une condition de visualisation périphérique P1, la scène était présentée dans un anneau dont le bord intérieur était fixé à 5° et le bord extérieur à 40°. Dans une condition de visualisation périphérique P2, la scène était présentée dans un anneau dont le bord intérieur était fixé à 12° et le bord extérieur à 40°. Enfin, dans une condition CO1, le stimulus P2 était réduit en tenant compte du facteur de magnification corticale. Pour cela, la scène était présentée dans un anneau dont le bord intérieur était fixé à 3,6° et le bord extérieur à 12°. Ainsi, les stimuli P2 et CO1 activaient la même surface V1 et contenaient la même quantité d'informations visuelles. Les résultats ont montré que les performances étaient plus faibles dans les conditions où seules les informations centrales étaient disponibles (C1 et C2) par rapport aux conditions où seules les informations périphériques étaient disponibles (P1 et P2), confirmant ainsi les résultats précédents de Larson et Loschky (2009), ainsi que les résultats de notre **Expérience 4**. Plus intéressant encore, les performances étaient moins bonnes dans la condition où seule l'information centrale était disponible (C1) par rapport à la condition où toutes les informations de la scène étaient présentées en vision centrale (C2). Cependant, pour les conditions P2 et CO1, les performances étaient similaires malgré le changement d'excentricité dans la présentation de la scène. Ces résultats suggèrent que les performances de catégorisation des scènes sont indépendantes de l'excentricité de la scène, à condition que la stimulation centrale et la stimulation périphérique activent toutes deux le même nombre de cellules de V1 et que la même quantité d'informations soit disponible en vision centrale et périphérique. Ceci n'est évidemment jamais le cas dans des conditions naturelles. Dans l'ensemble, ces résultats suggèrent que l'excentricité rétinienne joue un rôle minimal dans la perception de la scène. La vision périphérique prédominerait sur la vision centrale uniquement parce qu'elle couvre un

champ visuel plus large, permettant de présenter plus de contenu de la scène. Dans nos deux expériences, nous avons donc manipulé la largeur de nos stimuli de différentes excentricités en nous conformant à ces conditions naturelles de perception visuelle, c'est-à-dire, en ne cherchant pas à tout prix à ce que nos stimuli aient la même quantité d'information. Nous avons simplement manipulé la largeur des stimuli en fonction du facteur de magnification corticale de sorte que chaque stimulus active la même surface sur le cortex visuel primaire. Un neurone codant la vision périphérique recevant tout naturellement plus d'information du champ visuel qu'un neurone codant la vision centrale, il est donc logique que nos stimuli les plus périphériques contiennent le plus d'information visuelle - de plus faible résolution spatiale toutefois - de sorte qu'un objet perçu en vision centrale ne l'est pas forcément en vision périphérique.

Enfin, nous pensons réellement qu'il n'aurait pas été pertinent d'appliquer la même largeur à des stimuli de différentes excentricités. Par exemple, nous aurions pu appliquer la largeur du disque central (rayon de 1° d'angle visuel) aux anneaux périphériques. Mais ce contrôle expérimental aurait surpassé les capacités du système visuel en vision périphérique (non adapté à la vision des détails inférieur à 1° d'angle visuel) empêchant alors fortement la catégorisation. En revanche, si nous avons appliqué la largeur de l'Anneau 4 (de $7,74^\circ$ d'angle visuel) au rayon du disque central, ce dernier aurait alors contenu une partie du champ visuel périphérique. Par conséquent, si nous avons contrôlé la largeur des anneaux pour chaque excentricité rétinienne, nous n'aurions pas respecté les propriétés de la vision au naturel et nous aurions possiblement biaisé les stratégies utilisées par le système visuel pour catégoriser les scènes.

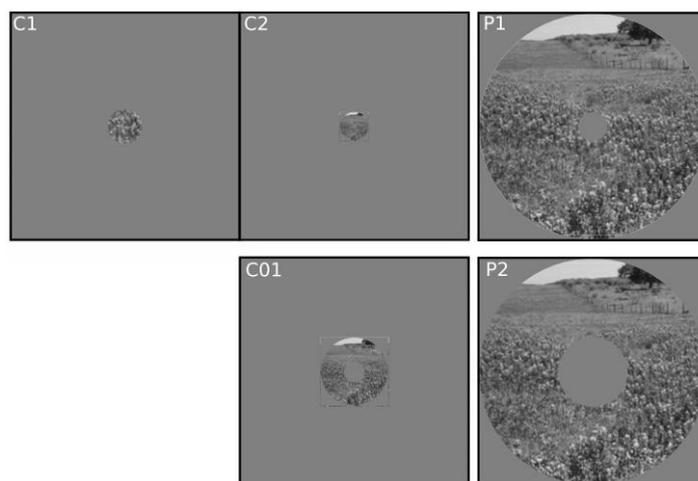


Figure 3.8 : Exemples de stimuli utilisés dans les différentes conditions expérimentales : C1, C2, P1, P2 et C01. Figure tirée de Geuzebroek et van den Berg (2018).

L'**Expérience 5** a été spécialement conçue pour tester si l'analyse rapide des informations grossières disponibles en vision périphérique pourrait permettre une première catégorisation de la scène qui serait ensuite validée par l'analyse des informations disponibles

en vision centrale (principalement une information détaillée en HFS). Pour cela, nous avons utilisé des séquences dynamiques composées du Disque central et des quatre anneaux de l'**Expérience 4**. Les stimuli étaient assemblés soit du Disque central à l'anneau le plus périphérique (Anneau 4) afin d'imposer une analyse de l'information visuelle *central-to-peripheral*, soit de l'anneau le plus périphérique (Anneau 4) au Disque central afin d'imposer une analyse de l'information visuelle *peripheral-to-central*. Dans cette expérience, nous avons également testé si la supériorité de la vision périphérique sur la vision centrale était amplifiée par des temps de présentation plus courts. Ainsi, les séquences étaient présentées pendant 100 ms ou 500 ms. Nous avons observé que les participants catégorisaient les séquences *peripheral-to-central* plus rapidement que les séquences *central-to-peripheral* pour les deux temps de présentation, cet avantage étant plus important pour le temps de présentation le plus court (100 ms). En accord avec un traitement *coarse-to-fine* des fréquences spatiales, il semblerait que la catégorisation de scènes suit également un traitement *peripheral-to-central*. Plus précisément, lorsque le temps de présentation des séquences est très court (100 ms), les informations grossières extraites rapidement dans le champ visuel périphérique (au cours des 20 premières ms) initieraient la catégorisation visuelle, qui serait ensuite affinée par des informations détaillées extraites plus lentement dans le champ visuel central.

Nous avons également observé que la présence d'un objet central accélérât le processus de catégorisation uniquement pour les séquences *central-to-peripheral* présentées pendant 500 ms. Cette absence d'effet pour des temps courts (100 ms) suggère que les informations détaillées de la vision centrale ne seraient même pas utilisées. Par conséquent, le poids de la vision centrale augmenterait pour des temps de présentation plus longs.

Enfin, nous avons conscience que pour une séquence de plus longue durée (500 ms), il est également possible que la catégorisation ne se soit basée que sur le premier anneau présenté pendant 100 ms. Cela pourrait d'ailleurs expliquer pourquoi les participants ont catégorisé plus lentement les séquences présentées pendant 100 ms que celles présentées pendant 500 ms. En effet, les résultats de l'**Expérience 4** et les travaux antérieurs sur la perception de scènes (Johnson & Olshausen, 2003 ; Thorpe, Fize & Marlot., 1996 ; VanRullen & Thorpe, 2001) démontrent la capacité du système visuel à catégoriser une scène en seulement 100 ms. Si la catégorisation ne repose que sur les 100 premières millisecondes de présentation, les participants devaient alors traiter cinq stimuli pour les séquences de 100 ms au lieu d'un seul stimulus pour les séquences présentées de 500 ms. La nécessité de cette accumulation rapide d'informations visuelles en seulement 100 ms pour les séquences de 100 ms pourrait alors expliquer un ralentissement de la réponse par rapport aux séquences de 500 ms où le système visuel ne devait traiter qu'un seul stimulus les 100 premières millisecondes.

4. Conclusion et limites

En conclusion, les résultats des **Expériences 4 et 5** confirment la plus grande efficacité et utilité de la vision périphérique lors de la catégorisation rapide des scènes, malgré la faible acuité visuelle inhérente à la vision périphérique et la valeur informative de la vision centrale.

En accord avec un traitement *coarse-to-fine* des fréquences spatiales, l'analyse rapide des informations en BFS disponibles en vision périphérique pourrait permettre une première catégorisation qui serait validée ou non par l'analyse plus lente des informations en HFS disponibles en vision centrale. Les résultats de ces deux expériences ont été soumis à la revue *Vision Research*. Les experts ont soulevé plusieurs biais méthodologiques que nous allons détailler ci-après.

Le premier biais concerne notre contrôle du facteur de magnification corticale lors de la construction des stimuli. Pour contrôler ce facteur dans nos expériences, la largeur et la surface des anneaux augmentaient de façon logarithmique (i.e. progressivement) avec l'excentricité. Une fonction logarithmique ne garantit toutefois pas que la surface de chaque stimulus ait la même surface d'activation sur le cortex visuel primaire. Il est possible que l'anneau le plus périphérique (Anneau 4) contenait plus d'informations visuelles, potentiellement plus utiles pour la catégorisation, que si nous avons calculé sa largeur sur la base d'une fonction de magnification corticale déterminée empiriquement. En conséquence, il est possible que l'avantage de la vision périphérique que nous avons observé dans nos expériences soit uniquement causé par ce choix méthodologique, car trop d'informations étaient potentiellement disponibles en vision périphérique (Anneau 4) qu'en vision centrale (Disque central).

Le second biais concerne les résultats de l'**Expérience 4** où nous avons observé de façon surprenante que la présence d'un objet au centre de la scène affectait aussi les performances de catégorisation dans plusieurs anneaux périphériques (Anneau 1 et Anneau 2) alors que l'objet n'était présent que dans le Disque central. Cet effet s'explique en fait simplement par la façon dont nous avons construit nos stimuli. Pour construire les scènes avec un objet en vision centrale, nous avons simplement centré les photographies originales sur un objet représentatif de la catégorie de la scène. Pour construire les scènes sans objet en vision centrale, nous avons déplacé le centre vers un emplacement différent, là où il n'y avait pas d'objet. En conséquence, les informations sémantiques contenues dans le Disque central n'étaient pas identiques entre les deux conditions expérimentales (i.e. Objet central ; Pas d'objet central), mais elles ne l'étaient pas non plus dans les anneaux périphériques. Ce choix méthodologique pourrait alors expliquer pourquoi les performances des Anneaux 1 et 2 différaient entre ces deux conditions.

Enfin, le dernier biais concerne l'hypothèse *zoom-out* proposée par Larson et al. (2014). En effet, ces auteurs ont observé une supériorité de la vision centrale pendant les 100 premières millisecondes de traitement, la supériorité de la vision périphérique apparaissant pour des temps plus longs. Pour interpréter ces résultats, les auteurs ont proposé l'hypothèse *zoom-out* selon laquelle l'attention est d'abord focalisée en vision centrale, puis se développe rapidement en vision périphérique. Dans notre **Expérience 4**, nous avons présenté nos stimuli pendant 100 ms (nous n'avons manipulé le temps de présentation que dans l'**Expérience 5**) et nous avons observé un avantage de la vision périphérique. Ce temps de présentation ne nous a donc pas permis de confronter l'hypothèse *zoom-out*.

Dans la suite de cette thèse, nous avons donc reconduit les **Expériences 4 et 5** en tenant compte des problèmes méthodologiques soulevés par les experts. Plus précisément, (1) nous avons construit de nouveaux stimuli qui activent la même surface corticale sur V1, (2) nous avons simplement retiré l'objet central de la photographie à l'aide d'un logiciel de retouche de photographies numériques pour construire les stimuli sans objet central et (3) nous avons

reproduit l'**Expérience 4** en testant également un temps plus court de présentation des stimuli isolés.

Enfin, l'avantage des séquences *peripheral-to-central* par rapport aux séquences *central-to-peripheral* que nous avons observé dans l'**Expérience 5** pourrait être en réalité causé par un effet de masquage visuelle. En effet, le stimulus périphérique, qui entoure le stimulus central, pourrait réduire la visibilité du stimulus central et agir comme un masque sur ce dernier. On parle dans le cas présent d'un effet de masquage latéral rétroactif ou masquage par métacontraste (ou *Metacontract masking*) qui est une forme de *visual backward masking* (Alpern, 1952 ; Breitmeyer & Ogmen, 2000 ; Breitmeyer et al., 2006 ; Enns & Lollo, 2000). Cet effet de masquage concernerait uniquement les séquences *central-to-peripheral*, où chaque stimulus de la séquence est suivi par un anneau plus large qui l'entoure (Figure 3.9). Spécifiquement, les anneaux périphériques, qui étaient présentés après le disque central dans les séquences *central-to-peripheral*, pouvaient réduire la visibilité du disque central. Nous avons donc également mené une expérience contrôle (**Expérience 8**) afin d'évaluer si l'avantage des séquences *peripheral-to-central* pourrait être biaisé par un effet de masquage rétroactif par métacontraste, c'est-à-dire par une visibilité réduite des stimuli présentés en vision centrale par les informations périphériques qui l'entourent ensuite.

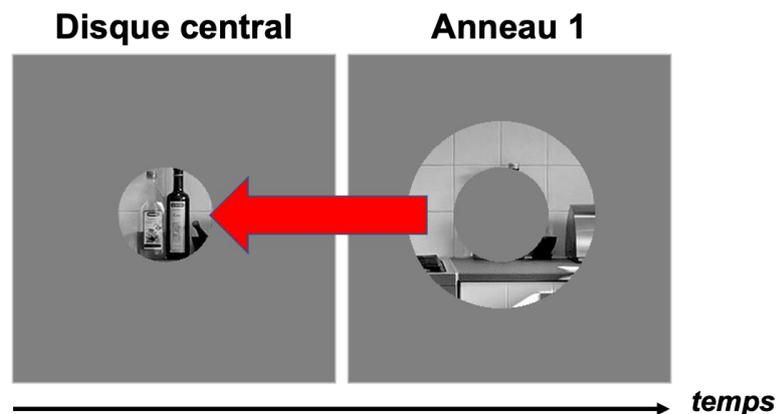


Figure 3.9 : Illustration de l'effet de masquage latéral rétroactif ou masquage par métacontraste (*Metacontract masking*) sur un extrait d'une séquence *central-to-peripheral*. La flèche rouge indique le sens du masquage, c'est-à-dire qu'un stimulus de la séquence *central-to-peripheral* est masqué par le stimulus qui l'entoure.

II. Expériences 6, 7 et 8 - Importance de la vision périphérique par rapport à la vision centrale : Contrôle de biais méthodologiques

L'objectif théorique était le même que celui des **Expériences 4 et 5** (i.e. étudier l'avantage de la vision périphérique par rapport à la vision centrale lors de la catégorisation rapide de scènes visuelles), mais nous avons apporté des modifications méthodologiques afin de pallier les biais soulignés plus haut.

Le premier biais concernait la fonction logarithmique que nous avons utilisée pour contrôler le facteur de magnification corticale lorsque nous avons construit nos stimuli de différentes excentricités. En effet, cette fonction n'était pas basée sur une approximation précise du facteur de magnification corticale. Nous avons donc répliqué nos expériences avec de nouveaux stimuli pour lesquels la largeur des anneaux a été calculée sur la base de la fonction de magnification corticale déterminée empiriquement par Wu et al. (2012, voir aussi Geuzebroek et al., 2018) afin que chaque stimulus active la même surface corticale de du cortex visuel primaire. A noter que l'utilisation de cette équation a eu pour effet de diminuer la quantité d'informations disponibles dans les anneaux les plus périphériques et d'augmenter la quantité d'informations disponibles dans le Disque central (Figure [3.10b](#)).

Le deuxième biais méthodologique concernait la manipulation de l'objet sémantiquement pertinent au centre de la scène. En effet, dans les **Expériences 4 et 5**, le fait de déplacer le centre des scènes entre la condition où il y avait un objet central et la condition où il n'y avait pas d'objet central a entraîné un effet inattendu dans lequel la présence de l'objet central affectait les performances de catégorisation dans plusieurs anneaux périphériques. Afin de remédier à ce biais méthodologique, nous avons créé des scènes contenant un objet en vision centrale toujours en centrant la photographie originale sur un objet représentatif de la catégorie de la scène, puis nous avons créé des scènes sans objet central en retirant cet objet à l'aide du logiciel de retouche de photographies numériques Photoshop (Figure [3.10a](#)). De cette façon, la position des informations sémantiques contenues dans les anneaux périphériques étaient identiques quelle que soit la condition expérimentale (objet central présent vs. objet central absent).

Ainsi, dans l'**Expérience 6**, nous avons testé la catégorisation de chaque stimulus qui composent les séquences dynamiques (i.e. Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4), de la même manière que l'Expérience 4. Les participants devaient catégoriser les stimuli comme étant une scène d'intérieur ou d'extérieur. Dans l'**Expérience 7**, nous avons testé la catégorisation des séquences dynamiques (*peripheral-to-central* et *central-to-peripheral*), de la même manière que l'Expérience 5. Enfin, nous avons mené une expérience contrôle (**Expérience 8**) afin d'évaluer si un avantage des séquences *peripheral-to-central* résulterait d'un effet de *Metacontract masking*.

1. **Expérience 6 – Catégorisation de larges scènes et de parties de scènes de différentes excentricités rétinienne**

L'objectif théorique de l'**Expérience 6** était le même que celui de l'**Expérience 4**, à savoir évaluer l'importance relative de la vision centrale par rapport à la vision périphérique lors du processus de catégorisation visuelle de scènes. Les participants devaient catégoriser les Disques entiers dans une Session 1, et ils devaient catégoriser les stimuli individuels (i.e. Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4) dans une Session 2. Toutefois, cette nouvelle expérience avait pour but de pallier les biais méthodologiques rencontrés dans l'**Expérience 4** : nos stimuli activaient désormais la même surface corticale sur le cortex visuel primaire et les informations sémantiques contenues au-delà du Disque central étaient identiques entre la condition Objet central présent et Objet central absent. Par ailleurs, dans cette expérience, nous avons également étudié l'effet des contraintes temporelles sur le traitement des informations visuelles disponibles en vision centrale et en vision périphérique en manipulant le temps de présentation des stimuli afin de tester l'hypothèse *zoom-out* (Larson et al., 2014). Nous avons donc présenté nos stimuli soit pendant un temps très court de 33 ms, soit pendant 100 ms comme dans l'**Expérience 5**. Nous tenons à faire remarquer que nous avons conservé le même écran entre l'**Expérience 4** et l'**Expérience 6**, mais nous avons changé d'ordinateur et le taux de rafraîchissement de ce nouvel ordinateur (60 Hz) ne nous a pas permis de présenter nos stimuli aussi rapidement que dans les précédentes expériences (nous allions jusqu'à seulement 20 ms dans l'**Expérience 5**).

1.1. Méthode

1.1.1. Participants

Trente-deux étudiants de l'Université Grenoble Alpes (28 femmes ; âge moyen : 20,63 \pm 2,68) ont participé à cette expérience. Ils ont été répartis en deux groupes ($n = 16$) dans lesquels les stimuli étaient présentés pendant 33 ms ou 100 ms. Nous avons fixé la taille de l'échantillon de chaque groupe en nous basant sur les résultats de l'**Expérience 4** et en particulier, sur l'interaction significative entre l'excentricité et la présence d'un objet central que nous avons observé pour les temps de réponses correctes moyens, avec une taille d'effet estimée par le calcul de l'éta-carré partiel (η^2) de 0.22. Nous avons ensuite converti la valeur de l'éta-carré partiel en d de Cohen afin de définir la taille d'échantillon. Ainsi, nous avons défini la taille de l'échantillon grâce à une analyse de puissance a priori pour test t de Student à échantillon dépendant réalisée avec GPower 3.1, avec une taille d'effet attendue de 1.061 (d de Cohen obtenu), une puissance désirée de 0.99 et un seuil alpha de 0.05. Grâce à cette analyse, nous avons obtenu une taille d'échantillon minimale de 16 participants par groupe expérimental, soit un total de 32 participants. Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Tous les participants ont donné leur consentement libre et éclairé avant de participer à cette expérience qui a été

menée en conformité avec le code d'éthique de l'*American Psychological Association* (APA) et approuvée par un comité et approuvée par un comité d'éthique (CERGA, Comité d'éthique de la recherche Grenoble Alpes, IRB00010290-2018-10-16-50). A la fin de l'expérience, chaque participant recevait un bon de 0,25 point par quart d'heure d'expérience à valoir lors des examens universitaires.

1.1.2. Stimuli

Nous avons travaillé sur la base des 50 photographies couleurs de scènes issues du site web Pixabay (<https://pixabay.com/fr/>) et précédemment utilisées dans l'**Expérience 4**. La moitié des photographies représentait des scènes d'intérieur (e.g., salon, cuisine, salle de bain) et l'autre moitié représentait des scènes d'extérieur (e.g., route, montagne, plage). Nous avons utilisé les images carrées de 1500×1500 pixels (Figure 3.10a) qui contenaient au centre un objet tenant dans un carré de 100×100 pixels. Sur la base de ces images, nous avons retiré l'objet du carré 100×100 pixels, ainsi que quelques éléments adjacents (certains de ces éléments pouvant s'étendre sur 200 pixels de large), à l'aide du logiciel de retouches de photographies Adobe Photoshop CS6. Pour cela, nous avons utilisé trois outils : l'outil 'remplissage d'après le contenu', le 'correcteur localisé' et le 'tampon de duplication'. Nous avons d'abord utilisé l'outil 'remplissage d'après le contenu', il permet de remplacer la zone sélectionnée de l'image avec du contenu similaire provenant des parties adjacentes de l'image. Nous avons par exemple utilisé cet outil pour remplacer une maison par du ciel déjà présent autour de la maison. Puis nous avons utilisé le 'correcteur localisé' afin de nettoyer précisément les imperfections créées par le premier outil. Cet outil permet de supprimer les défauts de l'image en ciblant très précisément les pixels à retoucher. Il fonctionne de la même façon que le premier outil. Nous avons ensuite utilisé le 'tampon de duplication', il permet de prélever une zone sélectionnée de l'image afin de la dupliquer sur d'autres parties de l'image. La taille et l'opacité de la zone dupliquée peuvent être définies à l'avance. Nous avons par exemple utilisé cet outil pour prolonger une étagère déjà existante. Enfin, afin d'obtenir une scène aussi naturelle que possible, nous avons recréé les ombres et les lumières à la place de l'objet avec les outils 'haute intensité' et 'basse densité'. Nous avons dû retirer 10 scènes d'intérieur et 10 scènes d'extérieur de nos stimuli car les éléments contenus dans ces scènes nous empêchaient de retirer l'objet sans que les retouches soient invisibles.

La manipulation des images a ensuite été réalisée à l'aide de MATLAB (Mathworks Inc., Sherborn, MA, USA). Toutes les images ont été converties en 256 niveaux de gris en faisant la moyenne des valeurs rouge, verte et bleue de chaque pixel. Nous avons également égalisé la luminance et le contraste des stimuli. Tous les stimuli ont été normalisés de façon à obtenir une luminance moyenne de 0,5 et un contraste RMS (Root Mean Square) de 0,2 pour des valeurs d'intensité des pixels comprises entre 0 et 1 (i.e. une luminance moyenne de 128 et un contraste RMS de 51 sur une échelle de 256 niveaux de gris). Ainsi, nous avons créé 80 scènes (40 scènes d'intérieur et 40 scènes d'extérieur) de taille 1500×1500 pixels. Les stimuli peuvent être téléchargés depuis ce lien <https://osf.io/ke6np/>.

Pour chaque scène, nous avons ensuite construit six stimuli qui révélait différentes parties de la scène en tenant compte du facteur de magnification corticale (Figure 3.10). Le premier stimulus (Disque entier) révélait la scène à travers un large disque d'un rayon de 15° d'angle visuel (i.e. 750 pixels) sur un fond gris de luminance moyenne de 0,5. Les quatre autres stimuli (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4) ont été construits à l'aide d'une équation dérivée empiriquement d'une mesure rétinotopique en IRMf (Wu et al., 2012 ; pour une procédure similaire, voir Geuzebroek & van den Berg, 2018) :

$$y = \frac{\left(\int_{r_{inner}}^{r_{outer}} (553.99 - 123.98 \ln(x)) dx\right)}{7.5}$$

Dans cette équation, y correspond à la surface du cortex visuel primaire (en millimètres carrés) activée par le stimulus et r_{outer} et r_{inner} correspondent respectivement aux rayons internes et externes (en degrés d'angle visuel) des stimuli (Figure 3.10b). Sur la base de cette équation, nous avons créé cinq stimuli de différentes excentricités mais qui activent la même surface sur le cortex visuel primaire (20% chacun, i.e. $\sim 136 \text{ mm}^3$). Le Disque central révélait la scène à travers un disque d'un rayon de seulement $1,66^\circ$ d'angle visuel (i.e. 83 pixels). Il était donc plus large que le Disque central utilisé dans l'Expérience 4. Pour le premier anneau (Anneau 1), adjacent au Disque central, les bords internes et externes étaient respectivement fixés à $1,66^\circ$ d'angle visuel (83 pixels) et $4,10^\circ$ d'angle visuel (205 pixels). Pour le second anneau (Anneau 2), adjacent à l'Anneau 1, les bords internes et externes étaient respectivement fixés à $4,10^\circ$ d'angle visuel (205 pixels) et $7,10^\circ$ d'angle visuel (355 pixels). Pour le troisième anneau (Anneau 3), adjacent à l'Anneau 2, les bords internes et externes étaient respectivement fixés à $7,10^\circ$ d'angle visuel (355 pixels) et $10,70^\circ$ d'angle visuel (535 pixels). Enfin, pour le quatrième anneau (Anneau 4), adjacent à l'Anneau 3, les bords internes et externes étaient respectivement fixés à $10,70^\circ$ d'angle visuel (535 pixels) et 15° d'angle visuel (750 pixels). Le premier stimulus, ou Disque entier, correspondait à la somme ou superposition de tous les autres stimuli. Tous les stimuli étaient présentés sur un fond gris de luminance moyenne de 0,5 pour des valeurs d'intensité des pixels comprises entre 0 et 1.

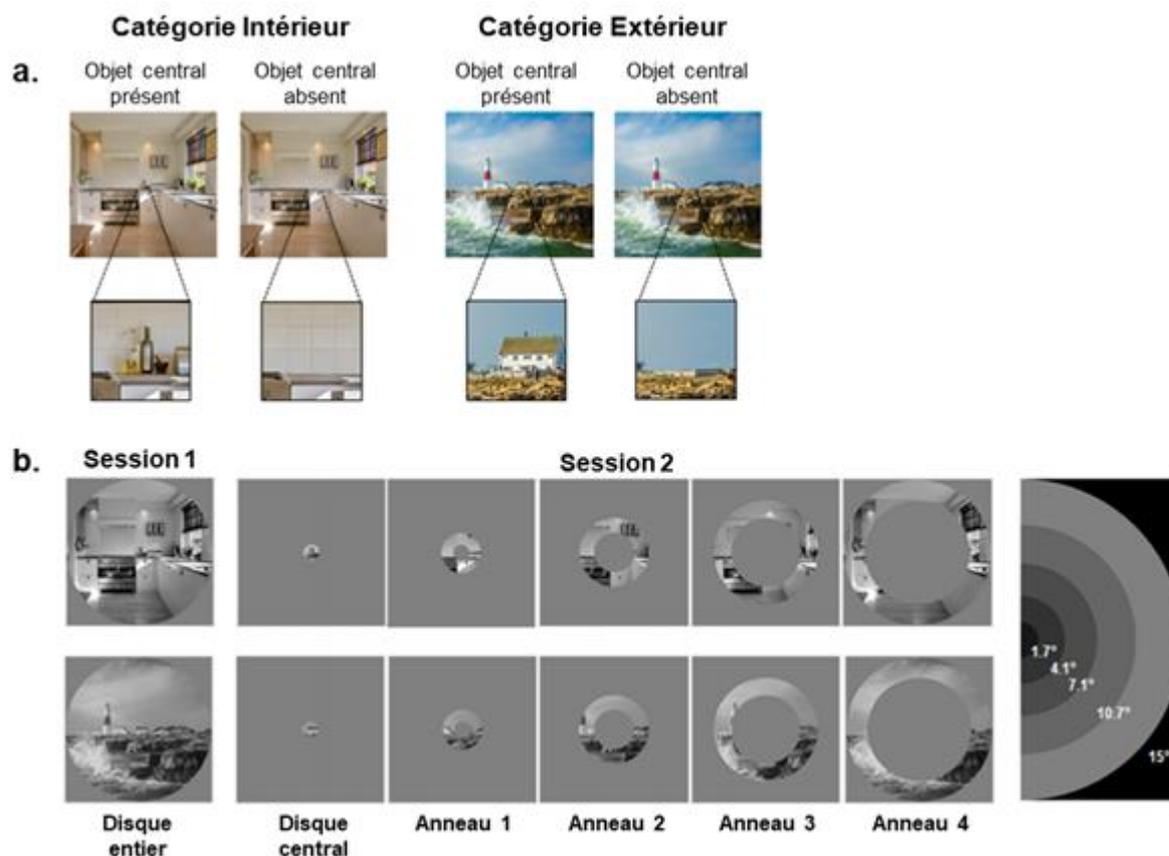


Figure 3.10 : (a) Exemple de scènes d'intérieur et d'extérieur de taille 1500 × 1500 pixels présentées avec ou sans objet sémantiquement relié à la catégorie de la scène tenant dans un carré central de taille 100 × 100 pixels. (b) Exemple de stimuli utilisés dans la Session 1 (Disque entier) et dans la Session 2 (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4) de l'Expérience 6. L'image de droite illustre les différentes excentricités (en degrés d'angle visuel) des bords intérieurs et extérieurs des stimuli qui augmentent selon une équation dérivée empiriquement de la mesure rétinotopique (Wu et al., 2012).

1.1.3. Procédure

L'expérience a été réalisée dans un box expérimental du Laboratoire de Psychologie et NeuroCognition (LPNC) de l'Université de Grenoble Alpes. Les participants étaient installés dans l'obscurité face à un écran d'ordinateur. Ils posaient leur menton sur une mentonnière installée sur l'axe sagittal médian du participant, à 70 cm de l'écran afin de respecter la taille angulaire de 30° d'angle visuel des stimuli. L'expérience était programmée grâce au logiciel E-prime 2.0 (E-prime Psychology Software Tools Inc., Pittsburgh, USA) et elle était présentée sur un écran DELL ULTRASHARP (60 Hz) de 30 pouces (61,4 × 40 cm), avec un affichage de résolution 2560 × 1600 pixels.

Tous les participants étaient répartis dans deux groupes (n = 16 dans chaque groupe) dans lesquels les stimuli étaient présentés pendant 33 ms ou 100 ms. De la même façon que dans l'Expérience 4, tous les participants réalisaient deux sessions expérimentales. Dans la

première session expérimentale (Session 1), les participants devaient catégoriser les Disques entiers. Dans la deuxième session expérimentale (Session 2), les participants devaient catégoriser les stimuli de différentes excentricités présentés aléatoirement (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4).

Pour chaque session expérimentale, un essai se déroulait comme suit (Figure 3.11) : un point de fixation central (noir sur un fond gris de luminance 0,5) était présenté pendant 500 ms, il était suivi de la présentation du stimulus (Session 1 : Disque entier ; Session 2 : Disque central ou anneaux périphériques) pendant 33 ms ou 100 ms selon le groupe expérimental, puis d'un écran gris (luminance : 0,5) de 1900 ms au cours duquel les participants pouvaient répondre. Les participants devaient catégoriser la scène comme étant soit une scène d'intérieur, soit une scène d'extérieur par l'intermédiaire des touches 1 et 2 du clavier d'ordinateur placé sur l'axe sagittal médian des participants. La moitié des participants appuyait sur la touche 1 avec l'index de la main droite pour la catégorie Intérieur et sur la touche 2 avec le majeur de la main droite pour la catégorie Extérieur. Les touches de réponses étaient contrebalancées pour l'autre moitié des participants. Les participants avaient pour consigne de répondre le plus correctement et rapidement possible. La Session 1 comprenait 80 essais (20 scènes d'intérieur et 20 scènes d'extérieur avec un objet au centre, 20 scènes d'intérieur et 20 scènes d'extérieur sans objet au centre) et durait environ 3 minutes. La Session 2 comprenait 400 essais (pour chaque type de stimulus, 20 scènes d'intérieur et 20 scènes d'extérieur avec un objet au centre, 20 scènes d'intérieur et 20 scènes d'extérieur sans objet au centre) et durait environ 16 minutes, entrecoupés de pauses tous les 80 essais.

Pour chaque essai, nous enregistrions l'exactitude de la réponse et le temps de réponse (en millisecondes). Avant de commencer chaque session expérimentale, les participants réalisaient quelques essais d'entraînement afin de se familiariser avec la tâche et les stimuli (10 essais).

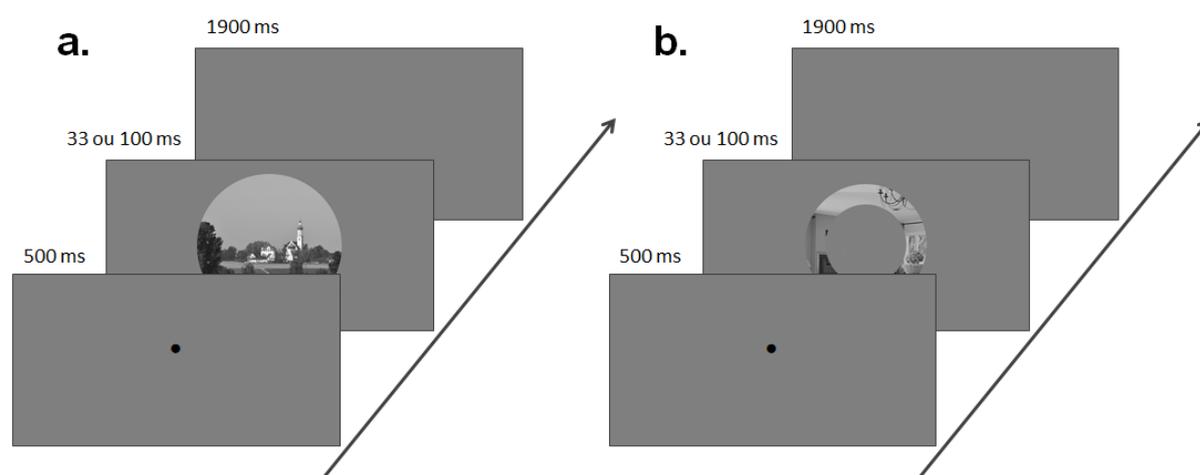


Figure 3.11 : Schéma de la procédure expérimentale de l'**Expérience 6**. Les stimuli étaient présentés pendant 33 ms ou 100 ms. Les participants devaient catégoriser **(a)** les Disques entier dans la Session 1 et **(b)** les stimuli de différentes excentricités (Disque central, Anneau 1, Anneau 2, Anneau 3, Anneau 4) dans la Session 2.

1.2. Hypothèses

Les hypothèses étaient identiques à celles de l'**Expérience 4**. Nous nous attendions donc à ce que les performances de catégorisation s'améliorent (i.e. à ce que les pourcentages d'erreurs et les temps de réponses diminuent) à mesure que l'information visuelle périphérique était disponible. Nous nous attendions également à ce que la présence d'un objet influence les différences de performances entre la catégorisation de l'information présente dans le disque central et celle de l'information présente dans l'anneau périphérique. En d'autres termes, nous nous attendions à une interaction entre l'excentricité de l'information visuelle disponible (centrale vs. périphérique) et la présence d'un objet en vision centrale. Par ailleurs, si la présence d'un objet pertinent en vision centrale améliore à lui seul la catégorisation de la scène, nous nous attendions à de meilleures performances pour la catégorisation des scènes entières qui contenaient un objet en vision centrale que pour les scènes entières qui n'en contenaient pas.

Enfin, sur la base de l'hypothèse d'un traitement *coarse-to-fine* prédominant lors de la catégorisation des scènes (Schyns & Oliva, 1994), nous nous attendions à ce que les informations grossières disponibles en vision périphérique soient davantage prises en compte lorsque le système visuel dispose de peu de temps. Ainsi, l'avantage de vision périphérique sur la vision centrale devrait être d'autant plus important que le temps de présentation est court. Alternativement, selon l'hypothèse du *zoom-out* de l'orientation de l'attention (Larson et al., 2014), les informations disponibles en vision centrale devraient être priorisées pour des temps de présentation courts et nous devrions observer de meilleures performances de catégorisation en vision centrale qu'en vision périphérique. Dans les deux cas, le temps de présentation des stimuli devrait interagir avec leur excentricité rétinienne.

1.3. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale de chaque session, les pourcentages d'erreurs moyens (%ERm) et les temps de réponses correctes moyens en millisecondes (TRm) ont été calculés. Afin de réduire l'effet des valeurs extrêmes sur les résultats, les temps de réponses supérieurs ou inférieurs à trois fois l'intervalle interquartile ont été supprimés. A l'aide de cette méthode, nous avons supprimé seulement 0,15% des temps de réponses pour la Session 1 et 0,49% des temps de réponses pour la Session 2. L'analyse des données a été effectuée à l'aide du logiciel Statistica 13.3 et du logiciel RStudio. Le seuil de significativité a été fixé à 0,05. La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2).

1.3.1 Session 1 : Catégorisation des Disques entiers

Pour l'analyse de la Session 1, une analyse de la variance à mesure répétées (ANOVA) de type $2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm avec comme facteur intra-sujet la Présence d'un objet central (Objet central présent ; Objet central absent), la Catégorie

de la scène (Intérieur ; Extérieur), et comme facteur inter-sujet le Temps de présentation des stimuli (33 ms ; 100 ms). Les pourcentages d'erreurs moyens (%ERm) et leurs erreurs-types ainsi que les temps de réponses moyens (TRm) et leurs erreurs-types sont présentés dans le Tableau 3.4.

Tableau 3.4 : Résultats de la Session 1 de l'Expérience 6. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Présence d'un objet central (Objet central présent ; Objet central absent), du Temps de présentation des stimuli (33 ms ; 100 ms) et de la Catégorie de la scène (Intérieur ; Extérieur).

		Objet central présent		Objet central absent	
		%ERm	TRm	%ERm	TRm
33 ms	Intérieur	4,06 ± 1,36	621 ± 21	4,37 ± 1,80	622 ± 24
	Extérieur	5,00 ± 1,27	629 ± 22	4,38 ± 1,54	643 ± 25
100 ms	Intérieur	5,31 ± 1,36	605 ± 21	3,75 ± 1,18	600 ± 24
	Extérieur	3,75 ± 1,27	605 ± 22	5,63 ± 1,54	621 ± 25

L'analyse de la variance (ANOVA) conduite sur les %ERm et sur les TRm n'a révélé aucun résultat significatif. Quelle que soit la condition expérimentale, les taux d'erreurs étaient bas (inférieurs à 5%, Figure 3.12). L'effet principal de la Présence d'un objet central n'était pas significatif, ni pour les %ERm (Objet présent : $4,53 \pm 0,96\%$; Objet absent : $4,53 \pm 1,16\%$; $F(1,29) < 1$), ni pour les TRm (Objet présent : 615 ± 20 ms ; Objet absent : 622 ± 23 ms ; $F(1,30) = 1,22$, $p = .277$). L'effet principal du Temps de présentation n'était pas non plus significatif, ni pour les %ERm ($F(1,30) < 1$), ni pour les TRm ($F(1,30) < 1$) et ce facteur n'interagissait pas avec la Présence d'un objet central (%ERm : $F(1,30) < 1$; TRm : $F(1,30) < 1$). Enfin, la Catégorie n'interagissait pas avec la Présence d'un objet central, ni pour les %ERm ($F(1,30) < 1$), ni pour les TRm ($F(1,30) = 2,34$, $p = .136$). Comme pour l'Expérience 4, ces résultats suggèrent que la présence d'un objet central sémantiquement relié à la catégorie de la scène n'améliore pas les performances de catégorisation, quel que soit le temps de présentation des stimuli. Toutefois, cette interprétation peut être à nouveau limitée par un potentiel effet plafond.

1.3.2. Session 2 : Catégorisation des portions de scènes (Disque central, Anneau 1, Anneau 2, Anneau 3 et Anneau 4)

Pour la Session 2, une analyse de la variance à mesure répétées (ANOVA) de type $5 \times 2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm avec comme facteur intra-sujet l'Excentricité (Disque central ; Anneau 1 ; Anneau 2 ; Anneau 3 ; Anneau 4), la Présence d'un objet central (Objet central présent ; Objet central absent), la Catégorie de la scène (Intérieur ; Extérieur) et le Temps de présentation des stimuli (33 ms ; 100 ms). Les pourcentages d'erreurs moyens (%ERm) et leurs erreurs-types ainsi que les temps de réponses moyens (TRm) et leurs erreurs-types sont présentés dans le Tableau 3.5.

Tableau 3.5 : Résultats de la Session 2 de l'Expérience 6. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de l'Excentricité des stimuli (Disque central ; Anneau 1 ; Anneau 2 ; Anneau 3 ; Anneau 4), de la Présence d'un objet central (Objet central présent ; Objet central absent), du Temps de présentation des stimuli (33 ms ; 100 ms) et de la Catégorie de la scène (Intérieur ; Extérieur).

		Objet central présent		Objet central absent			
		%ERm	TRm	%ERm	TRm		
33 ms	Disque central	Intérieur	9,38 ± 2,28	673 ± 22	22,50 ± 3,96	737 ± 27	
		Extérieur	7,50 ± 1,80	685 ± 24	18,13 ± 2,52	742 ± 25	
	Anneau 1	Intérieur	9,06 ± 1,36	661 ± 27	6,25 ± 1,78	658 ± 22	
		Extérieur	3,75 ± 1,36	683 ± 26	3,13 ± 1,00	667 ± 23	
	Anneau 2	Intérieur	7,19 ± 1,93	644 ± 25	5,31 ± 1,42	658 ± 27	
		Extérieur	2,50 ± 1,01	657 ± 27	3,13 ± 1,24	654 ± 26	
	Anneau 3	Intérieur	4,38 ± 1,57	649 ± 23	5,31 ± 1,69	657 ± 24	
		Extérieur	3,44 ± 1,15	659 ± 23	5,00 ± 1,58	649 ± 28	
	Anneau 4	Intérieur	5,00 ± 2,45	676 ± 29	6,88 ± 2,22	697 ± 26	
		Extérieur	4,38 ± 1,55	647 ± 26	3,13 ± 1,66	660 ± 23	
	100 ms	Disque central	Intérieur	12,19 ± 2,28	698 ± 22	23,13 ± 3,96	733 ± 27
			Extérieur	12,19 ± 1,80	719 ± 24	20,63 ± 2,52	756 ± 25
Anneau 1		Intérieur	6,25 ± 1,36	713 ± 27	8,13 ± 1,78	687 ± 22	
		Extérieur	4,06 ± 1,36	693 ± 26	3,44 ± 1,00	688 ± 23	
Anneau 2		Intérieur	8,44 ± 1,93	696 ± 25	5,63 ± 1,42	697 ± 27	
		Extérieur	3,45 ± 1,01	688 ± 27	4,38 ± 1,24	680 ± 26	
Anneau 3		Intérieur	8,13 ± 1,57	675 ± 23	6,88 ± 1,69	702 ± 24	
		Extérieur	2,50 ± 1,15	660 ± 23	4,38 ± 1,58	670 ± 28	
Anneau 4		Intérieur	11,56 ± 2,45	709 ± 29	11,25 ± 2,22	695 ± 26	
		Extérieur	4,06 ± 1,55	684 ± 26	5,14 ± 1,66	666 ± 23	

- Analyse des pourcentages d'erreurs moyens

L'analyse de la variance (ANOVA) menée sur les %ERm a révélé un effet principal de l'Excentricité ($F(4,120) = 65,79, p < .001, \eta^2 = .69$; Figure 3.12a). Les tests de contraste linéaire et quadratique réalisés entre les différentes excentricités étaient significatifs (Contraste linéaire : $F(1,30) = 84,09, p < .001$; Contraste quadratique : $F(1,30) = 114,24, p < .001$; Disque central : $15,74 \pm 2,13\%$; Anneau 1 : $5,51 \pm 1,32\%$; Anneau 2 : $5,00 \pm 1,47\%$; Anneau 3 : $5,00 \pm 1,44\%$; Anneau 4 : $6,42 \pm 1,89\%$). Nous avons ensuite effectué des comparaisons planifiées entre chaque stimulus afin d'identifier quel anneau conduisait à de meilleures performances. Ces analyses ont montré une différence significative entre le Disque central et l'Anneau 1 ($F(1,30) = 146,13, p < .001$), mais pas de différences significative ni entre l'Anneau 1 et l'Anneau 2 ($F(1,30) < 1$), ni entre l'Anneau 2 et l'Anneau 3 ($F(1,30) < 1$), ni entre l'Anneau 3 et l'Anneau 4 ($F(1,30) = 3,268, p = .081$). Ces résultats suggèrent que les erreurs diminuent de façon exponentielle avec l'augmentation de l'excentricité.

L'effet principal de la Présence de l'objet central était également significatif ($F(1,30) = 19,45, p < .001$), $\eta^2 = .39$). Les participants faisaient en moyenne moins d'erreurs pour

catégoriser une scène avec un objet central ($6,49 \pm 2,20\%$) par rapport à une scène sans objet central ($8,59 \pm 2,27\%$). La Présence de l'objet central interagissait avec l'Excentricité ($F(4,120) = 25,04, p < .001, \eta^2 = .45$). Nous avons effectué des comparaisons planifiées entre chaque stimulus afin d'identifier quel anneau conduisait à de meilleures performances en fonction de la présence ou non d'un objet central. Lorsqu'il y avait un objet central, ces analyses ont montré une différence significative entre le Disque central et l'Anneau 1 ($F(1,30) = 23,28, p < .001$), mais aucune différence significative ni entre l'Anneau 1 et l'Anneau 2 ($F(1,30) < 1$), ni entre l'Anneau 2 et l'Anneau 3 ($F(1,30) = 1,30, p = .264$), ni entre l'Anneau 3 et l'Anneau 4 ($F(1,30) = 2,44, p = .129$). Lorsqu'il n'y avait pas d'objet central, ces analyses ont à nouveau montré une différence significative entre le Disque central et l'Anneau 1 ($F(1,30) = 163,05, p < .001$), mais aucune différence significative ni entre l'Anneau 1 et l'Anneau 2 ($F(1,30) < 1$), ni entre l'Anneau 2 et l'Anneau 3 ($F(1,30) < 1$), ni entre l'Anneau 3 et l'Anneau 4 ($F(1,30) = 1,21, p = .279$). De manière critique, pour le Disque central, les participants faisaient moins d'erreurs pour catégoriser une scène contenant un objet central ($10,39 \pm 1,35\%$; $F(1,30) = 37,32, p < .001$) que pour catégoriser une scène qui n'en contenait pas ($21,09 \pm 2,19\%$). Mais malgré la présence d'un objet central qui améliorerait significativement les performances de catégorisation, les participants faisaient toujours moins d'erreurs lorsque l'information était disponible en vision périphérique qu'en vision centrale.

L'effet principal du Temps de présentation n'était quant à lui pas significatif ($F(1,30) = 1,33, p = .258$) et ce facteur n'interagissait ni avec l'Excentricité ($F(4,120) = 1,42, p = 0,231$), ni avec l'interaction Excentricité \times Présence d'un objet central ($F(4,120) < 1$). Enfin, l'effet principal de la Catégorie était significatif ($F(1,30) = 18,85, p < .001, \eta^2 = .39$). Les participants faisaient moins d'erreurs pour catégoriser des scènes d'extérieur ($5,93 \pm 1,20\%$) que des scènes d'intérieur ($9,14 \pm 2,79\%$). Par contre, la Catégorie n'interagissait pas avec les autres facteurs manipulés (Excentricité \times Catégorie : $F(4,120) < 1$; Excentricité \times Temps de présentation \times Catégorie : $F(4,120) < 1$; Excentricité \times Présence d'un objet central \times Catégorie : $F(4,120) = 1,27$; Excentricité \times Présence d'un objet central \times Temps de présentation \times Catégorie : $F(1,24) < 1$).

Nous avons également comparé les performances de chacune des excentricités de la Session 2 (Disque central, Anneau 1, Anneau 2, Anneau 3, Anneau 4) aux performances obtenues dans la Session 1 (Disque entier). Ces comparaisons ont montré que les performances de catégorisation de l'Anneau 4 et du Disque central étaient significativement différentes de la catégorisation du Disque entier (Anneau 4 : $t(32) = -2,41, p = .022$; Disque central : $t(32) = -12,60, p < .001$). Les performances des autres excentricités (Anneau 3, Anneau 2 et Anneau 1) n'étaient pas significativement différentes de la catégorisation du Disque entier ($ts(32) < -1,62, ps > .114$). Ces résultats suggèrent que la catégorisation des scènes est désavantagée lorsque seules les informations visuelles centrales ou que seules les informations visuelles les plus périphériques sont disponibles.

- **Analyse des temps de réponses correctes moyens**

L'analyse de la variance (ANOVA) menée sur les TRm a révélé un effet principal de l'Excentricité ($F(4,120) = 29,32, p < .001, \eta^2 = .49$; Figure [3.12b](#)). Similairement à l'analyse

des %ERm, le test de contraste linéaire et quadratique effectué entre les différentes excentricités a révélé une diminution à la fois linéaire et quadratique des temps de réponses à mesure que l'excentricité augmentait (Contraste linéaire : $F(1,30) = 61,47, p < .001$; Contraste quadratique : $F(1,30) = 37,41, p < .001$; Disque central : 718 ± 32 ms ; Anneau 1 : 679 ± 33 ms ; Anneau 2 : 672 ± 35 ms ; Anneau 3 : 565 ± 33 ms ; Anneau 4 : 679 ± 34 ms). Nous avons ensuite effectué des comparaisons planifiées entre chaque stimulus afin d'identifier quel anneau conduisait à de meilleures performances. Ces analyses ont montré une différence significative entre le Disque central et l'Anneau 1 ($F(1,30) = 55,86, p < .001$) et entre l'Anneau 3 et l'Anneau 4 ($F(1,30) = 5,13, p = .031$), mais pas de différences significative ni entre l'Anneau 1 et l'Anneau 2 ($F(1,30) = 2,03, p = .164$), ni entre l'Anneau 2 et l'Anneau 3 ($F(1,30) = 2,17, p = .151$). Ces résultats suggèrent que les temps de réponses suivent une courbe en forme de 'U'.

Comme pour l'analyse des %ERm, l'effet principal de la Présence de l'objet central était significatif ($F(1,30) = 12,71, p = .001, \eta^2 = .30$). Les participants étaient plus rapides pour catégoriser une scène avec un objet central (677 ± 52 ms) que pour catégoriser une scène sans objet central (688 ± 52 ms). La Présence de l'objet central interagissait avec l'Excentricité ($F(4,120) = 10,32, p < .001, \eta^2 = .26$). Nous avons effectué des comparaisons planifiées entre chaque stimulus afin d'identifier quel anneau conduit à de meilleures performances en prenant en compte la présence d'un objet central. Lorsqu'il y avait un objet central, ces analyses ont montré une différence significative entre l'Anneau 3 et l'Anneau 4 ($F(1,30) = 4,32, p = .050$), mais aucune différence significative ni entre le Disque central et l'Anneau 1 ($F(1,30) = 2,42, p = .130$), ni entre l'Anneau 1 et l'Anneau 2 ($F(1,30) = 2,10, p = .157$), ni entre l'Anneau 2 et l'Anneau 3 ($F(1,30) = 2,28, p = .141$). Lorsqu'il n'y avait pas d'objet central, ces analyses ont montré une différence significative entre le Disque central et l'Anneau 1 ($F(1,30) = 60,80, p < .001$), mais aucune différence significative ni entre l'Anneau 1 et l'Anneau 2 ($F(1,30) < 1$), ni entre l'Anneau 2 et l'Anneau 3 ($F(1,30) < 1$), ni entre l'Anneau 3 et l'Anneau 4 ($F(1,30) = 1,87, p = .181$). De manière critique, pour le Disque central, les comparaisons planifiées ont montré que les participants étaient plus rapides pour catégoriser une scène contenant un objet central (694 ± 21 ms ; $F(1,30) = 37,32, p < .001$) qu'une scène qui n'en contenait pas (742 ± 25 ms). L'effet principal du Temps de présentation n'était pas significatif ($F(1,30) < 1$) et ce facteur n'interagissait ni avec l'Excentricité ($F(4,120) = 1,37, p = .248$), ni avec l'interaction Excentricité \times Présence d'un objet central ($F(4,120) = 2,16, p = .078$).

Enfin, l'effet principal de la Catégorie n'était pas significatif ($F(1,30) = 1,48, p = .232$), mais ce facteur interagissait avec l'Excentricité ($F(4,120) = 5,47, p < .001$). Les comparaisons planifiées ont montré que les participants étaient plus rapides pour catégoriser des scènes d'extérieur que des scènes d'intérieur uniquement pour l'Anneau 4 ($F(1,30) = 8,53, p = 0,006$; Intérieur : 694 ± 26 ms ; Extérieur : 664 ± 24 ms), mais pas pour les autres excentricités (Disque central : $F(1,30) = 2,86, p = .101$; Anneau 1 : $F(1,30) < 1$; Anneau 2 : $F(1,30) < 1$; Anneau 3 : $F(1,30) = 2,548, p = .121$). Par contre, la Catégorie n'interagissait pas avec les autres facteurs manipulés (Excentricité \times Temps de présentation \times Catégorie : $F(4,120) = 1,28, p = .28$; Excentricité \times Présence d'un objet central \times Catégorie : $F(4,30) < 1$; Excentricité \times Présence d'un objet central \times Temps de présentation \times Catégorie : $F(1,30) < 1$).

Comme pour l'analyse des %ERm, nous avons comparé les performances de chacune des excentricités de la Session 2 (Disque central, Anneau 1, Anneau 2, Anneau 3, Anneau 4)

aux performances obtenues dans la Session 1 (Disque entier). Ces comparaisons ont montré que les performances de catégorisation de toutes les excentricités (Anneau 4, Anneau 3, Anneau 2, Anneau 1, Disque central) étaient significativement différentes de la catégorisation du Disque entier à la fois lorsqu'il n'y avait pas d'objet et lorsqu'il y en avait un ($t_s(32) > -4,17, p_s < .001$). Ainsi, une présentation partielle d'une scène allongeait systématiquement les temps de réponses.

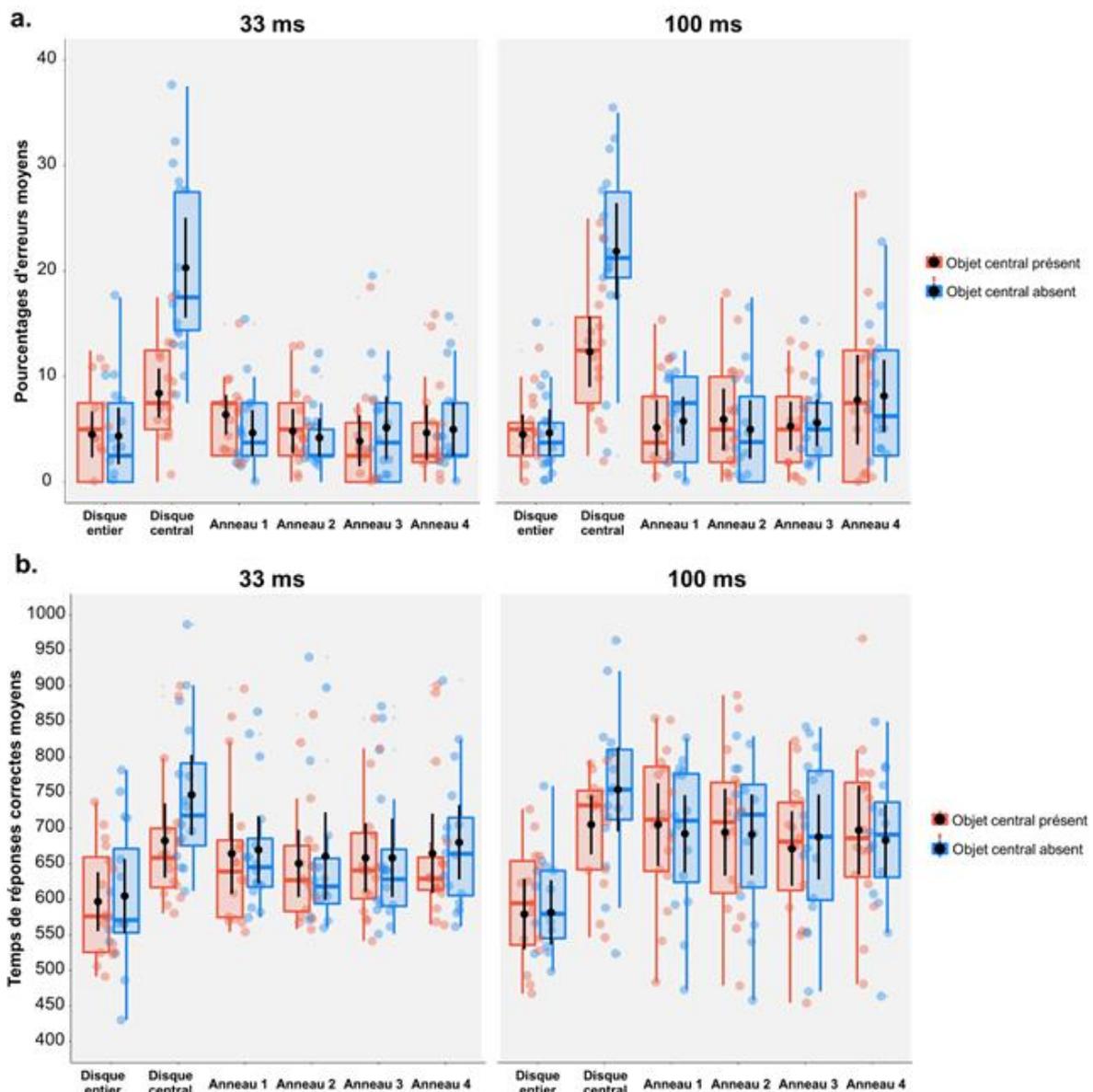


Figure 3.12 : Résultats de la Session 1 (catégorisation du Disque entier) et de la Session 2 (catégorisation du Disque central, de l'Anneau 1, de l'Anneau 2, de l'Anneau 3 et de l'Anneau 4) de l'Expérience 6 : (a) Pourcentages d'erreurs moyens et (b) Temps de réponses correctes moyens en ms en fonction de l'Excentricité des stimuli (Disque entier ; Disque central ; Anneau 1 ; Anneau 2 ; Anneau 3 ; Anneau 4), de la Présence d'un objet central (Objet central présent ; Objet central absent) et du Temps de présentation des stimuli (33 ms ; 100 ms). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

1.4. Discussion

Les résultats de l'**Expérience 6** ont montré que les performances de catégorisation des participants s'amélioraient (i.e. que les pourcentages d'erreurs moyens et les temps de réponses correctes moyens diminuaient) à mesure que l'information était présentée en vision périphérique. Nous avons également observé que les taux d'erreurs obtenues en vision périphérique (Anneau 1, Anneau 2 et Anneau 3 de la Session 2) ne différaient pas de ceux obtenus lors de la catégorisation de la scène entière (Disque entier de la Session 1). Ce dernier résultat diffère de celui de l'**Expérience 4** où seules les performances obtenues pour l'anneau le plus périphérique (Anneau 4) ne différaient pas de celles obtenues pour la catégorisation de la scène entière. Cette différence peut toutefois s'expliquer par ce nouveau contrôle du facteur de magnification corticale dans la présente expérience. En effet, la largeur des Anneaux 1 et 2 était plus grande dans l'**Expérience 6** que dans l'**Expérience 4**, tandis que la largeur des Anneaux 3 et 4 était plus petite. Dans tous les cas, nos résultats restent cohérents avec les travaux précédents (Larson & Loschky, 2009 ; Loschky et al., 2019) indiquant que la vision périphérique joue un rôle plus important que la vision centrale pour catégoriser rapidement une scène.

Dans cette expérience, nous avons une nouvelle fois testé si la supériorité de la vision périphérique sur la vision centrale observée par Larson et Loschky (2009) pouvait être expliquée par l'absence d'informations sémantiquement pertinentes en vision centrale. Les résultats de la Session 1 (catégorisation des Disques entiers) ont montré que la simple présence d'un objet central sémantiquement relié à la catégorie de la scène n'améliorait pas les performances de catégorisation. Ainsi, les informations en vision centrale - même pertinentes - seraient moins considérées pour catégoriser une scène lorsque toutes les informations sont disponibles. Ce résultat corrobore celui que nous avons observé dans l'**Expérience 4** et suggère que le système visuel se baserait en priorité sur les informations en provenance de la périphérie qui seraient suffisantes pour catégoriser des scènes d'intérieur ou d'extérieur. Toutefois, nous pouvons à nouveau noter les taux d'erreurs extrêmement bas (effet plafond) qui amènent à penser que la tâche de catégorisation était trop facile dans la Session 1, augmentant ainsi la difficulté à détecter l'effet attendu. Une fois de plus, dans la Session 2, les taux d'erreurs obtenus dans la condition de catégorisation du Disque central étaient significativement plus importants que ceux obtenus dans la condition de catégorisation de la scène entière (Disque entier de la Session 1) et dans ce cas, la présence d'un objet sémantiquement relié à la catégorie de la scène améliorait significativement les performances de catégorisation du Disque central.

Par ailleurs, dans la Session 2, nous avons observé que les participants étaient plus précis et plus rapides pour catégoriser le disque central lorsqu'il y avait un objet sémantiquement relié au centre. Cependant, les performances observées pour les anneaux périphériques demeuraient meilleures que celles observées pour le disque central (les performances s'amélioraient à mesure que l'excentricité augmentait indépendamment de la présence d'un objet central dans la scène). Ces résultats sont cohérents avec les résultats de l'**Expérience 4** et avec notre hypothèse selon laquelle l'efficacité de la vision centrale repose en partie sur la présence d'un contenu visuel spécifique, comme un objet pertinent à catégoriser, tout en démontrant la supériorité de

la vision périphérique sur la vision centrale indépendamment de leurs contenus informatifs respectifs.

Enfin, contrairement à nos attentes, le temps de présentation des stimuli n'interagissait pas avec l'excentricité. Ce résultat contraste avec celui observé par Larson et al. (2014) et suggère plutôt que l'information visuelle est automatiquement et très rapidement extraite de la vision périphérique pour reconnaître l'essentiel d'une scène. Les divergences que nous observons entre nos résultats et ceux de Larson et al. (2014) peuvent être dues à des différences dans la procédure expérimentale. Tout d'abord, dans leur expérience, les auteurs ont utilisé un paradigme Fenêtre-Scotome. Afin de contrôler les effets confondus inhérentes aux propriétés naturelles de la vision centrale et périphérique (magnification corticale en faveur de la vision centrale vs. quantité d'informations visibles en faveur de la vision périphérique), les auteurs ont construit leurs stimuli en utilisant un rayon critique, c'est-à-dire le rayon qui produit des performances égales dans les deux conditions Fenêtre et Scotome. Dans leur expérience, ce rayon critique était de $5,54^\circ$. Par conséquent, leur stimulus Fenêtre était un disque central d'un rayon de $5,54^\circ$ d'angle visuel et leur stimulus Scotome était un anneau d'un rayon interne de $5,54^\circ$ et d'un rayon externe de $10,95^\circ$. Or, en nous appuyant sur l'équation empiriquement dérivée de la mesure rétinotopique (Wu et al., 2012), nous avons calculé que la surface corticale activée par un stimulus Fenêtre était estimée à 62% (344 mm^3) et celle d'un stimulus Scotome était estimée à 38% (213 mm^3). Par conséquent, le stimulus Fenêtre était surreprésenté au niveau cortical par rapport au stimulus Scotome. De plus, dans l'étude de Larson et al. (2014), le temps de traitement du stimulus était manipulé à l'aide d'un masque. Le stimulus était toujours présenté pendant 24 ms, suivi d'un intervalle inter-stimulus présenté pendant 0 ms, 71 ms, 165 ms, 295 ms ou 353 ms, puis d'un masque présenté pendant 24 ms. Après le masque, il y avait un écran blanc présenté pendant 750 ms, puis un écran avec le nom de la catégorie (e.g., plage) présenté jusqu'à ce que les participants décident si le stimulus présenté correspondait à cette catégorie. Par conséquent, les participants devaient systématiquement attendre le dernier écran avec le nom de la catégorie avant de répondre. Compte tenu du long intervalle de temps entre la stimulation visuelle et la réponse, il est possible que des processus plus attentionnels aient été impliqués, contrairement à notre paradigme dans lequel nous privilégions une réponse dès l'apparition du stimulus afin d'évaluer des processus plus automatiques.

Dans l'ensemble, les résultats de l'**Expérience 6** sont cohérents avec ceux de l'**Expérience 4** malgré tous les changements méthodologiques que nous avons effectués. Ils vont donc dans le sens d'une plus grande efficacité et utilité de la vision périphérique par rapport à la vision centrale pour catégoriser rapidement des scènes (en tant que scènes d'intérieur et d'extérieur) et cela malgré la faible acuité visuelle inhérente à la vision périphérique. Ainsi, cette expérience permet de confirmer à nouveau que le traitement de l'information de basse résolution spatiale (BFS) disponible en vision périphérique serait privilégiée lors de la catégorisation rapide de scènes.

Dans la continuité de cette expérience, nous avons mené une nouvelle expérience (**Expérience 7**) dans le but de reproduire l'**Expérience 5** en considérant tous les changements méthodologiques. Plus précisément, l'**Expérience 7** avait pour objectif de déterminer si l'analyse rapide d'une information grossière disponible en vision périphérique pourrait permettre une première catégorisation que validerait ou non l'analyse des informations

disponibles en vision centrale (principalement une information détaillée en HFS). Pour tester cette hypothèse, nous avons utilisé des séquences dynamiques permettant d'imposer une analyse *peripheral-to-central* ou *central-to-peripheral* de l'information visuelle, en assemblant les stimuli utilisés dans la Session 2 de l'**Expérience 6** en allant respectivement du stimulus le plus périphérique au stimulus central ou en allant du stimulus central au stimulus le plus périphérique. En utilisant les stimuli de cette expérience, nous avons pu à nouveau considérer l'influence de la présence d'un contenu sémantique pertinent en vision centrale, mais nous avons aussi pu contrôler le facteur de magnification corticale. Dans cette expérience, nous avons également manipulé le temps de présentation des séquences afin de tester l'hypothèse que le traitement *peripheral-to-central* serait d'autant plus privilégié que le système visuel dispose de peu de temps pour traiter l'information. Les séquences étaient présentées soit pendant 500 ms (chaque stimulus dans la séquence étant présenté pendant 100 ms, comme dans l'**Expérience 6**), soit 165 ms (chaque stimulus était présenté pendant 33 ms de façon à ce que la séquence totale ait la même durée de présentation qu'un seul stimulus de l'**Expérience 6**).

2. **Expérience 7 - Catégorisation de larges scènes dynamiques en expansion ou en contraction**

2.1. Méthode

2.1.1. Participants

Trente-deux étudiants de l'Université Grenoble Alpes (24 femmes ; âge moyen : 19,66 \pm 1,66) ont participé à cette expérience. Ils ont été répartis en deux groupes ($n = 16$) qui devaient respectivement catégoriser les séquences présentées pendant 165 ms ou 500 ms. La taille de l'échantillon a été calculée comme dans l'**Expérience 6**. Nous avons obtenu une taille d'échantillon minimale de 16 participants par groupe expérimental, soit un total de 32 participants. Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Tous les participants ont donné leur consentement libre et éclairé avant de participer à cette expérience qui a été menée en conformité avec le code d'éthique de l'*American Psychological Association* (APA) et approuvée par un comité et approuvée par un comité d'éthique (CERGA, Comité d'éthique de la recherche Grenoble Alpes, IRB00010290-2018-10-16-50). A la fin de l'expérience, chaque participant recevait un bon de 0,25 point par quart d'heure d'expérience à valoir lors des examens universitaires.

2.1.2. Stimuli et procédure

Les stimuli qui composaient les séquences dynamiques dans l'**Expérience 7** étaient les mêmes que ceux utilisés dans la Session 2 de l'**Expérience 6** (Disque central, Anneau 1,

Anneau 2, Anneau 3 et Anneau 4, Figure 3.10b). Ainsi, nous avons utilisé 80 stimuli (40 scènes d'intérieur et 40 scènes d'extérieur). A partir du Disque central et des quatre anneaux de chaque scène, nous avons créé deux types de séquences (Figure 3.13) : séquences *peripheral-to-central* et séquences *central-to-peripheral*. Pour les séquences *peripheral-to-central*, les cinq stimuli étaient assemblés en allant de l'information périphérique à l'information centrale (dans l'ordre : l'Anneau 4, l'Anneau 3, l'Anneau 2, l'Anneau 1, puis le Disque central). Pour les séquences *central-to-peripheral*, les cinq stimuli étaient assemblés en allant de l'information centrale à l'information périphérique (dans l'ordre : le Disque central, puis l'Anneau 1, l'Anneau 2, l'Anneau 3 et l'Anneau 4). L'expérience a été programmée avec le logiciel E-Prime 2.0 (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) et les stimuli étaient présentés sur le même écran de 30 pouces que dans l'Expérience 6. Afin de respecter la taille angulaire des stimuli à 30° d'angle visuel, les participants étaient installés à une distance de 70 cm de l'écran, le menton positionné sur une mentonnière.

L'expérience a été réalisée dans un box expérimental du Laboratoire de Psychologie et NeuroCognition (LPNC) de l'Université de Grenoble Alpes, dans les mêmes conditions expérimentales que l'Expérience 6. Les participants étaient répartis dans deux groupes (n = 16 dans chaque groupe) qui devaient respectivement catégoriser les séquences présentées pendant 165 ms (chaque stimulus étant présenté pendant 33 ms, comme le temps de présentation le plus court de l'Expérience 6) ou pendant 500 ms (chaque stimulus étant présenté pendant 100 ms, comme le temps de présentation le plus long de l'Expérience 6). Les séquences *peripheral-to-central* et *central-to-peripheral* étaient présentées aléatoirement. Chaque essai commençait par un point de fixation central (noir sur un fond gris de luminance 0,5) présenté pendant 500 ms, suivi d'une séquence de cinq stimuli (séquence *peripheral-to-central* ou *central-to-peripheral*) présentée pendant 165 ms ou 500 ms selon le groupe expérimental, suivi d'un écran gris présenté pendant 1900 ms au cours duquel les participants pouvaient répondre. Notons que contrairement à l'Expérience 5, nous n'avons pas présenté de masque après la présentation de chaque stimulus qui compose la séquence afin d'éviter un effet de *metaccontrast masking* du masque sur le dernier stimulus de la séquence (Alpern, 1952 ; Breitmeyer et al., 2000, 2006 ; Enns & Lollo, 2000). Pour rappel, il s'agit d'un effet de masquage réduisant la visibilité d'un stimulus cible présenté brièvement, par la présence d'un second stimulus (le masque) qui est présenté spatialement autour du premier stimulus cible. Quel que soit le groupe expérimental, les participants devaient catégoriser la séquence comme étant soit une scène d'intérieur, soit une scène d'extérieur par l'intermédiaire des touches 1 et 2 du clavier d'ordinateur placé sur l'axe sagittal médian des participants. La moitié des participants de chaque groupe appuyait sur la touche 1 avec l'index de la main droite pour la catégorie Intérieur et sur la touche 2 avec le majeur de la main droite pour la catégorie Extérieur. Les touches de réponse étaient contrebalancées pour l'autre moitié des participants. Les participants avaient pour consigne de répondre le plus correctement et rapidement possible, dès l'apparition du premier stimulus de la séquence. L'expérience comprenait 160 essais (pour chaque type de séquence *peripheral-to-central* et *central-to-peripheral* : 20 scènes d'intérieur et 20 scènes d'extérieur avec un objet au centre, 20 scènes d'intérieur et 20 scènes d'extérieur sans objet au centre). L'expérience durait environ 8 minutes, avec une pause au bout de 100 essais. Pour chaque essai, l'exactitude de la réponse et le temps de réponse (en millisecondes) étaient enregistrés. Les participants étaient

entraînés au préalable avec des stimuli construits à partir de photographies non utilisées lors de l'expérience (20 essais).

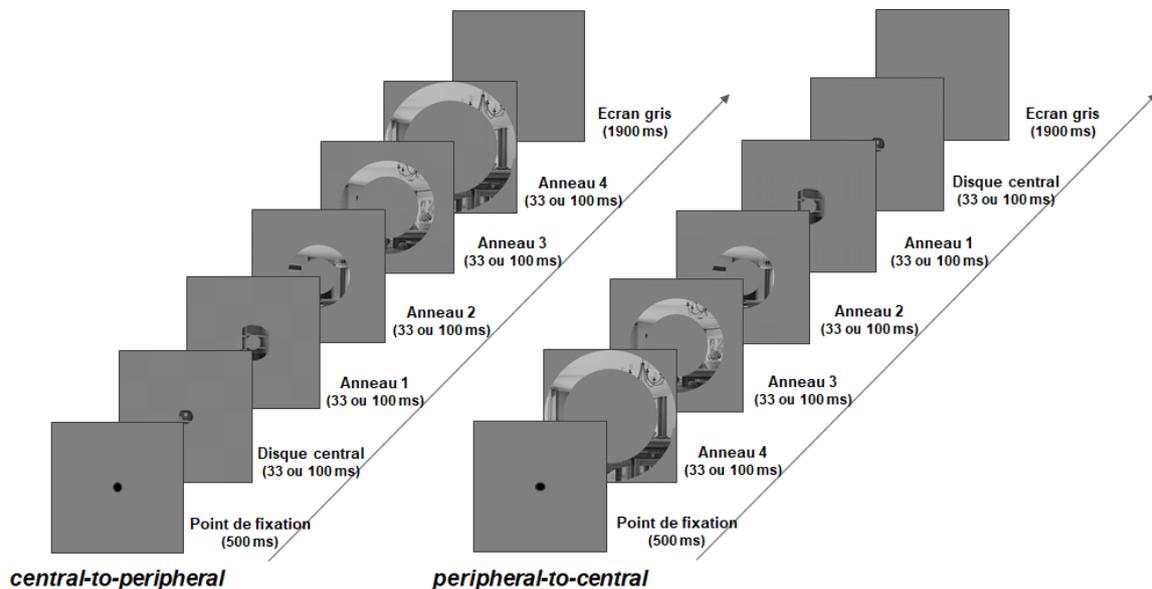


Figure 3.13 : Schéma de la procédure expérimentale de l'Expérience 7. Le Disque central, l'Anneau 1, l'Anneau 2, l'Anneau 3 et l'Anneau 4 sont assemblés du centre vers la périphérie (séquences *central-to-peripheral*) ou de la périphérie vers le centre (séquences *peripheral-to-central*). Les séquences étaient présentées pendant 165 ms (chaque stimulus était présenté pendant 33 ms) ou 500 ms (chaque stimulus était présenté pendant 100 ms). Les participants devaient catégoriser les séquences le plus rapidement et le plus précisément possible, dès l'apparition du premier stimulus de la séquence.

2.2. Hypothèses

Les hypothèses étaient identiques à celles de l'Expérience 5, à savoir que nous nous attendions à ce que les séquences *peripheral-to-central* soient catégorisées plus rapidement et plus précisément que les séquences *central-to-peripheral* dans les deux conditions de temps de présentation. Par ailleurs, en partant de l'hypothèse qu'une stratégie d'analyse *coarse-to-fine* prédomine lors de la catégorisation de scènes, nous nous attendions à ce que l'avantage des séquences *peripheral-to-central* soit plus prononcé pour les séquences présentées pendant 165 ms que pendant 500 ms. En effet, les informations en BFS disponibles en vision périphérique pourraient être davantage prises en compte lors de présentations courtes, tandis que l'analyse plus lente des informations en HFS disponibles en vision centrale pourrait prendre l'avantage lors de présentations plus longues.

Enfin, sur la base des résultats de l'Expérience 6 indiquant que la présence d'un objet central facilite la catégorisation de scènes, nous nous attendions à ce que la présence d'un objet central facilite la catégorisation des séquences *central-to-peripheral* par rapport aux séquences *peripheral-to-central*. En d'autres termes, nous nous attendions à une interaction entre la séquence (*peripheral-to-central* vs *central-to-peripheral*) et la présence d'un objet central (objet central présent vs. objet central absent).

2.3. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale, les pourcentages d'erreurs moyens (%ERm) et les temps de réponses correctes moyens en millisecondes (TRm) ont été calculés. Afin de réduire l'effet des valeurs extrêmes sur les résultats, les temps de réponses supérieurs ou inférieurs à trois fois l'intervalle interquartile ont été supprimés. A l'aide de cette méthode, nous avons supprimé seulement 0,25% des temps de réponse. L'analyse des données a été effectuée à l'aide du logiciel Statistica 13.3 et du logiciel RStudio. Le seuil de significativité a été fixé à 0,05. La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2).

Une analyse de la variance à mesure répétées (ANOVA) de type $2 \times 2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm avec comme facteur intra-sujet la Séquence (*peripheral-to-central* ; *central-to-peripheral*), la Présence d'un objet central (Objet central présent ; Objet central absent) et la Catégorie de la scène (Intérieur ; Extérieur) et comme facteur inter-sujet le Temps de présentation de chaque séquence (165 ms ; 500 ms). Les pourcentages d'erreurs moyens (%ERm) et leurs erreurs-types ainsi que les temps de réponses moyens (TRm) et leurs erreurs-types sont présentés dans le Tableau 3.6.

Tableau 3.6 : Résultats de l'Expérience 7. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Séquence (*peripheral-to-central* ; *central-to-peripheral*), de la Présence d'un objet central (Objet central présent ; Objet central absent), du Temps de présentation de chaque séquence (165 ms ; 500 ms) et de la Catégorie de la scène (Intérieur ; Extérieur).

			Objet central présent		Objet central absent	
			%ERm	TRm	%ERm	TRm
<i>peripheral-to-central</i>	165 ms	Intérieur	3,45 ± 1,60	625 ± 22	5,07 ± 1,61	638 ± 20
		Extérieur	4,46 ± 1,51	635 ± 20	3,45 ± 1,67	639 ± 23
	500 ms	Intérieur	6,56 ± 1,60	598 ± 22	7,37 ± 1,61	578 ± 20
		Extérieur	7,81 ± 1,51	595 ± 20	6,60 ± 1,67	598 ± 20
<i>central-to-peripheral</i>	165 ms	Intérieur	4,08 ± 1,70	663 ± 25	4,10 ± 1,25	653 ± 22
		Extérieur	5,07 ± 1,69	651 ± 23	2,83 ± 1,23	656 ± 20
	500 ms	Intérieur	7,89 ± 1,70	646 ± 25	5,36 ± 1,25	664 ± 22
		Extérieur	7,53 ± 1,69	635 ± 23	5,31 ± 1,23	663 ± 20

- Analyse des pourcentages d'erreurs moyens

L'analyse de la variance (ANOVA) menée sur les %ERm n'a ni montré d'effet principal de la Séquence ($F(1,30) < 1$; *central-to-peripheral* : $5.27 \pm 1.50\%$; *peripheral-to-central* : $5.71 \pm 1.64\%$; Figure 3.14a), ni d'effet principal du Temps de présentation ($F(1,30) = 3.16$, $p = .085$) et ces deux facteurs n'interagissaient pas entre eux ($F(1,30) < 1$). De plus, il n'y avait ni d'effet principal de la Présence d'un objet central ($F(1,30) = 1.56$, $p = .220$), ni d'interaction de ce facteur avec la Séquence ($F(1,30) = 3.09$, $p = .089$) et ces deux facteurs n'interagissaient pas

avec le Temps de présentation ($F(1,30) < 1$). Enfin, il n'y avait ni d'effet principal de la Catégorie ($F(1,30) < 1$), ni d'interaction entre la Catégorie et les autres facteurs manipulés (Séquence \times Catégorie : $F(1,30) < 1$; Séquence \times Temps de présentation \times Catégorie : $F(1,30) < 1$; Séquence \times Présence d'un objet central \times Catégorie : $F(1,30) < 1$; Séquence \times Présence d'un objet central \times Temps de présentation \times Catégorie : $F(1,30) < 1$).

- **Analyse des temps de réponses correctes moyens**

L'analyse de la variance (ANOVA) menée sur les TRm a révélé un effet principal de la Séquence ($F(1,30) = 65,88$, $p < .001$, $\eta^2 = .69$; Figure 3.14b). Les participants étaient en moyenne plus rapides pour catégoriser des séquences *peripheral-to-central* (613 ± 29 ms) par rapport aux séquences *central-to-peripheral* (654 ± 30 ms). L'effet principal du Temps de présentation n'était pas significatif ($F(1,30) < 1$) mais ce facteur interagissait avec la Séquence ($F(1,30) = 14,52$, $p < .001$). Les comparaisons planifiées ont montré que les participants catégorisaient les séquences *peripheral-to-central* plus rapidement que les séquences *central-to-peripheral* à la fois lorsqu'elles étaient présentées pendant 165 ms (*peripheral-to-central* : 634 ± 41 ms ; *central-to-peripheral* : 656 ± 43 ms ; $F(1,30) = 9,27$, $p < .005$) et pendant 500 ms (*peripheral-to-central* : 592 ± 41 ms ; *central-to-peripheral* : 652 ± 43 ms ; $F(1,30) = 71,12$, $p < .001$). Plus précisément, la différence entre les deux types de séquences était plus grande lorsqu'elles étaient présentées pendant 500 ms que pendant 165 ms. Par ailleurs, l'effet principal de la Présence d'un objet central n'était pas significatif ($F(1,30) = 2,03$, $p < .164$) et ce facteur n'interagissait pas avec la Séquence ($F(1,30) = 2,08$, $p = .162$). Cependant, l'analyse a révélé une interaction significative entre la Séquence, le Temps de présentation et la Présence d'un objet central ($F(1,30) = 7,75$, $p = .009$, $\eta^2 = .20$). Nous avons testé l'interaction entre la Séquence et le Temps de présentation pour chaque condition de la Présence d'un objet central. Cette interaction était significative lorsqu'il n'y avait pas d'objet central ($F(1,30) = 19,44$, $p < .001$) mais pas lorsqu'il y avait un objet central ($F(1,30) = 2,13$, $p = .154$). Plus précisément, lorsqu'il y avait un objet central, les participants catégorisaient les séquences *peripheral-to-central* plus rapidement que les séquences *central-to-peripheral* indépendamment du Temps de présentation des stimuli (500 ms : *peripheral-to-central* : 597 ± 29 ms, *central-to-peripheral* : 640 ± 33 ms ; 165 ms : *peripheral-to-central* : 630 ± 29 ms, *central-to-peripheral* : 657 ± 33 ms). Cependant, lorsqu'il n'y avait pas d'objet central, les participants catégorisaient plus rapidement les séquences *peripheral-to-central* que les séquences *central-to-peripheral* uniquement lorsqu'elles sont présentées pendant 500 ms (500 ms : *peripheral-to-central* : 588 ± 30 ms, *central-to-peripheral* : 663 ± 29 ms, $F(1,30) = 63,37$, $p < .001$; 165 ms : *peripheral-to-central* : 638 ± 30 ms, *central-to-peripheral* : 655 ± 29 ms, $F(1,30) = 2,98$, $p = .095$). Ces résultats suggèrent que l'avantage des séquences *peripheral-to-central* était réduit lorsqu'il n'y avait pas d'objet central pour les séquences présentées pendant 165 ms. Par ailleurs, l'effet principal de la Catégorie n'était pas significatif ($F(1,30) < 1$), mais l'interaction entre la Catégorie et la Séquence était significative ($F(1,30) = 4,85$, $p = .035$, $\eta^2 = .14$). Les comparaisons planifiées montraient que la différence entre les séquences *central-to-peripheral* et les séquences *peripheral-to-central* était significative à la fois pour la catégorie Intérieur

(*central-to-peripheral* : 657 ± 23 ms ; *peripheral-to-central* : 610 ± 21 ms ; $F(1,30) = 53,52$, $p < .001$) et pour la catégorie Extérieur (*central-to-peripheral* : 651 ± 21 ms ; *peripheral-to-central* : 617 ± 21 ms ; $F(1,30) = 46,84$, $p < .001$). Toutefois, cette fois-ci, l'avantage des séquences *peripheral-to-central* était plus grand pour la catégorie Extérieur que pour la catégorie Intérieur. Enfin, la Catégorie n'interagissait pas avec les autres facteurs manipulés (Séquence \times Catégorie : $F(1,30) < 1$; Séquence \times Temps de présentation \times Catégorie : $F(1,30) < 1$; Séquence \times Présence d'un objet central \times Catégorie : $F(1,30) < 1$; Séquence \times Présence d'un objet central \times Temps de présentation \times Catégorie : $F(1,30) = 1,77$, $p = .193$).

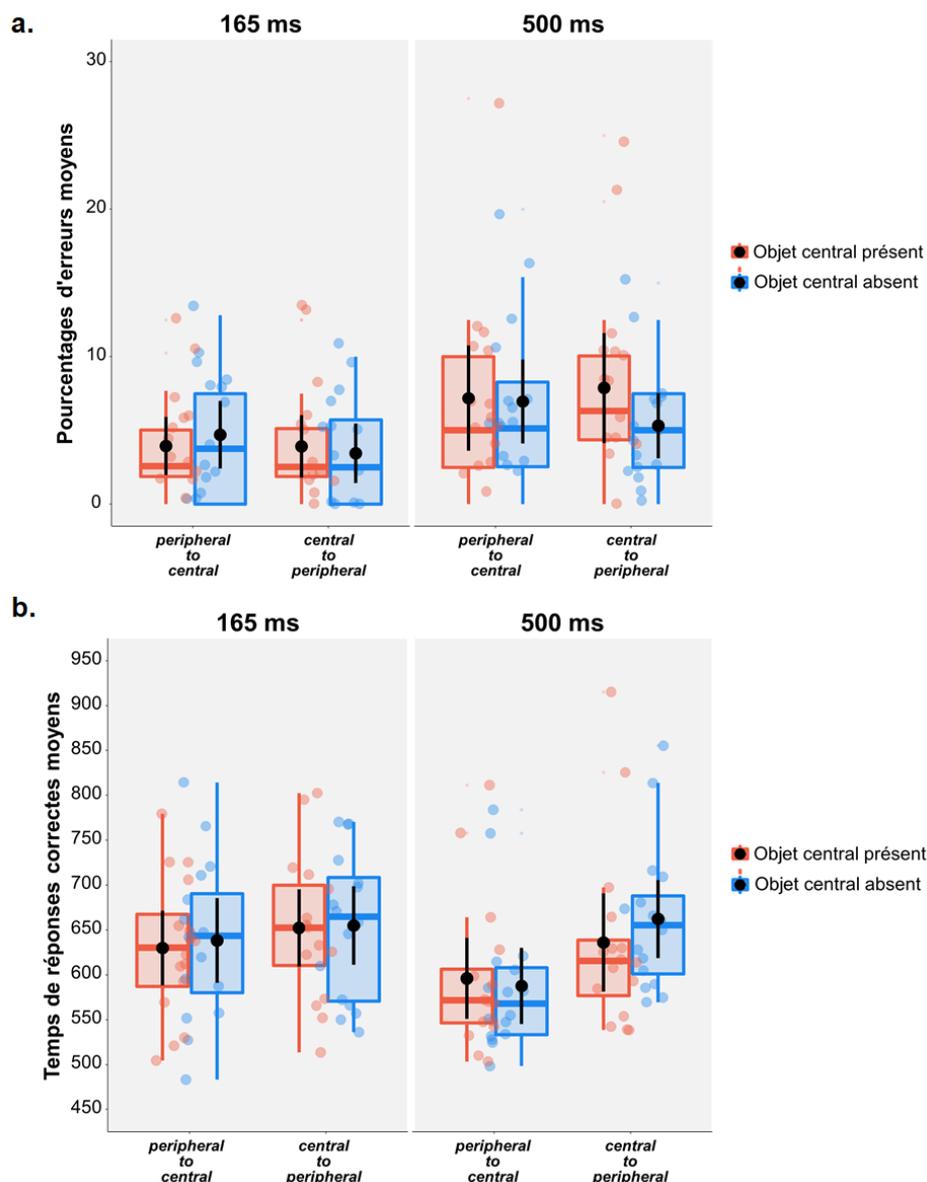


Figure 3.14 : Résultats de l'Expérience 7 : (a) Pourcentages d'erreurs moyens et (b) Temps de réponses correctes moyens en ms en fonction de la Séquence (*peripheral-to-central* ; *central-to-peripheral*), de la Présence d'un objet central (Objet central présent ; Objet central absent), et du Temps de présentation (165 ms ; 500 ms). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

2.4. Discussion

L'**Expérience 7** avait pour but de reproduire l'**Expérience 5** en testant si l'analyse rapide des informations disponibles en vision périphérique pouvait permettre une première catégorisation de la scène qui serait ensuite validée par l'analyse des informations disponibles en vision centrale. Pour cela, nous avons utilisé des séquences dynamiques composées du disque central et des quatre anneaux assemblés soit du disque central à l'anneau le plus périphérique (Anneau 4) afin d'imposer une analyse de l'information visuelle *central-to-peripheral*, soit de l'anneau le plus périphérique (Anneau 4) au disque central afin d'imposer une analyse de l'information visuelle *peripheral-to-central*. Nous avons également testé si la supériorité de la vision périphérique était améliorée par des temps de présentations plus courts. Ainsi, les séquences étaient présentées pendant 165 ms ou 500 ms de façon à ce que chaque stimulus qui compose la séquence soit respectivement présenté pendant 33 ms ou 100 ms. Les résultats ont effectivement montré un avantage des séquences *peripheral-to-central*, mais contre toute attente, il était plus important pour le temps de présentation le plus long (500 ms). Nous avons observé que ce résultat était en fait dû à un effet inattendu lorsqu'il n'y avait pas d'objet central pour le temps de présentation le plus court (165 ms). Ce résultat sera discuté plus en détail dans la Discussion générale de ce Chapitre.

Pris dans leur ensemble, les résultats des **Expériences 5** et **7** suggèrent un avantage de la vision périphérique sur la vision centrale lors de la catégorisation rapide des scènes. Cependant, comme nous l'a suggéré Lester Loschky lors de son expertise de l'article, l'avantage des séquences *peripheral-to-central* pourrait être dû à un effet de masquage visuel. En effet, les anneaux périphériques présentés après le disque central (séquences *central-to-peripheral*) pourraient avoir provoqué un masquage latéral rétroactif du traitement d'un stimulus périphérique sur un stimulus central adjacent. Cet effet connu sous le nom de masquage par métacontraste ou *metacontrast masking* est un type de *visual backward masking* (Alpern, 1952 ; Breitmeyer & Ogmen, 2000 ; Breitmeyer et al., 2006 ; Enns & Lollo, 2000). Cet effet se caractérise par une baisse de la visibilité d'une cible, non pas lorsque le masque se superpose à la cible, mais lorsqu'il épouse étroitement les contours de la cible. Cet effet de masquage apparaîtrait uniquement sous certaines contraintes temporelles (Alpern, 1952). La visibilité de la cible serait en fait réduite lorsque le masque est présenté juste après la cible. Dans nos expériences, cet effet de masquage concernerait donc uniquement les séquences *central-to-peripheral*, où chaque stimulus de la séquence est suivi par un anneau plus large qui l'entoure. Par conséquent, l'Anneau 1 qui entoure le Disque central pourrait réduire sa visibilité lorsqu'il est présenté juste après, puis l'Anneau 2 pourrait réduire celle de l'Anneau 1 et ainsi de suite. Habituellement, la réduction de la visibilité de la cible dépend de la durée entre la présentation de la cible et celle du masque (*Stimulus Onset Asynchrony*, SOA) et en suivant une fonction en forme de U. Le masquage serait plus important pour des SOA relativement courts (allant de 30 à 80 ms ; Breitmeyer et al., 2006). Dans nos Expériences, ils devraient donc apparaître surtout dans le cas où les séquences sont présentées très rapidement (150 ms dans l'**Expérience 5** et 165 ms dans l'**Expérience 7**). Pourtant, les résultats de l'**Expérience 7** ont montré que l'avantage des séquences *peripheral-to-central* était justement plus important pour les séquences les plus longues (500 ms). De plus, si des anneaux périphériques peuvent avoir

provoqué un effet de masquage par métacontraste sur le traitement du disque central, il devient difficile d'expliquer pourquoi cet effet se manifesterait lorsqu'un objet central est présent (condition pour laquelle nous avons observé un avantage des séquences *peripheral-to-central*), mais pas lorsqu'il est absent (condition pour laquelle nous n'avons pas observé l'avantage des séquences *peripheral-to-central*). Cependant, compte tenu de l'existence d'un tel biais méthodologique, nous avons mené une expérience contrôle (**Expérience 8**) afin de quantifier cet effet de masquage lors de la catégorisation des séquences *central-to-peripheral* et *peripheral-to-central* et d'évaluer si cet effet est réellement plus important pour un stimulus périphérique qui entoure un stimulus central que l'inverse.

Dans cette expérience contrôle (**Expérience 8**), nous avons présenté les stimuli de nos expériences que nous avons masqué ou non (Figure 3.15). Dans une condition sans masque, les participants devaient catégoriser soit le Disque central, soit l'Anneau 4 (anneau le plus périphérique) comme étant soit une scène d'intérieur, soit une scène d'extérieur. Dans une autre condition avec masque, les participants devaient catégoriser soit des séquences composées de cinq stimuli assemblés du centre à la périphérie dans lesquelles le Disque central d'une scène était présenté en premier, suivi par les Anneaux 1, 2, 3 et 4 d'un masque (séquences *central-to-mask*), soit des séquences composées de cinq stimuli assemblés de la périphérie au centre dans lesquelles l'Anneau 4 d'une scène était présenté en premier, suivi par les Anneaux 3, 2, 1 et le Disque central d'un masque (séquences *peripheral-to-mask*). Ici aussi, nous avons fait varier le temps de présentation des stimuli à catégoriser, et donc le temps de présentation de la séquence de masques.

Notre objectif était ainsi de tester si la présence d'un masque latéral perturbe réellement les performances de catégorisation d'une scène et d'évaluer si cet effet est réellement plus important lorsqu'un stimulus périphérique entoure un stimulus central que l'inverse.

3. **Expérience 8 – Contrôle d'un effet de masquage latéral rétroactif**

3.1. Méthode

3.1.1. Participants

Trente-deux étudiants de l'Université Grenoble Alpes (28 femmes ; âge moyen : 20,63 \pm 2,68) ont participé à cette expérience. Ils ont été répartis en deux groupes ($n = 16$) qui devaient respectivement catégoriser les stimuli scènes présentés pendant 33 ms ou 165 ms. La taille de l'échantillon a été calculée comme dans l'**Expérience 6**. Nous avons obtenu une taille d'échantillon minimale de 16 participants par groupe expérimental, soit un total de 32 participants. Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Tous les participants ont donné leur consentement libre et éclairé avant de participer à cette expérience qui a été menée en conformité avec le code

d'éthique de l'*American Psychological Association* (APA) et approuvée par un comité et approuvée par un comité d'éthique (CERGA, Comité d'éthique de la recherche Grenoble Alpes, IRB00010290-2018-10-16-50). A la fin de l'expérience, chaque participant recevait un bon de 0,25 point par quart d'heure d'expérience à valoir lors des examens universitaires.

3.1.2. Stimuli et procédure

Les stimuli du Disque central et de l'Anneau 4 étaient exactement les mêmes que ceux utilisés dans les [Expériences 6](#) et [7](#). Nous avons construit les masques sur la base d'une image dont le spectre d'amplitude décroît avec la fréquence spatiale (bruit distribué en $1/F$) et à l'intérieur duquel nous avons découpé un large disque d'un rayon de 15° d'angle visuel (750 pixels). Puis, nous avons construit des stimuli de type masque qui révélaient différentes parties de ce large disque en utilisant l'équation empiriquement dérivée d'une mesure rétinotopique en IRMf (Wu et al., 2012). Comme dans les [Expériences 6](#) et [7](#), nous avons créé cinq stimuli d'excentricités différentes mais avec la même surface d'activation sur le cortex visuel primaire (20%, i.e. $\sim 136 \text{ mm}^3$). Tous les stimuli ont été normalisés de façon à obtenir une luminance moyenne de 0,5 et un contraste RMS (*Root Mean Square*) de 0,2 pour des valeurs d'intensité des pixels comprises entre 0 et 1 (i.e. une luminance moyenne de 128 et un contraste RMS de 51 sur une échelle de 256 niveaux de gris). Pour les séquences *central-to-mask*, les cinq stimuli étaient assemblés en allant du centre vers la périphérie : le Disque central d'une scène était présenté en premier, puis l'Anneau 1, l'Anneau 2, l'Anneau 3 et l'Anneau 4 du masque. Pour les séquences *peripheral-to-mask*, les cinq stimuli étaient assemblés en allant de la périphérie vers le centre : l'Anneau 4 d'une scène était présenté en premier, puis l'Anneau 3, l'Anneau 2, l'Anneau 1 et le Disque central du masque. L'expérience a été programmée avec le logiciel E-Prime 2.0 (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) et les stimuli étaient présentés sur le même écran de 30 pouces que dans les [Expériences 6](#) et [7](#). Afin de respecter la taille angulaire des stimuli à 30° d'angle visuel, les participants étaient installés à une distance de 70 cm de l'écran, le menton positionné sur une mentonnière.

L'expérience a été réalisée dans un box expérimental du Laboratoire de Psychologie et NeuroCognition (LPNC) de l'Université de Grenoble Alpes, dans les mêmes conditions expérimentales que les [Expériences 6](#) et [7](#). Tous les participants étaient répartis dans deux groupes ($n = 16$ dans chaque groupe) qui devaient respectivement catégoriser les stimuli présentés pendant 33 ms ou pendant 100 ms. Ainsi, les stimuli sans masque (Disque central ; Anneau 4) étaient présentés pendant 33 ou pendant 100 ms et les séquences (*central-to-mask* ; *peripheral-to-mask*) étaient présentées pendant 165 ms (le stimulus à catégoriser ainsi que chaque masque de la séquence était présenté pendant 33 ms) ou pendant 500 ms (le stimulus à catégoriser ainsi que chaque masque était présenté pendant 100 ms). La Figure [3.15](#) montre le schéma de la procédure expérimentale d'un essai pour chaque condition expérimentale. Tous les stimuli expérimentaux étaient présentés aléatoirement aux participants. Chaque essai commençait par un point de fixation central (noir sur un fond gris de luminance 0,5) présenté pendant 500 ms, immédiatement suivi d'un stimulus (Disque central ou Anneau 4) présenté pendant 33 ms ou 100 ms, ou d'une séquence (*central-to-mask* ou *peripheral-to-mask*)

présentée pendant 165 ms ou 500 ms selon le groupe expérimental, puis d'un écran gris présenté pendant 1900 ms au cours duquel les participants pouvaient répondre. Quel que soit le groupe expérimental, les participants devaient catégoriser le stimulus ou la séquence comme étant soit une scène d'intérieur, soit une scène d'extérieur par l'intermédiaire des touches 1 et 2 du clavier d'ordinateur placé sur l'axe sagittal médian des participants. La moitié des participants de chaque groupe appuyait sur la touche 1 avec l'index de la main droite pour la catégorie Intérieur et sur la touche 2 avec le majeur de la main droite pour la catégorie Extérieur. Les touches de réponse étaient contrebalancées pour l'autre moitié des participants. Les participants avaient pour consigne de répondre le plus correctement et rapidement possible, dès l'apparition du stimulus. L'expérience comprenait 320 essais (pour chaque type de stimulus et chaque type de séquence, il y avait 20 scènes d'intérieur et 20 scènes d'extérieur avec un objet au centre, 20 scènes d'intérieur et 20 scènes d'extérieur sans objet au centre). L'expérience durait environ 15 minutes, avec une pause tous les 80 essais. Pour chaque essai, l'exactitude de la réponse et le temps de réponse (en millisecondes) étaient enregistrés. Les participants étaient entraînés au préalable avec des stimuli construits à partir de photographies non utilisées lors de l'expérience (20 essais).

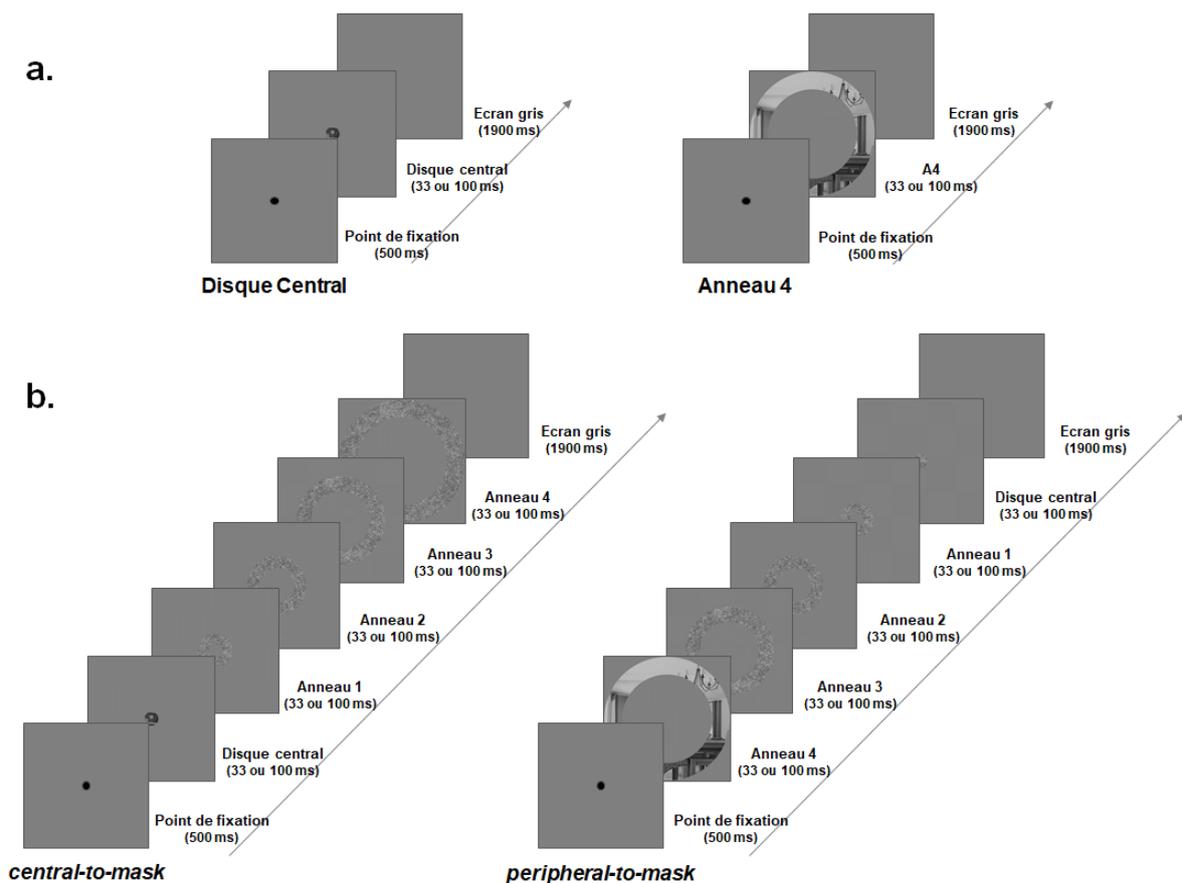


Figure 3.15 : Schéma de la procédure expérimentale de l'Expérience 8. (a) Les participants devaient catégoriser le Disque central ou l'Anneau 4 d'une scène présentée pendant 33 ms ou 100 ms. (b) Les participants devaient catégoriser des séquences *central-to-mask* et *peripheral-to-mask*. Les séquences *central-to-mask* étaient composées de cinq stimuli assemblés du centre vers la périphérie dans laquelle le Disque central d'une scène était présenté en premier, puis l'Anneau 1, l'Anneau 2, l'Anneau 3 et l'Anneau 4.

l'Anneau 4 du masque. Les séquences *peripheral-to-mask* étaient composées de cinq stimuli assemblés de la périphérie vers le centre dans laquelle l'Anneau 4 d'une scène était présenté en premier, puis l'Anneau 3, l'Anneau 2, l'Anneau 1 et le Disque central du masque. Les séquences étaient présentées pendant 165 ms (chaque stimulus était présenté pendant 33 ms), ou 500 ms (chaque stimulus était présenté pendant 100 ms). Les participants devaient répondre le plus correctement et rapidement possible, dès l'apparition du premier stimulus de la séquence.

3.2. Hypothèses

Si le traitement des séquences *central-to-peripheral* et *peripheral-to-central* de l'**Expérience 7** implique un effet de masquage par métacontraste, les performances des participants devraient être moins bonnes lorsque les stimuli de la présente expérience sont masqués que lorsqu'ils ne le sont pas. Par ailleurs, si cet effet de masquage est plus important pour un stimulus périphérique qui entoure un stimulus central (séquences *central-to-peripheral*) que l'inverse (séquences *peripheral-to-central*), l'excentricité du stimulus à catégoriser de la présente expérience devrait interagir avec la présence d'un masque de telle sorte que la différence de performances entre la condition Masque présent et la condition Masque absent devrait être plus grande pour la catégorisation du Disque central que pour la catégorisation de l'Anneau 4. Ainsi, sur la base des résultats observés dans l'**Expérience 7**, si un avantage des séquences *peripheral-to-central* est dû à un effet de masquage par métacontraste, on devrait observer une interaction significative entre l'excentricité du stimulus à catégoriser (Disque central ; Anneau 4) et la présence d'un masque (Masque présent ; Masque absent).

3.3. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale, les pourcentages d'erreurs moyens (%ERm) et les temps de réponses correctes moyens en millisecondes (TRm) ont été calculés. Afin de réduire l'effet des valeurs extrêmes sur les résultats, les temps de réponses supérieurs ou inférieurs à trois fois l'intervalle interquartile ont été supprimés. A l'aide de cette méthode, nous avons supprimé seulement 0,80% des temps de réponse. L'analyse des données a été effectuée à l'aide du logiciel Statistica 13.3 et du logiciel RStudio. Le seuil de significativité a été fixé à 0,05. La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2).

Une analyse de la variance à mesure répétées (ANOVA) de type $2 \times 2 \times 2 \times 2$ a été réalisée sur les valeurs de %ERm et les TRm avec comme facteur intra-sujet l'Excentricité du stimulus à catégoriser (Disque central ; Anneau 4), la Présence d'un masque (Masque présent ; Masque absent), la Présence d'un objet central (Objet central présent ; Objet central absent) et la Catégorie de la scène (Intérieur ; Extérieur) et comme facteur inter-sujet le Temps de présentation des stimuli (33 ms ; 100 ms). Les pourcentages d'erreurs moyens (%ERm) et leurs erreurs-types ainsi que les temps de réponses moyens (TRm) et leurs erreurs-types sont présentés dans le Tableau [3.7](#).

Tableau 3.7 : Résultats de l'**Expérience 8**. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de l'Excentricité des stimuli (Disque central ; Anneau 4), de la Présence d'un objet central (Objet central présent ; Objet central absent), de la Présence d'un masque (Masque présent ; Masque absent), et du Temps de présentation des stimuli (33 ms ; 100 ms).

		33 ms		165 ms		
		%ERm	TRm	%ERm	TRm	
Disque central	Masque présent	Objet central présent	15,16 ± 2,80	773 ± 30	5,50 ± 1,96	806 ± 31
		Objet central absent	30,00 ± 3,34	842 ± 42	21,50 ± 2,60	711 ± 31
	Masque absent	Objet central présent	9,38 ± 1,90	697 ± 27	6,00 ± 2,89	746 ± 36
		Objet central absent	20,31 ± 2,52	787 ± 35	24,33 ± 3,45	686 ± 28
Anneau 4	Masque présent	Objet central présent	14,38 ± 2,62	773 ± 44	5,83 ± 2,70	719 ± 45
		Objet central absent	19,06 ± 2,80	780 ± 44	6,50 ± 2,89	719 ± 45
	Masque absent	Objet central présent	7,34 ± 1,92	725 ± 38	6,67 ± 1,98	701 ± 40
		Objet central absent	8,44 ± 1,79	731 ± 40	4,67 ± 1,85	699 ± 41

Etant donné que les %ERm étaient très élevés pour la catégorisation du Disque central sans objet central (taux d'erreurs supérieurs à 20% ; Tableau 3.7) mais que les TRm étaient très courts dans cette même condition (temps de réponses inférieurs à 650 ms), nous avons d'abord mené une analyse de la variance (ANOVA) sur les %ERm et sur les TRm afin de nous assurer qu'il n'y avait pas de *speed-accuracy trade-off* (c'est à dire, des résultats contraires entre les temps de réponses et les taux d'erreurs) dans cette condition spécifique. Nous avons effectivement observé un *speed-accuracy trade-off*. En effet, pour l'analyse des %ERm, l'interaction entre l'Excentricité et la Présence d'un objet central était significative ($F(1,29) = 175,87, p < .001$). Les comparaisons planifiées ont montré qu'il y avait significativement plus d'erreurs pour le Disque central que pour l'Anneau 4 lorsqu'il n'y avait pas d'objet central ($F(1,29) = 162,69, p < .001$; Disque Central : $24,00 \pm 2,10\%$, Anneau 4: $9,67 \pm 1,65\%$), mais aucune différence significative entre le Disque central et l'Anneau 4 lorsqu'un objet central était présent ($F(1,29) < 1$; Disque central : $9,01 \pm 7,69\%$, Anneau 4 : $8,55 \pm 1,63\%$). En revanche, pour l'analyse des TRm, les performances étaient inversées. L'interaction entre l'Excentricité et la Présence d'un objet central était bien significative ($F(1,29) = 75,67, p < .001$), mais les comparaisons planifiées ont montré cette-fois ci que les participants étaient plus rapides pour catégoriser le Disque central que l'Anneau 4 lorsqu'il n'y avait pas d'objet central ($F(1,29) = 75,67, p < .001$; Disque central : 621 ± 23 ms, Anneau 4 : 668 ± 22 ms) et aucune différence significative entre le Disque central et l'Anneau 4 lorsqu'un objet central était présent ($F(1,29) = 1,95, p = .173$; Disque central : 662 ± 17 ms, Anneau 4 : 675 ± 22 ms). Ainsi, dans la condition spécifique où il n'y avait pas d'objet central, les participants faisaient plus d'erreurs pour catégoriser le Disque central que l'Anneau 4, mais ils étaient plus rapides.

Nous avons donc mené une deuxième analyse sans considérer la condition Objet central absent. Notons à juste titre que dans l'**Expérience 7**, nous avons observé un avantage des séquences *peripheral-to-central* uniquement lorsqu'un objet central était présent. Afin de tester si l'avantage des séquences *peripheral-to-central* de l'**Expérience 7** est effectivement dû à un effet de *metacontrast masking*, il semble donc cohérent de ne considérer que la condition où l'objet central est présent. Ainsi, la nouvelle analyse de la variance (ANOVA) incluait comme

facteur intra-sujet l'Excentricité du stimulus à catégoriser (Disque central ; Anneau 4), la Présence d'un masque (Masque présent ; Masque absent) et la Catégorie de la scène (Intérieur ; Extérieur) et comme facteur inter-sujet le Temps de présentation des stimuli (33 ms ; 100 ms).

- **Analyse des pourcentages d'erreurs moyens**

L'analyse de la variance (ANOVA) menée sur les %ERm n'a pas montré d'effet principal significatif de l'Excentricité ($F(1,29) < 1$; Anneau 4 : $8,55 \pm 1,89\%$; Disque central : $9,01 \pm 2,18\%$), mais l'effet principal de la Présence d'un masque était significatif ($F(1,29) = 12,67$, $p = .001$, $\eta^2 = .30$; Figure 3.16a). Les participants faisaient moins d'erreurs pour catégoriser un stimulus lorsqu'il n'y avait pas de séquence de masque qui le suivait ($7,22 \pm 1,70\%$) que lorsqu'il y en avait une ($10,34 \pm 2,33\%$). Ce résultat suggère clairement un masquage par métacontraste. Cependant, la Présence d'un masque n'interagissait pas avec l'excentricité ($F(1,29) < 1$), ce qui indique que cet effet de masquage n'était pas plus grand lorsqu'un stimulus périphérique entourait un stimulus central que l'inverse. Le Temps de présentation n'interagissait pas avec l'Excentricité et la Présence d'un masque ($F(1,29) < 1$). Enfin, il n'y avait ni d'effet principal de la Catégorie ($F(1,29) < 1$), ni d'interaction entre la Catégorie, l'Excentricité et la Présence d'un masque ($F(1,29) < 1$).

- **Analyse des temps de réponses correctes moyens**

L'analyse de la variance (ANOVA) menée sur les TRm n'a pas montré d'effet principal significatif de l'Excentricité ($F(1,29) = 1,95$, $p = .173$; Anneau 4 : 675 ± 22 ms ; Disque central : 662 ± 17 ms), mais l'effet principal de la Présence d'un masque était à nouveau significatif ($F(1,29) = 45,68$, $p < .001$, $\eta^2 = .61$; Figure 3.16b). Les participants étaient plus rapides pour catégoriser un stimulus lorsqu'il n'y avait pas de séquence de masque qui le suivait (702 ± 32 ms) que lorsqu'il y en avait une (744 ± 36 ms). Encore une fois, ce résultat suggère un effet de masquage par métacontraste. Comme pour l'analyse des %ERm, la Présence d'un masque n'interagissait pas avec l'excentricité ($F(1,29) = 2,73$, $p = .109$), ce qui indique à nouveau que cet effet de masquage n'était pas plus grand lorsqu'un stimulus périphérique entourait un stimulus central que l'inverse. Le temps de présentation n'interagissait pas avec l'Excentricité et la Présence d'un masque ($F(1,29) < 1$). Enfin, l'effet principal de la Catégorie était significatif ($F(1,29) = 4,30$, $p = .047$, $\eta^2 = .13$), à savoir que les participants étaient plus rapides pour la catégorie Extérieur (717 ± 33 ms) que pour la catégorie Intérieur (730 ± 41 ms). La Catégorie interagissait avec l'Excentricité ($F(1,29) = 5,88$, $p = .022$). Les comparaisons planifiées ont montré que les participants catégorisaient plus rapidement le Disque central que l'Anneau 4 pour les scènes d'intérieur (Anneau 4 : 743 ± 31 ms ; Disque central : 717 ± 21 ms ; $F(1,29) = 4,83$, $p = .036$), mais pas pour les scènes d'extérieur ($F(1,29) < 1$). Enfin, la Catégorie n'interagissait pas avec l'Excentricité et la Présence d'un masque ($F(1,29) = 2,47$, $p = .127$).

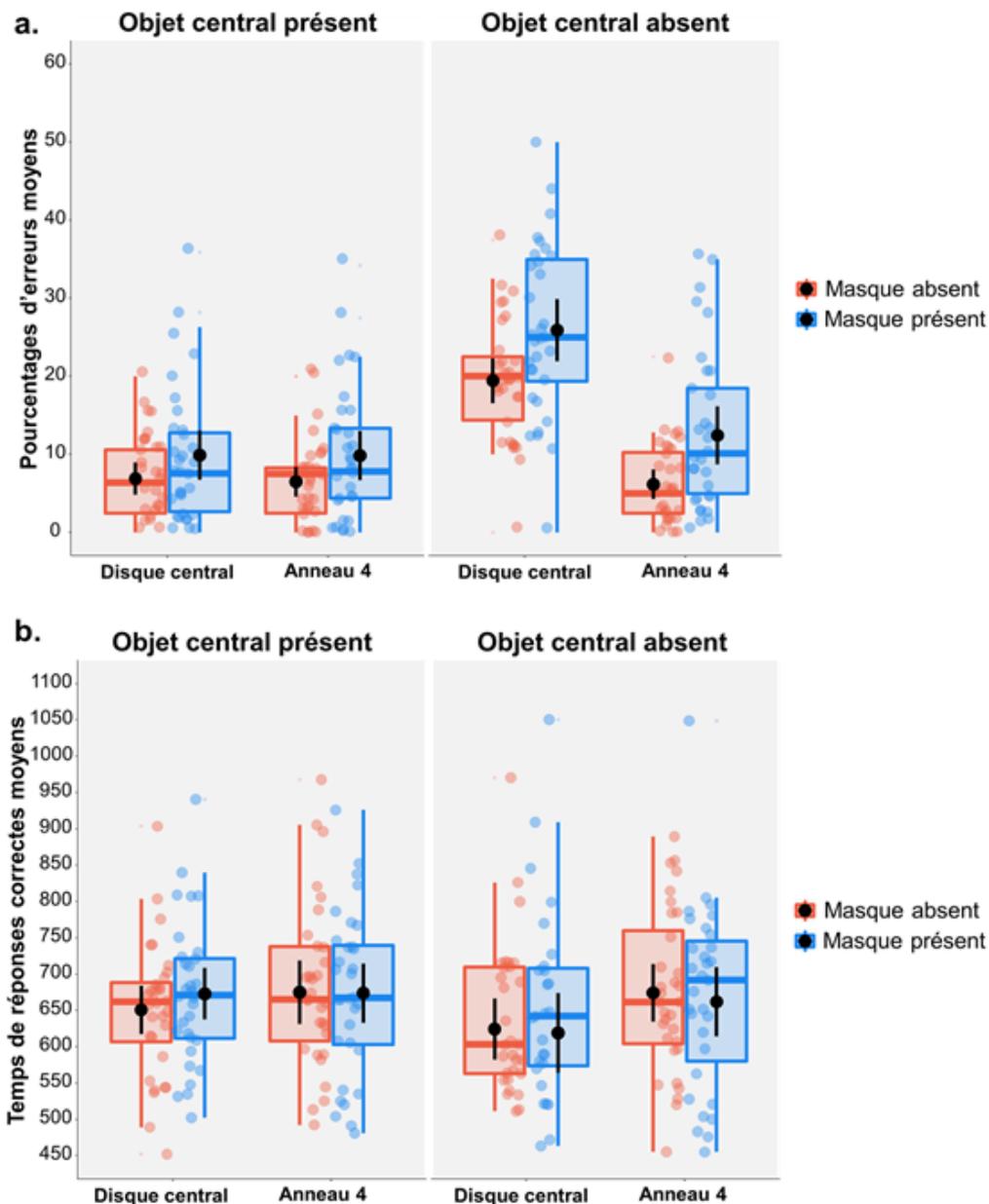


Figure 3.16 : Résultats de l'Expérience 8 : (a) Pourcentages d'erreurs moyens et (b) Temps de réponses correctes moyens en ms en fonction de l'Excentricité des stimuli (Disque central ; Anneau 4), de la Présence d'un objet central (Objet central présent ; Objet central absent), des Stimuli à catégoriser (Disque central ; Anneau 4) et de la Présence d'un masque (Masque absent ; Masque présent). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

4. Discussion générale

Les **Expériences 6** et **7** et **8** avaient le même objectif que les **Expériences 4** et **5**, à savoir d'étudier l'avantage de la vision périphérique sur la vision centrale lors de la catégorisation rapide de scènes. Nous avons mentionné à plusieurs reprises que les informations visuelles périphériques seraient principalement encodées en BFS alors que les informations visuelles centrales seraient principalement encodées en HFS. Sur la base de cette propriété visuelle, mais aussi de l'hypothèse d'une prédominance d'un traitement *coarse-to-fine* des fréquences spatiales, nous nous sommes demandés si la catégorisation rapide de scènes suit également un traitement *coarse/peripheral-to-central/fine*. Nous avons reconduit les **Expériences 4** et **5** afin de remédier aux problèmes méthodologiques soulevés par les experts lors de la soumission de ces deux expériences à la revue *Vision Research*.

Pour cela, nous avons mené trois expériences dans lesquelles nous avons utilisé de grandes photographies de scènes à partir desquelles nous avons construit un disque central et quatre anneaux circulaires révélant différentes régions des scènes, de différentes excentricités. Dans une première expérience (**Expérience 6**), les participants devaient catégoriser les scènes entières (Disque entier) et les stimuli individuels (Disque central, Anneau 1, Anneau 2, Anneau 2 et Anneau 4) comme étant des scènes d'intérieur ou d'extérieur. Dans une deuxième expérience (**Expérience 7**), les participants devaient catégoriser des séquences dynamiques composées du disque central et des quatre anneaux. Les stimuli étaient assemblés soit de la vision périphérique à la vision centrale (séquences *peripheral-to-central*), soit de la vision centrale à la vision périphérique (séquences *central-to-peripheral*). L'**Expérience 8** a été spécialement conçue pour évaluer si nos résultats pouvaient être biaisés par un effet de masquage par métacontraste. Dans ces trois expériences, nous avons manipulé le contenu sémantique disponible dans la partie centrale de la scène. Ces expériences ont révélés deux résultats principaux : (1) les performances de catégorisation s'amélioraient à mesure que la scène était dévoilée en vision périphérique (**Expérience 6**) et (2) les performances de catégorisation étaient meilleures lors de la catégorisation des séquences *peripheral-to-central* par rapport aux séquences *central-to-peripheral*, mais uniquement lorsqu'un objet pertinent était présent dans le disque central (**Expérience 7**), alors même que la présence de cet objet améliorait les performances de catégorisation du disque central et des séquences *central-to-peripheral* présentées pendant 500 ms.

Dans l'**Expérience 6**, nous avons également observé que l'absence d'un objet sémantiquement relié à la catégorie de la scène était davantage préjudiciable à la catégorisation des scènes d'intérieur que des scènes d'extérieur. Ce résultat soutient celui de l'**Expérience 4** et suggère que l'absence d'un objet pertinent dans la vision centrale pourrait inciter les participants à considérer une scène d'intérieur comme une étant une scène d'extérieur. Cependant, même si la catégorisation des scènes pouvait être facilitée par la présence d'un objet en son centre, nous avons systématiquement observé un avantage de la vision périphérique. De plus, dans l'**Expérience 6**, les performances obtenues dans la vision périphérique proche (pour les Anneaux 1, 2 et 3 de la Session 2) ne différaient pas des performances obtenues lors de la catégorisation des scènes entières (Disques entiers de la Session 1), suggérant que cette partie de la vision périphérique pourrait être suffisante pour catégoriser des scènes. Par conséquent,

l'**Expérience 6** confirme la plus grande efficacité et utilité de la vision périphérique par rapport à la vision centrale lors de la catégorisation de scènes (Larson & Loschky, 2009), malgré la faible acuité visuelle inhérente à la vision périphérique et la valeur informative de la vision centrale. En contraste avec les résultats obtenus dans l'**Expérience 4** pour laquelle nous n'avons pas contrôlé le facteur de magnification corticale de la meilleure façon, ces résultats suggèrent qu'il existerait une large zone de vision périphérique pour laquelle la catégorisation est optimale (de l'Anneau 1 à l'Anneau 3). Au-delà de cette zone, la reconnaissance deviendrait moins efficace (Anneau 4). A noter que ce résultat nous servira à justifier le choix des stimuli que nous avons utilisés dans les expériences du **Chapitre 4**.

L'**Expérience 7** a été spécialement conçue pour tester si l'analyse rapide des informations grossières disponibles en vision périphérique pourrait permettre une première catégorisation de la scène qui serait ensuite validée par l'analyse plus lente des informations détaillées disponibles en vision centrale. Pour cela, nous avons assemblés les stimuli de l'**Expérience 6** soit du Disque central à l'anneau le plus périphérique (Anneau 4) afin d'imposer une analyse de l'information *central-to-peripheral*, soit de l'anneau le plus périphérique (Anneau 4) au Disque central afin d'imposer une analyse de l'information *peripheral-to-central*. Dans cette expérience, nous avons également testé si la supériorité de la vision périphérique était amplifiée pour des temps de présentation courts. Ainsi, les séquences étaient présentées pendant 165 ms ou 500 ms. Nous avons observé que les participants catégorisaient les séquences *peripheral-to-central* plus rapidement que les séquences *central-to-peripheral*, mais cet effet était en fait plus grand pour le temps de présentation le plus long (500 ms). Ce résultat était en fait dû à un effet totalement inattendu. En l'absence d'un objet central, l'avantage des séquences *peripheral-to-central* n'était plus observé pour la durée de présentation la plus courte (165 ms). En comparaison à la condition dans laquelle il y avait un objet central, les temps de réponses augmentaient considérablement pour catégoriser les séquences *peripheral-to-central*, comme si l'absence d'informations pertinentes dans la vision centrale qui suivait le traitement périphérique était préjudiciable à la catégorisation rapide (165 ms) des scènes. Au premier abord, ce résultat contredit nos hypothèses, mais il pourrait toutefois être interprété dans le cadre d'un traitement *coarse-to-fine* lors de la catégorisation de scènes. Selon cette hypothèse, l'extraction rapide d'informations grossières dans le champ visuel périphérique devrait fournir la forme et la structure globale de la scène pour une catégorisation rapide. Cette catégorisation pourrait être affinée, confirmée ou infirmée par le traitement des informations détaillées disponibles dans le champ visuel central. Cette catégorisation rapide basée sur la périphérie serait alors confirmée par la présence d'un objet pertinent dans la partie centrale de la scène. Ainsi, nous pensons que lorsqu'aucune information pertinente n'est présente en vision centrale, le système visuel aurait besoin de plus d'informations, retardant le temps de catégorisation. En suivant cette logique, l'avantage des séquences *peripheral-to-central* présentées plus longtemps, indépendamment de la présence d'un objet central, suggère que les informations détaillées de la vision centrale ne seraient même pas utilisées. Comme nous l'avons souligné auparavant, il est tout à fait possible que pour des séquences plus longues (500 ms), la catégorisation ne repose que sur le premier anneau présenté pendant 100 ms.

Alternativement, l'avantage des séquences *peripheral-to-central* pourrait être causé par un biais méthodologique. En effet, les anneaux périphériques présentés après le Disque central

dans les séquences *central-to-peripheral* pourraient avoir provoqué un effet de masquage par métacontraste sur le traitement du Disque central. Nous avons donc mené une expérience contrôle (**Expérience 8**) afin de quantifier ce potentiel effet de masquage lors du traitement des séquences *central-to-peripheral* et *peripheral-to-central* et d'évaluer si cet effet est réellement plus important lorsqu'un stimulus périphérique entoure un stimulus central que l'inverse. Dans cette expérience, les participants devaient catégoriser le Disque central ou l'Anneau 4, présentés seuls ou suivis d'une séquence de masques de différentes excentricités induisant un effet de masquage par métacontraste similaire à celui qui pourrait être induit par les séquences *central-to-peripheral* ou *peripheral-to-central*. Les résultats ont montré un effet principal de la présence d'un masque, suggérant qu'une séquence de masques présentée après le Disque central ou l'Anneau 4 peut réduire leur visibilité. Par contre, il n'y avait pas d'interaction significative entre l'excentricité du stimulus à catégoriser (le Disque central et l'Anneau 4) et la présence d'un masque. Ce résultat suggère que l'effet de masquage par métacontraste observé lorsque le stimulus périphérique entourait le stimulus central (séquences *central-to-peripheral*) était similaire à celui observé lorsque le stimulus central était entouré par un stimulus périphérique (séquences *peripheral-to-central*). Par conséquent, même si notre paradigme provoquait intrinsèquement un effet de masquage, ce biais méthodologique serait aussi fort pour les séquences *central-to-peripheral* que pour les séquences *peripheral-to-central*. Il pourrait donc difficilement expliquer à lui seul l'avantage des séquences *peripheral-to-central*.

Une autre explication alternative pour l'avantage des séquences *peripheral-to-central* présentées pendant 500 ms peut être liée aux mouvements oculaires. En effet, pour les séquences les plus longues, il est probable que les participants fixaient séquentiellement chaque stimulus. Malheureusement, nous n'avons pas pu enregistrer les mouvements oculaires dans cette expérience afin de contrôler la position du regard des participants sur le point de fixation central tout au long de la présentation de la séquence. Par conséquent, il ne peut pas être totalement exclu que les participants aient réalisé un mouvement oculaire lorsque la séquence était présentée pendant 500 ms. Nous pensons cependant qu'il est peu probable qu'un tel comportement se soit produit et puisse expliquer l'avantage des séquences *peripheral-to-central*. La première raison est que les participants ne connaissaient pas à l'avance l'excentricité du premier stimulus (soit le Disque central, soit l'Anneau 4). Il aurait été alors inefficace de faire une saccade sans altérer la catégorisation de scènes. Les participants auraient également pu initier une saccade d'un stimulus au suivant dans la séquence, mais cette stratégie aurait été contraignante. En effet, les mouvements oculaires saccadés sont généralement initiés entre 100 et 150 ms (Fisher & Weber, 1993). Si les participants faisaient une saccade lors de la présentation d'un stimulus (présenté pendant 100 ms), cela aurait soit empêché sa perception, soit imposé d'importantes contraintes sur le temps restant pour traiter les informations visuelles dans le stimulus suivant. On peut enfin remarquer que dans le **Chapitre 4** de cette thèse, nous présenterons des expériences où nous avons pu mesurer les mouvements oculaires des participants pendant qu'ils catégorisaient des stimuli présentés brièvement en vision périphérique. Nous n'avons enregistré que 0,76% de saccades dans cette condition, ce qui nous amène à penser que les participants fixent généralement bien le centre même pour catégoriser des stimuli excentrés.

De nouvelles expériences sur l'utilisation de la vision centrale et de la vision périphérique pourraient aider à affiner ces résultats. Par exemple, dans nos expériences, nous n'avons pas filtré le contenu en fréquences spatiales de nos stimuli périphériques car nous considérons que la vision périphérique agit comme un filtre passe-bas naturel. Dans de futures expériences, nous pourrions appliquer un filtre passe-bas qui tienne compte de l'excentricité rétinienne afin d'étudier explicitement le rôle des BFS en vision périphérique. Toutefois, nous pensons qu'il est peu probable qu'une telle manipulation expérimentale puisse avoir un impact sur la catégorisation de nos stimuli comme en témoigne une étude de Loschky et al. (2005). Dans cette étude, les auteurs demandaient à des participants d'explorer visuellement de grandes photographies de scènes. De temps à autre, ils filtraient en BFS les informations présentes en vision périphérique au cours d'une fixation grâce à la technique du *gaze-contingent* (technique permettant de changer l'affichage de l'écran d'ordinateur en fonction de l'endroit où l'observateur regarde). Les participants devaient appuyer sur un bouton dès qu'ils détectaient une dégradation dans la scène visuelle. Les résultats ont montré que les participants ne remarquaient pas de dégradation lorsque leur durée de fixation était inférieure à 100 ms. Ces résultats confortent une fois de plus l'hypothèse que seule l'information en BFS est perçue par le système visuel en vision périphérique lorsque les stimuli visuels sont présentés rapidement sans possibilité de mouvements oculaires, comme c'est le cas dans nos expériences. Le filtrage passe-bas de ces informations est donc superflu. Des expériences supplémentaires peuvent également être nécessaires pour évaluer dans quelle mesure une stratégie de traitement préférentiel *peripheral-to-central* peut dépendre des exigences de la tâche. Dans le contexte du traitement des fréquences spatiales, il est probable qu'une utilisation flexible des informations centrales et périphériques puisse se produire, à l'image de la flexibilité de l'analyse *coarse-to-fine*, en fonction de différents facteurs tels que les contraintes de la tâche visuelle (e.g., des tâches de catégorisation des expressions faciales pourraient nécessiter l'utilisation préférentielle des informations détaillées disponibles au centre ; Schyns & Oliva, 1999) ou le niveau de catégorisation de scènes (e.g., des tâches de catégorisation sous-ordonnées pourraient elles aussi nécessiter une plus grande utilisation des informations détaillées disponibles au centre ; Collin, 2006 ; Collin & McMullen, 2005). Par ailleurs, les résultats de l'**Expérience 6** ont montré que les participants faisaient moins d'erreurs pour catégoriser les scènes d'extérieur que les scènes d'intérieur. De plus, les temps de réponses étaient plus rapides pour les scènes d'extérieur que d'intérieur uniquement pour le stimulus plus périphérique (Anneau 4), comme si la vision périphérique était spécifiquement pertinente pour la reconnaissance des scènes d'extérieur. Dans l'**Expérience 7**, l'avantage des séquences *peripheral-to-central* était plus grand pour les scènes d'extérieur. Les boutons de réponse étant contrebalancés entre les participants, ces résultats suggèrent un biais de réponse en faveur des scènes d'extérieur, peut-être parce qu'une tâche de catégorisation 'extérieur vs. Intérieur' avec de grands stimuli favorise la catégorisation des scènes d'extérieur. Ainsi, de futures expériences utilisant différentes tâches de catégorisation de scènes pourraient être menées.

D'autres études portant sur la façon dont les informations issues de la vision centrale et de la vision périphérique interagissent et s'influencent mutuellement lors de la catégorisation rapide de scènes pourraient aider à affiner nos résultats. C'est cette perspective expérimentale que nous avons spécifiquement choisi d'aborder dans le **Chapitre 4** de cette thèse. Pour ce

faire, nous avons évalué les effets d'interférence sémantique de la vision centrale sur la vision périphérique et réciproquement en présentant simultanément un disque central et un anneau périphérique appartenant soit à la même catégorie sémantique (congruent), soit à des catégories sémantiques différentes (incongruent)

Pour conclure sur ce **Chapitre 3**, les résultats de nos cinq expériences confirment la plus grande utilité et efficacité de la vision périphérique lors de la catégorisation rapide de scènes, malgré la faible acuité visuelle inhérente à la vision périphérique et la valeur informative de la vision centrale. En accord avec un traitement *coarse-to-fine* des fréquences spatiales, l'analyse rapide des informations en BFS disponibles en vision périphérique pourrait permettre une première catégorisation qui serait validée ou non par les informations en HFS disponibles en vision centrale.

Les résultats des Expériences 6, 7 et 8 ont été publiés dans la revue *Vision Research* (Annexe 1) sous le titre : “Rapid scene categorization: From coarse peripheral vision to fine central vision”.

Chapitre 4 - Influence des informations sémantiques et physiques en vision périphérique sur la catégorisation des scènes en vision centrale

Selon les principaux modèles prédictifs de la reconnaissance visuelle décrits dans le **Chapitre 1** (Bar, 2003 ; Bar, 2007 ; Kauffmann, Ramanoël & Peyrin 2014 ; Roux-Sibilon, 2020), des informations rudimentaires en BFS seraient suffisantes et utiles pour déclencher des mécanismes prédictifs qui guideraient ensuite une analyse visuelle plus détaillée en HFS. Étant donné que les BFS sont majoritairement disponibles en vision périphérique, nous avons fait l'hypothèse générale que le système visuel utiliserait très rapidement les informations de faible résolution spatiale (les BFS) disponibles en vision périphérique afin d'accéder à une première représentation de la scène et de générer des prédictions sur son contenu. Ces prédictions guideraient ensuite la reconnaissance de l'information présente en vision centrale. Cette partie expérimentale avait donc pour objectif principal de tester cette hypothèse en étudiant si les informations présentes en vision périphérique influencent effectivement la catégorisation des scènes en vision centrale. Avant de réaliser mes propres expériences pour répondre à cet objectif de recherche, j'ai participé activement au cours de ma thèse à la réalisation de deux autres expériences menées par l'équipe de Carole Peyrin au Laboratoire de Psychologie et NeuroCognition qui ont été publiées dans les revues *Journal of Vision* (Roux-Sibilon, Trouilloud, Kauffmann, Guyader, Mermillod & Peyrin, 2019) et *Journal of Cognitive Neuroscience* (Peyrin, Roux-Sibilon, Trouilloud, Khazaz, Joly, Pichat, Boucart, Krainik & Kauffmann, 2021).

Dans la première expérience de Roux-Sibilon et al. (2019), nous avons étudié l'interférence sémantique d'une scène présentée en vision périphérique sur la catégorisation d'un objet présenté en vision centrale (Figure 4.1a). Le contexte de la scène en vision périphérique pouvait être congruent (e.g., une chaise dans une scène d'intérieur) ou incongruent avec l'objet (e.g., une chaise dans une scène d'extérieur). Dans une première expérience, nous avons manipulé le temps de présentation du contexte périphérique avant l'apparition de l'objet en vision centrale (soit il apparaissait en même temps que l'objet, soit 30 ms avant, soit 150 ms avant). Les résultats ont montré que les participants étaient plus rapides pour catégoriser l'objet lorsque celui-ci se trouvait dans un contexte congruent que dans un contexte incongruent uniquement lorsque le contexte de la scène en vision périphérique était présenté avant l'objet (30 ms et 150 ms). Par ailleurs, l'effet de congruence était d'autant plus fort que les participants avaient du temps pour traiter les informations périphériques avant le début de la présentation de l'objet (150 ms). Ces résultats suggèrent donc que les participants avaient traité

automatiquement les informations en vision périphérique, même si celles-ci n'étaient pas pertinentes pour la tâche. Ils sont également cohérents avec l'hypothèse que l'information en vision périphérique est utilisée pour guider les traitements en vision centrale, plutôt qu'avec l'hypothèse que ces deux informations sont traitées simultanément de façon purement ascendante. Plus il serait possible d'accumuler des informations en vision périphérique, plus le signal prédictif serait fort et influencerait la reconnaissance de l'objet en vision centrale.

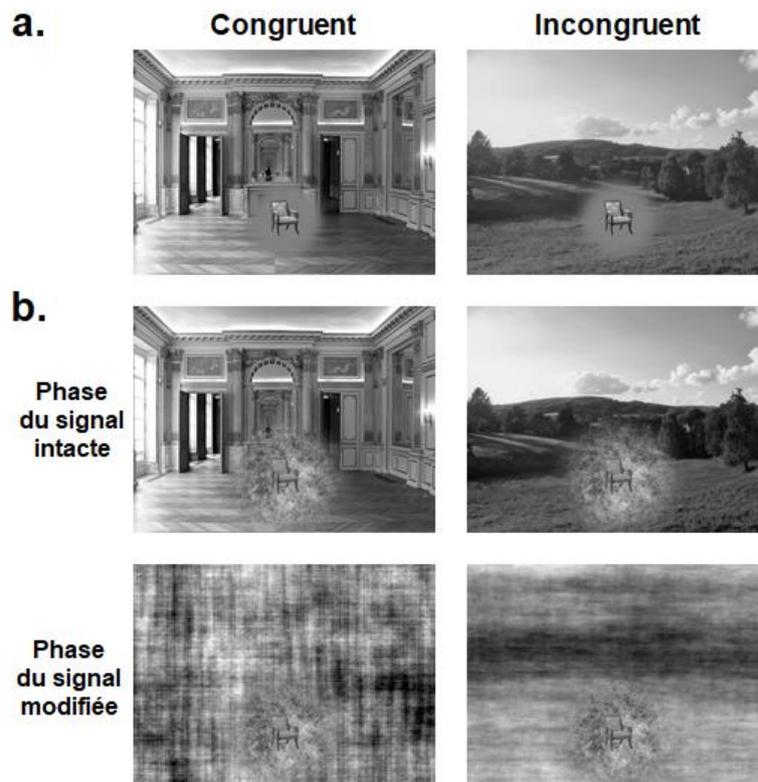


Figure 4.1 : Exemples de stimuli utilisés dans l'étude de Roux-Sibilon et al. (2019) : **(a)** Dans l'expérience 1, Un objet est présenté au centre d'une large scène. Le contexte de la scène périphérique est congruent (e.g., une chaise dans une scène d'intérieur) ou incongruent avec l'objet (e.g., une chaise dans une scène d'extérieur). **(b)** Dans l'expérience 2, la visibilité de l'objet central est réduite. L'information contenue dans le spectre de phase de la scène périphérique est soit intacte, soit modifiée supprimant ainsi les informations concernant la configuration spatiale de la scène.

Dans l'étude de Peyrin et al. (2021), nous avons à nouveau utilisé un paradigme d'interférence sémantique, mais cette fois-ci adapté des images hybrides de Kauffmann et al. (2017) et des expériences de **Chapitre 2** de cette thèse. Nous avons utilisé les deux scènes qui composaient à l'origine une image hybride, mais nous avons présenté une scène en vision centrale et l'autre en vision périphérique (Figure 4.2). Les deux scènes étaient soit sémantiquement congruentes, soit sémantiquement incongruentes et pouvaient partager ou non les mêmes propriétés physiques (spectre d'amplitude et configuration spatiale similaires). Les participants devaient catégoriser la scène présentée en vision centrale comme étant soit une scène artificielle, soit une scène naturelle, tout en ignorant la scène présentée en vision

périphérique. Les résultats ont montré que la catégorisation d'une scène en vision centrale était gênée par la présence en vision périphérique d'une scène sémantiquement incongruente (qui devait être ignorée) et ce, tout particulièrement lorsque les deux scènes étaient physiquement similaires (Figure 4.2). L'interprétation des auteurs était que les informations sémantiques, mais aussi physiques des BFS disponibles en vision périphérique étaient rapidement utilisées pour générer des prédictions sur la catégorie de la scène présentée en vision centrale. Ainsi, lorsque les deux scènes sont sémantiquement incongruentes, les prédictions sémantiques venant de la périphérie sont erronées et gênent la catégorisation de la scène au centre. Mais si l'information physique indique pourtant que les deux scènes se ressemblent et qu'elles devraient être donc de la même catégorie, cela aurait comme effet d'augmenter l'incertitude sur la valeur réelle de la prédiction sémantique et donc le conflit entre ce qui est prédit en périphérie et ce qui doit être catégoriser au centre. Il en résulterait alors un effet d'interférence sémantique encore plus grand lorsque les deux scènes sont physiquement similaires. Au niveau cortical, cet effet d'interférence était associé à une activation plus forte du cortex orbitofrontal, ce qui est cohérent avec les précédentes études menées dans ce contexte de recherche (Kauffmann, Bourgin et al., 2015 ; Mu & Li, 2013). Cet effet était en fait observé lorsque les deux scènes partageaient les mêmes propriétés physiques, soulignant l'importance du contenu sémantique, mais aussi physique, des informations disponibles en vision périphérique pour générer des prédictions.

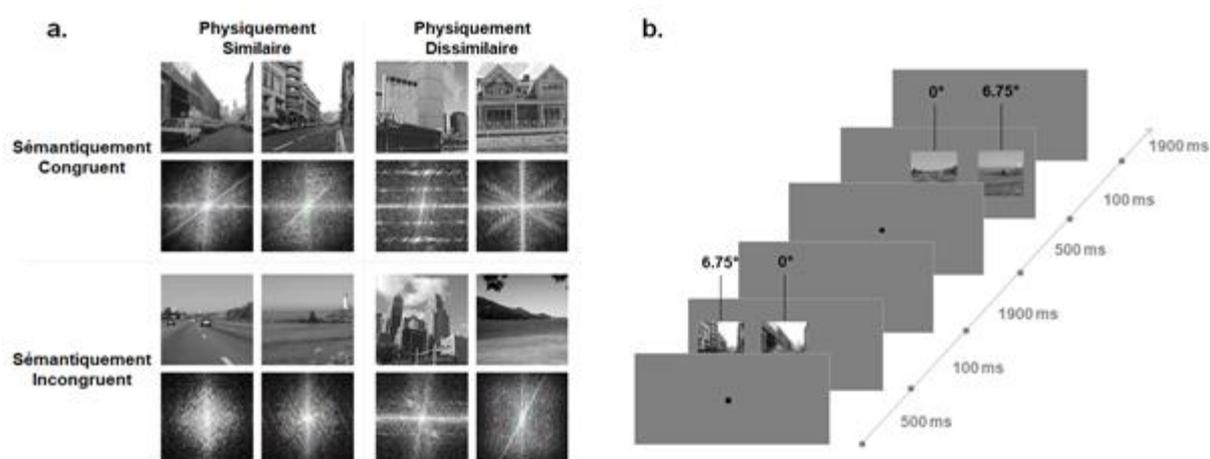


Figure 4.2 : (a) Exemples de paires de scènes avec leur spectre d'amplitude utilisés dans Peyrin et al. (2021). Les deux scènes pouvaient être sémantiquement congruentes ou sémantiquement incongruentes, mais aussi physiquement similaires ou physiquement dissimilaires. (b) Schéma de la procédure expérimentale.

Cependant, cette étude ne nous a pas permis de déterminer la nature des caractéristiques physiques utilisées pour générer les prédictions. En effet, dans le domaine de Fourier, une image est décrite à la fois par son spectre d'amplitude et son spectre de phase. Le spectre d'amplitude résume la distribution du contraste de luminance en fonction des orientations et des fréquences spatiales. Le spectre de phase résume quant à lui des informations sur la position relative des objets dans une scène, c'est-à-dire la configuration spatiale de la scène. Étant donné que les

scènes physiquement similaires que nous avons utilisées étaient appariées à la fois en termes de spectre d'amplitude et de spectre de phase, il ne nous a pas été possible de déterminer si c'était l'information contenue dans le spectre d'amplitude, le spectre de phase ou les deux qui étaient utilisées pour générer des prédictions.

Une réponse possible peut être apportée par la deuxième expérience que nous avons menée dans l'étude de Roux-Sibilon et al. (2019) et dans laquelle nous avons directement abordé la façon dont les caractéristiques physiques présentent dans les régions périphériques d'une scène peuvent influencer la catégorisation d'un objet en vision centrale (Figure 4.1b). Dans cette expérience, nous avons manipulé l'information contenue dans le spectre de phase de la scène périphérique. Cette information était soit intacte, soit modifiée (c'est-à-dire que la distribution de la fréquence spatiale et de l'orientation spatiale était préservée tandis que la phase du signal était modifiée supprimant ainsi les informations concernant la configuration spatiale de la scène). Afin d'augmenter les effets d'interférence, nous avons également rendu la tâche de catégorisation d'objets plus difficile en manipulant le niveau de visibilité des objets centraux (7 niveaux allant du moins visible au plus visible). Les résultats ont effectivement montré que l'effet d'interférence sémantique était le plus important pour les objets peu visibles, mais cet effet n'était observé que lorsque la scène périphérique était intacte, ce qui suggère que les informations contenues dans le spectre de phase - et non celles contenues dans le spectre d'amplitude - étaient automatiquement utilisées en vision périphérique pour générer des prédictions sur la catégorie de l'objet en vision centrale.

Ces résultats ont permis à notre équipe de recherche de proposer une extension au modèle Fréquentiel Prédicatif (modèle **FP**) incluant la vision périphérique : le modèle **FP-Peri** (Roux-Sibilon, 2020). Dans le modèle initial, le système visuel s'appuie sur une première représentation grossière issue des BFS de la scène pour générer rapidement des prédictions sur son contenu. Dans la révision du modèle (Figure 4.3), les informations du champ visuel central et celles du champ visuel périphérique sont traitées de manière principalement indépendante, et en parallèle. Le système de reconnaissance reçoit des BFS provenant de la rétine centrale et de la rétine périphérique, et des HFS provenant uniquement de la rétine centrale. Le système prédictif reçoit des BFS provenant de la rétine centrale mais surtout de la rétine périphérique. Ces prédictions viennent alors guider le traitement ultérieur des informations plus détaillées issues des HFS.

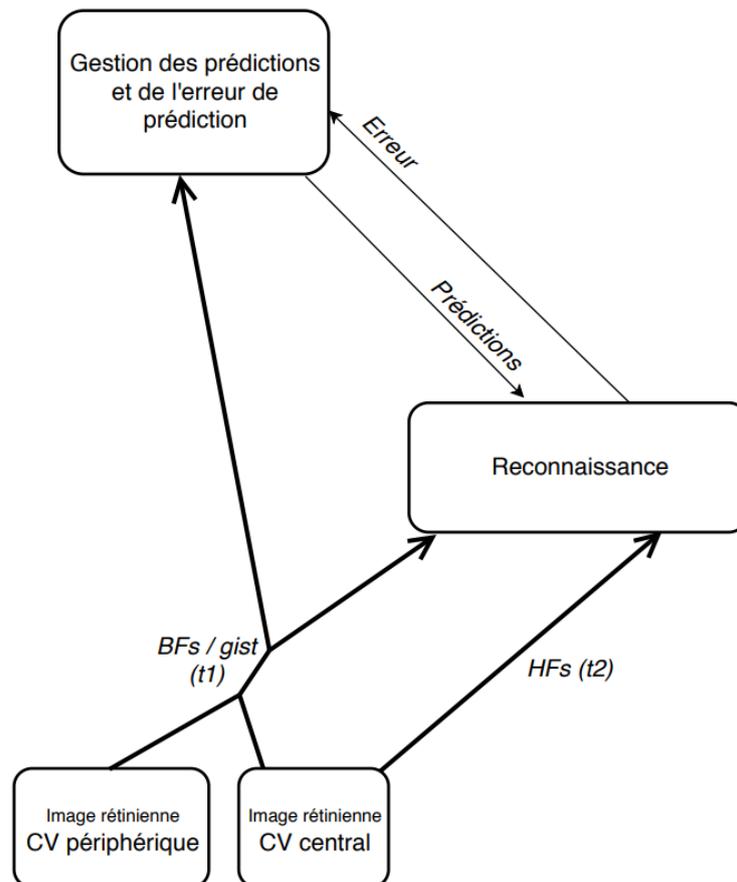


Figure 4.3 : Modèle Fréquentiel Prédicatif incluant la vision périphérique : modèle **FP-Peri**. CV = champ visuel. Figure tirée de la thèse d'Alexia Roux-Sibilon (2020).

Si nos travaux démontrent bien l'influence de la vision périphérique sur la vision centrale, l'inverse est également vrai. Plusieurs études ont montré que la catégorisation du contexte d'une scène peut être influencée par la congruence sémantique d'un objet présent dans la scène (Davenport, 2007 ; Davenport & Potter, 2004 ; Joubert et al., 2008). Lukavský (2019) s'est explicitement intéressé à l'influence réciproque entre la vision centrale et la vision périphérique lors de la catégorisation de scènes (Figure 4.4). En se basant à nouveau sur un paradigme d'interférence sémantique, cet auteur présentait des stimuli composés d'un disque central et d'un anneau périphérique. Le disque central et l'anneau périphérique pouvaient appartenir à la même catégorie sémantique (congruent) ou à deux catégories sémantiques différentes (incongruent). Les participants devaient soit catégoriser le disque central tout en ignorant l'anneau périphérique, soit catégoriser l'anneau périphérique tout en ignorant le disque central. Les stimuli étaient présentés très brièvement pendant 33 ms, suivis par un écran blanc et un masque, le délai entre un stimulus et le masque étant de 200 ms. Les résultats ont montré que les participants faisaient plus d'erreurs lorsque le disque central et l'anneau périphérique étaient incongruents que congruents, quelle que soit la position du stimulus à catégoriser (disque central ou anneau périphérique). Ce résultat indique que l'information à ignorer était traitée automatiquement et qu'elle interférait avec le traitement de l'information à catégoriser. Par

ailleurs, cet effet d'interférence était aussi fort pour la catégorisation du disque central avec un distracteur périphérique que pour la catégorisation de l'anneau périphérique avec un distracteur central. La catégorisation de la scène reposerait donc autant sur des informations visuelles centrales que périphériques.

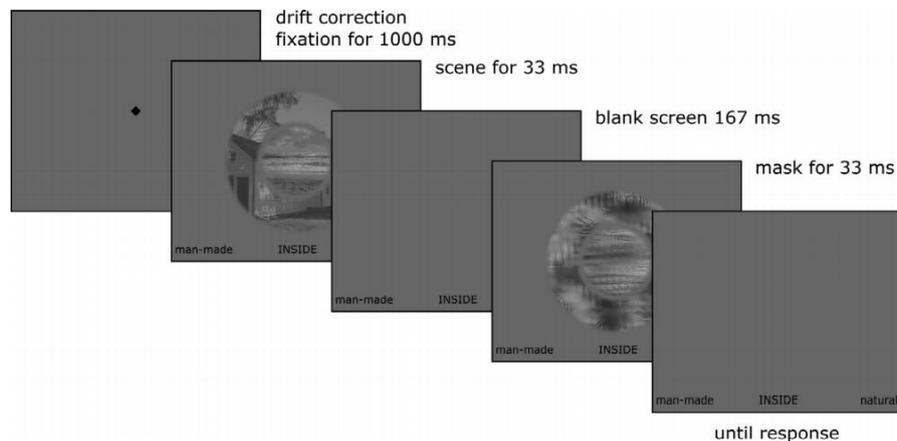


Figure 4.4 : Exemple de stimuli utilisés dans Lukavský (2019). Les deux scènes pouvaient être sémantiquement congruentes ou sémantiquement incongruentes. Figure tirée de Lukavský (2019).

Il est toutefois possible que l'absence de différence entre une interférence sémantique centrale vs. périphérique s'explique dans cette expérience par le temps relativement long entre la présentation des stimuli et le masque (200 ms). Au bout de 200 ms, les informations centrales et périphériques seraient déjà toutes deux traitées, empêchant de déterminer la précedence d'une information sur l'autre. Lukavský a donc mené une deuxième expérience avec la même procédure, mais il a manipulé ce temps (33 ms, 200 ms et 400 ms). Même avec le temps le plus court (33 ms), l'effet d'interférence sémantique était aussi fort pour la catégorisation du disque central avec un distracteur périphérique que pour la catégorisation de l'anneau périphérique avec un distracteur central. Selon l'auteur, ces résultats contredisaient l'hypothèse *zoom-out* proposée par Larson et al. (2014) selon laquelle l'attention est d'abord focalisée en vision centrale, puis se développe rapidement en vision périphérique. En suivant cette logique, on peut aussi avancer que ces résultats contredisent également notre hypothèse d'une précedence de la vision périphérique sur la vision centrale telle que le suggère l'avantage des séquences *central-to-peripheral* dans notre [Expérience 7](#). Lukavský (2019) a plutôt interprété ces résultats dans le cadre du modèle d'accumulation de preuves (Mirolli et al., 2010) ou du modèle *random walk* (Gold & Schadlen, 2007) qui postulent que tous les stimuli perçus fournissent des preuves soutenant plus ou moins plusieurs hypothèses possibles. Les preuves sont alors progressivement accumulées par le système visuel jusqu'à ce que le seuil de décision de l'une des hypothèses soit atteint. Puisque la vision centrale et la vision périphérique s'influencent mutuellement, Lukavský (2019) postule que des preuves provenant de différentes parties du champ visuel sont progressivement rassemblées lors de la reconnaissance visuelle de la scène. Les informations provenant de la vision périphérique et de la vision centrale ne seraient pas traitées de manière

isolée mais en parallèle, de sorte que les informations extraites en vision périphérique influencent l'analyse du contenu central et vice-versa. Ce qui est important à retenir dans les deux expériences de Lukavský (2019), c'est que la vision périphérique influence les traitements réalisés en vision centrale. Il convient toutefois de noter que dans ces expériences, l'excentricité du disque central était de 5,54° d'angle visuel tandis que l'anneau périphérique s'étendait de 5,54° à 11,08° d'excentricité. Si on considère le facteur de magnification corticale, le disque central était donc surreprésenté par rapport à l'anneau périphérique. Par conséquent, nous pensons qu'un traitement prioritaire de l'information périphérique pour déclencher des mécanismes prédictifs est toujours plausible après avoir considéré cette limite.

Dans l'ensemble, toutes les études de ce début de chapitre sont cohérentes avec l'idée que l'extraction d'information de faible résolution en vision périphérique pourrait déclencher des mécanismes prédictifs qui guideraient le traitement très détaillé des informations présentes en vision centrale. L'objectif de ce chapitre expérimental était de tester directement cette hypothèse en utilisant un paradigme d'interférence sémantique inspiré de Lukavský (2019), mais en considérant le facteur de magnification corticale. Pour cela, nous avons réutilisé les stimuli utilisés dans l'**Expérience 6** du **Chapitre 3** où chaque stimulus active la même surface du cortex visuel primaire. Nous avons mené deux expériences (**Expériences 9 et 10**) dans lesquelles nous avons présenté simultanément un disque central et un anneau périphérique qui pouvaient soit appartenir à la même catégorie sémantique (congruent) soit à des catégories sémantiques différentes (incongruent). Dans l'**Expérience 9** de cette thèse, les participants devaient catégoriser le disque central (la cible) tout en ignorant l'anneau périphérique (le distracteur) ou catégoriser l'anneau périphérique (la cible) tout en ignorant le disque central (le distracteur) de la même manière que dans l'expérience de Lukavský (2019). Sur la base de notre hypothèse que l'information visuelle périphérique serait utilisée pour déclencher rapidement des mécanismes prédictifs, nous nous attendions à observer un effet d'interférence sémantique de la vision périphérique sur la vision centrale plus important que celui de la vision centrale sur la vision périphérique. Dans l'**Expérience 10**, nous avons étudié la nature du signal physique en vision périphérique qui influence le traitement en vision centrale (Figure 4.7). Nous avons utilisé soit un anneau périphérique d'origine (condition Intact), soit un anneau périphérique dont le signal contenu dans le spectre de phase était préservé et le signal contenu dans le spectre d'amplitude était modifié (condition Phase), soit un anneau périphérique dont le signal contenu dans le spectre d'amplitude était préservé et le signal contenu dans le spectre de phase était modifié (condition Amplitude). Les participants devaient cette fois-ci toujours catégoriser le disque central tout en ignorant l'anneau périphérique. Nous nous attendions ici aussi à observer un effet d'interférence sémantique pour la condition Intact, les deux autres conditions étant manipulées à titre exploratoire. Ainsi, si l'information contenue dans le spectre de phase (c'est-à-dire la configuration spatiale de la scène est utilisée), nous devrions observer un effet d'interférence dans la condition Phase. Similairement, si l'information contenue dans le spectre d'amplitude (c'est-à-dire à dire la distribution de l'information en fréquences spatiales et en orientations) est utilisée pour générer des prédictions, nous devrions observer un effet d'interférence dans la condition Amplitude.

1. Expérience 9 - Influences mutuelles des informations provenant de la vision périphérique et de la vision centrale lors de la catégorisation de scènes

1.1. Méthode

1.1.1. Participants

Vingt-quatre étudiants de l'Université Grenoble Alpes (19 femmes ; âge moyen : 20,83 \pm 2,47) ont participé à cette expérience. Aucune donnée statistique de référence n'étant disponible pour estimer la taille de l'échantillon, nous nous sommes donc basés sur un échantillon de 20 participants en nous basant sur la taille d'échantillon minimale de 20 participants dans l'**Expérience 1** où nous avons testé l'interférence sémantique distinctement en vision centrale et vision périphérique. Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Ils ont tous donné leur consentement libre et éclairé avant de participer à cette expérience qui a été menée en conformité avec le code d'éthique de l'American Psychological Association (APA) et approuvée par un comité d'éthique (CERGA, Comité d'éthique de la recherche Grenoble Alpes, IRB00010290-2018-10-16-50). A la fin de l'expérience, chaque participant recevait un bon de 0,25 point par quart d'heure d'expérience à valoir lors des examens universitaires.

1.1.2. Stimuli

Nous avons utilisé le Disque central et l'Anneau 3 de l'**Expérience 6** du Chapitre 3. Pour rappel, le disque central révélait la scène à travers un disque de seulement 1,66° d'angle visuel de rayon (i.e. 83 pixels). Pour l'anneau 3, les bords internes et externes étaient respectivement fixés à 7,10° d'angle visuel (355 pixels) et 10,70° d'angle visuel (535 pixels). Nous avons utilisé spécifiquement l'Anneau 3 pour deux raisons. Premièrement, il s'agissait du stimulus qui obtenait les meilleures performances de catégorisation en vision périphérique (aussi bien sur les pourcentages d'erreurs moyens et sur les temps de réponses correctes moyens). Deuxièmement, nous avons montré dans l'**Expérience 6** qu'il existerait une large zone de vision périphérique pour laquelle la catégorisation est optimale (i.e. la catégorisation de l'Anneau 1, Anneau 2 et Anneau 3). Au-delà de cette zone, la reconnaissance visuelle était moins efficace (i.e. la catégorisation de l'Anneau 4). Le Disque central et l'Anneau 3 étaient présentés simultanément (Figure 4.5). Ils pouvaient être soit sémantiquement congruents (e.g., une scène centrale d'intérieur assemblée à la même scène périphérique d'intérieur), soit sémantiquement incongruents (e.g., une scène centrale d'intérieur assemblées à une scène périphérique d'extérieur). Les stimuli étaient présentés sur un fond gris de luminance moyenne de 0,5 pour des valeurs d'intensité des pixels comprises entre 0 et 1.

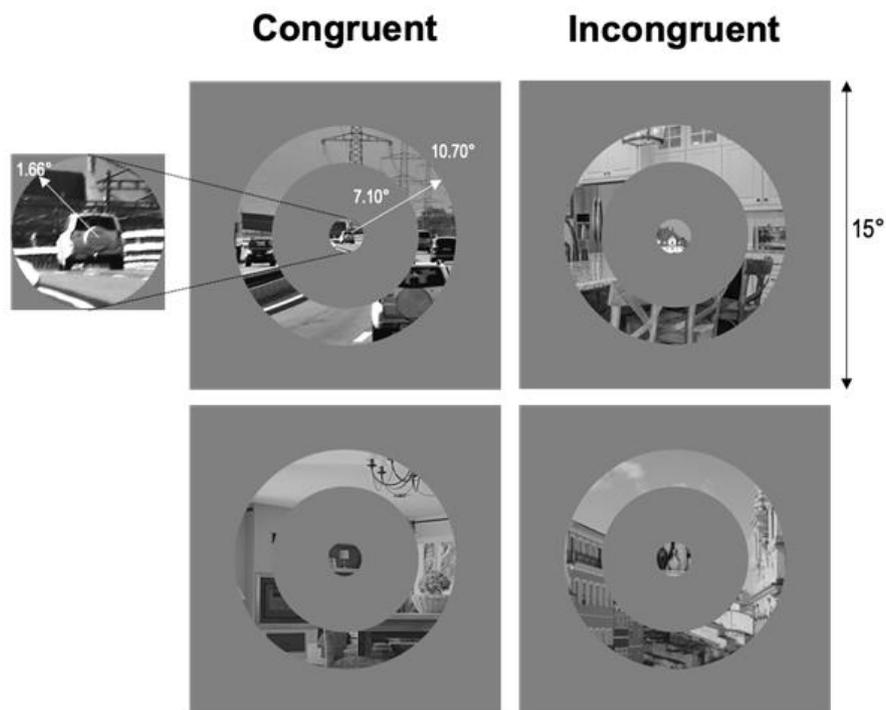


Figure 4.5 : Exemple de stimuli utilisés dans l'[Expérience 9](#). Le Disque central et l'Anneau périphérique étaient présentés simultanément. Ils pouvaient être soit sémantiquement congruents (e.g., une scène centrale d'intérieur et la même scène périphérique d'intérieur), soit sémantiquement incongruents (e.g., une scène centrale d'intérieur et une scène périphérique d'extérieur).

1.1.3. Procédure

L'expérience a été réalisée dans un box expérimental du Laboratoire de Psychologie et NeuroCognition (LPNC) de l'Université de Grenoble Alpes. Les participants étaient installés dans l'obscurité face à un écran d'ordinateur. Les stimuli ont été affichés à l'aide du Psychtoolbox (Brainard, 1997 ; Pelli, 1997) implémenté dans MATLAB R2019 (MathWorks, Natick, MA, USA) sur un écran DELL ULTRASHARP (60 Hz) de 30 pouces (61,4 × 40 cm), avec un affichage de résolution 2560 × 1600 pixels et un taux de rafraîchissement de 60 Hz. Les participants posaient leur menton sur une mentonnière installée sur l'axe sagittal du participant à 70 cm de l'écran afin de respecter la taille angulaire des stimuli.

Tous les participants réalisaient deux sessions expérimentales. Dans une première session expérimentale, les participants devaient catégoriser le Disque central (la cible) et ignorer l'Anneau périphérique (le distracteur). Dans une deuxième session expérimentale, les participants devaient à l'inverse catégoriser l'Anneau périphérique (la cible) et ignorer le Disque central (le distracteur). Les deux sessions expérimentales étaient contrebalancées entre chaque participant.

Pour chaque session, un essai se déroulait comme suit : un point de fixation central (noir sur un fond gris de luminance 0,5) était présenté pendant 500 ms et suivi de la présentation du Disque central et de l'Anneau périphérique pendant 100 ms et d'un écran gris (luminance : 0,5)

de 1900 ms au cours duquel les participants pouvaient répondre. Les participants devaient catégoriser la scène cible comme étant soit une scène d'intérieur, soit une scène d'extérieur par l'intermédiaire des touches 1 et 2 du clavier d'ordinateur placé sur l'axe sagittal médian des participants. La moitié des participants appuyait sur la touche 1 avec l'index de la main droite pour la catégorie Intérieur et sur la touche 2 avec le majeur de la main droite pour la catégorie Extérieur. Les touches de réponse étaient contrebalancées pour l'autre moitié des participants. Les participants avaient pour consigne de répondre le plus correctement et le plus rapidement possible. Chaque session expérimentale comprenait 160 essais (40 stimuli congruents appartenant à la catégorie Intérieur, 40 stimuli congruents appartenant à la catégorie Extérieur et 80 stimuli incongruents). L'expérience totale durait environ 45 minutes, entrecoupés de pauses tous les 80 essais. Pour chaque essai, nous avons enregistré l'exactitude de la réponse et le temps de réponse (en millisecondes). Avant de commencer chaque session expérimentale, les participants réalisaient quelques essais d'entraînement (25 essais pour la Session 1, 25 essais pour la Session 2) en utilisant des stimuli qui n'étaient pas présentés dans l'expérience.

A ce stade du travail de thèse, nous avons pu enregistrer les mouvements oculaires afin de nous assurer que les participants fixaient le centre de l'écran pendant toute la durée d'un essai. Les mouvements oculaires ont été enregistrés à l'aide d'un *eye-tracker* Eyelink 1000 (SR Research) avec un taux d'échantillonnage de 1000 Hz et une résolution spatiale de 0,01° d'angle visuel. Pour chaque participant, seul l'œil droit a été enregistré en utilisant le mode "réflexion pupillaire-cornéenne". Le logiciel Eyelink détectait automatiquement les saccades avec les seuils suivants : vitesse >30°/s, accélération >8000°/s² et déplacement saccadique > 0,15°. Les fixations étaient détectées lorsque la pupille était visible et qu'aucune saccade n'était en cours. Les clignements étaient détectés pendant l'occlusion partielle ou totale de la pupille. Chaque session était précédée d'une procédure de calibration au cours de laquelle les participants devaient orienter leur regard vers neuf points distincts apparaissant séquentiellement dans une grille de 3 × 3 qui occupait la totalité de l'écran. Une correction de la position du regard était effectuée tous les 10 essais et un nouveau calibrage était effectué au milieu de l'expérience.

1.2. Hypothèses

Sur la base des travaux ayant utilisé un paradigme d'interférence sémantique lors de la catégorisation de scènes (Mu et Li, 2013 ; Kauffmann et al., 2015 ; Kauffman et al., 2017 ; Lukavsky, 2019 ; Roux-Sibilon et al., 2019 ; Peyrin et al., 2021), nous nous attendions tout d'abord à observer un effet d'interférence sémantique quelle que soit la position de la cible (centrale ou périphérique), c'est-à-dire à de moins bonnes performances de catégorisation dans le cas où le disque central et l'anneau périphérique sont sémantiquement incongruents que congruents.

Par ailleurs, nous avons fait l'hypothèse dans cette thèse que le système visuel utiliserait très rapidement les BFS issues de la vision périphérique afin de former une première représentation de la scène et de générer des prédictions sur son contenu, prédictions qui guideraient le traitement des HFS en vision centrale. Ainsi, nous nous attendions à observer un effet d'interférence sémantique de la vision périphérique sur la vision centrale plus important

que celui de la vision centrale sur la vision périphérique. Autrement dit, nous nous attendions à une interaction entre la congruence sémantique et la position de la cible.

1.3. Analyse des résultats

Les données ont été analysées à l'aide de l'environnement R (R Core Team, 2019) et du package lme4 (Bates, Mächler, et al., 2015) pour effectuer les analyses sur les pourcentages d'erreurs moyens (%ERm) et les temps de réponse corrects moyens (TRm ; Tableau 4.1). Pour analyser le %ERm, nous avons utilisé un modèle de régression logistique mixte⁵ (la variable dépendante étant une variable binaire). La réponse du participant était codée 1 lorsqu'elle était incorrecte et 0 lorsqu'elle était correcte. Pour analyser les TRm, nous avons utilisé un modèle mixte linéaire pour la Congruence (congruent vs. incongruent) et pour la Position de la cible (centrale vs. périphérique) dans un plan intra-sujet 2×2 . La Congruence, la Position de la cible et leur terme d'interaction ont été introduits dans chaque modèle en tant qu'effets fixes, c'est-à-dire en tant qu'effets des variables d'intérêts. Les interceptes pour les participants, ainsi que les pentes aléatoires pour l'effet de la Congruence, la Position de la cible et leur interaction ont été spécifiés comme effets aléatoires, c'est-à-dire comme les effets des variables auxquelles nous voulons généraliser nos résultats. Pour tester nos hypothèses, nous avons calculé les effets principaux de la Congruence, de la Position de la cible ainsi que leur interaction pour les %ERm et pour les TRm.

Nous avons utilisé la méthode proposée par Bates, Kliegl et al. (2015) pour construire des modèles mixtes parcimonieux empêchant les problèmes de convergence. Le graphique d'*outliers* n'a pas révélé de déviations de l'homoscédasticité ou de la normalité des données. Les valeurs de p ont été obtenues avec un test de Wald pour l'analyse des %ERm et par approximation de Satterthwaite pour l'analyse des TRm avec le package lmerTest (Kuznetsova et al., 2017). Le seuil de significativité a été fixé à 0,5 pour les analyses des %ERm et des TRm. Les tailles d'effet ont été estimées à l'aide du d_z de Cohen (Lakens, 2013).

Dans cette expérience, les mouvements oculaires ont pu être enregistrés pour contrôler que la position du regard sur chaque essai était maintenue dans un rayon de $1,7^\circ$ d'excentricité par rapport au centre de l'écran (i.e. le rayon du disque central). Pour chaque participant et chaque session expérimentale, nous avons supprimé les essais pour lesquels les saccades étaient initiées avec une amplitude supérieure à $1,7^\circ$ autour de la croix de fixation. Cela nous a conduit à supprimer 0,78% des essais. Les données d'un participant qui déplaçait son regard au-delà de $1,7^\circ$ d'excentricité par rapport au centre de l'écran dans une majorité d'essais ont également été

⁵ A ce stade de ma thèse, j'ai pu bénéficier d'une formation en statistique sur les modèles mixtes. Nous avons donc décidé d'analyser toutes les données du Chapitre 4 (Expériences 9 et 10) à l'aide de modèles mixtes. Les modèles mixtes permettent de modéliser les effets fixes (comme dans une analyse classique) et les effets aléatoires. Les effets fixes sont liés à nos variables indépendantes d'intérêts et les effets aléatoires sont liés à la variabilité des participants ou des stimuli. Dans les Expériences 9 et 10, les effets aléatoires incluent l'intercept (l'ordonnée à l'origine) de chaque participant et la variabilité des réponses de chaque participant.

supprimées des analyses. Les analyses statistiques ont donc été réalisées sur 23 participants (18 femmes, âge moyen : $20,77 \pm 2,43$).

Tableau 4.1 : Résultats de l'**Expérience 9**. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (TRm ; en millisecondes) et erreurs-types en fonction de la Congruence sémantique (Congruent ; Incongruent) et de la Position de la cible (Cible centrale/Distracteur périphérique ; Cible périphérique/Distracteur central).

	Cible centrale/Distracteur périphérique		Cible périphérique/Distracteur central	
	%ERm	TRm	%ERm	TRm
Congruent	$4,42 \pm 0,63$	618 ± 19	$6,14 \pm 0,94$	683 ± 31
Incongruent	$7,77 \pm 1,02$	657 ± 21	$13,57 \pm 2,37$	718 ± 29

Les analyses effectuées sur les %ERm ont révélé un effet principal de la Congruence, $\beta = 0,67$, $z = 5,49$, $p < .001$ (Figure 4.6). Les participants faisaient en moyenne plus d'erreurs lorsque les deux scènes étaient sémantiquement incongruentes ($10,10 \pm 6,28\%$) que congruentes ($5,22 \pm 4,64\%$). Ce résultat indique un effet d'interférence sémantique. L'effet principal de la Position de la cible était également significatif, $\beta = 0,35$, $z = 2,41$, $p = .016$. Plus précisément, les participants faisaient moins d'erreurs pour catégoriser le disque central ($6,09 \pm 4,99\%$) que l'anneau périphérique ($9,24 \pm 6,04\%$). Cependant, l'interaction entre la Congruence et la Position de la cible n'était pas significative, $\beta = 0,10$, $z = 0,45$, $p = .656$, suggérant que l'effet d'interférence sémantique était aussi important pour catégoriser une scène cible en vision centrale avec un distracteur en vision périphérique que pour catégoriser une scène cible en vision périphérique avec un distracteur en vision centrale.

Les analyses effectuées sur les TRm ont révélé un effet principal de la Congruence, $\beta = 35,25$, $t(6670,97) = 8,08$, $p < .001$, $dz = 1.685$ (Figure 4.6). Les participants étaient en moyenne plus lents lorsque les deux scènes étaient sémantiquement incongruentes (683 ± 44 ms) que congruentes (648 ± 45 ms), indiquant une nouvelle fois un effet d'interférence sémantique. L'effet principal de la Position de la cible était également significatif, $\beta = 59,78$, $t(21,88) = 3.98$, $p < .001$, $dz = .830$. Les participants étaient plus rapides pour catégoriser le Disque central (637 ± 39 ms) que l'Anneau périphérique (695 ± 49 ms). Comme pour l'analyse des %ERm, l'interaction entre la Congruence et la Position de la cible n'était pas significative, $\beta = -5,25$, $t(22,06) = -0,368$, $p = .717$, $dz = .077$, suggérant à nouveau que l'effet d'interférence sémantique était aussi important pour catégoriser une scène cible en vision centrale avec un distracteur en vision périphérique que pour catégoriser une scène cible en vision périphérique avec un distracteur en vision centrale.

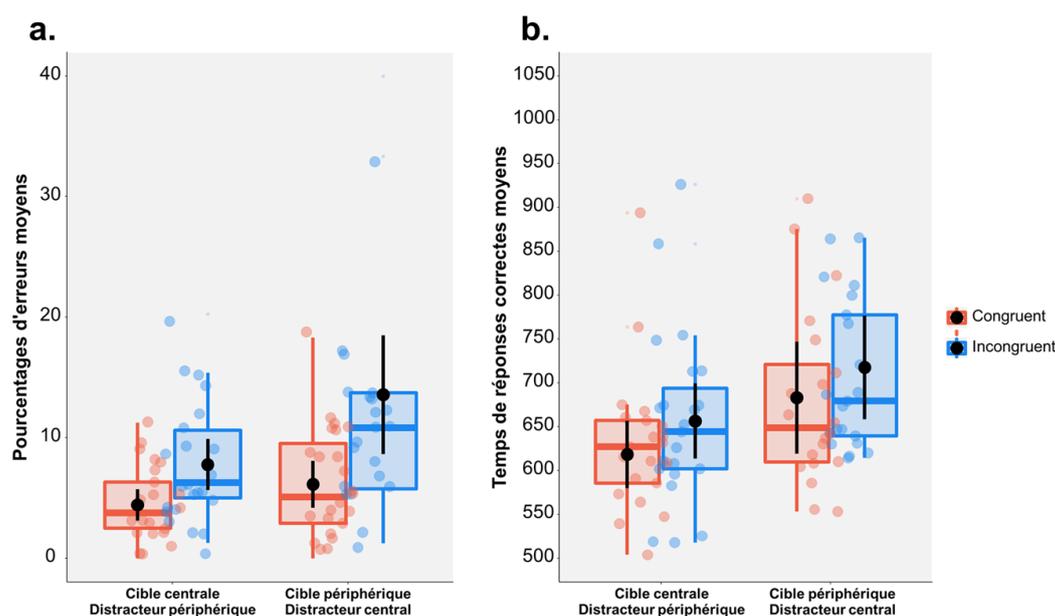


Figure 4.6 : Résultats de l'**Expérience 9** : **(a) Pourcentages d'erreurs moyens** et **(b) Temps de réponses correctes moyens** en fonction de la Congruence sémantique (Congruent ; Incongruent) et de la Position de la cible (Cible centrale/Distracteur périphérique ; Cible périphérique/Distracteur central). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

1.4. Discussion

Conformément à notre hypothèse, les résultats ont montré un effet d'interférence sémantique. En effet, les participants étaient plus rapides et faisaient moins d'erreurs pour catégoriser des scènes congruentes que des scènes incongruentes. Cet effet d'interférence sémantique indique que l'information sémantique conflictuelle contenue dans la scène distractive à ignorer était quand même traitée et perturbait la catégorisation de la scène cible. Bien que les informations provenant de la scène distractive ne soient pas pertinentes pour la tâche, elles sont automatiquement traitées et interfèrent avec la catégorisation de la scène cible. Par ailleurs, sur la base de notre hypothèse selon laquelle les informations en vision périphérique sont utilisées pour générer un signal prédictif qui guiderait le traitement en vision centrale, nous nous attendions à ce que l'effet d'interférence de la scène distractive soit plus important pour la catégorisation du disque central que de l'anneau périphérique. Cependant, l'effet d'interférence de l'anneau périphérique sur la catégorisation du disque central était aussi important que l'effet d'interférence du disque central sur la catégorisation de l'anneau périphérique. Ces résultats sont similaires à ceux de Lukavský (2019) qui n'avait pas observé d'interaction entre la congruence sémantique et la position de la cible. Nous reviendrons sur l'interprétation de ce résultat dans la discussion générale de ce chapitre.

Il faut toutefois noter que notre tâche imposait aux participants de maintenir le regard fixé au centre de l'écran tout au long de l'essai (de la présentation du point de fixation jusqu'à leur réponse), ce qui peut avoir ajouté une difficulté à ignorer l'information centrale. De plus,

même si les participants n'avaient aucune difficulté à déplacer leur attention entre les deux sessions expérimentales (cible centrale vs. cible périphérique), ils avaient de meilleures performances (à la fois sur les pourcentages d'erreurs et sur les temps de réponse) lorsqu'ils catégorisaient le disque central avec un distracteur périphérique que lorsqu'ils catégorisaient l'anneau périphérique avec un distracteur central (indépendamment de la congruence sémantique). Ce résultat suggère qu'il est difficile d'ignorer les informations en vision centrale. Néanmoins, le fait de maintenir le regard fixé au centre de l'écran est inhérent aux conditions d'étude de la vision.

Globalement, les résultats de cette expérience sont cohérents avec les résultats précédents de Lukavský (2019) montrant une influence réciproque entre la vision centrale et la vision périphérique pour catégoriser rapidement des scènes. Les informations provenant de la vision centrale, de haute résolution spatiale, ne peuvent pas être ignorées et influencent donc le traitement des informations en vision périphérique. Néanmoins, l'influence de la vision périphérique sur les traitements réalisés en vision centrale semble être aussi importante que l'influence de la vision centrale, malgré la faible résolution spatiale inhérente à la vision périphérique. Bien que non pertinente pour la tâche, la scène périphérique était également traitée automatiquement et interférait avec la catégorisation de la scène centrale. Pour aller plus loin, nous avons souhaité déterminer la nature des caractéristiques physiques utilisées dans les informations périphériques pour influencer les traitements en vision centrale. Pour cela, nous avons mené une nouvelle expérience (**Expérience 10**) en utilisant le même paradigme expérimental, mais nous avons utilisé soit un anneau périphérique dont le signal était totalement préservé (comme dans l'**Expérience 9**), soit un anneau périphérique dont seul le spectre de phase était préservé (et le spectre d'amplitude brouillé), soit un anneau périphérique dont seul le spectre d'amplitude était préservé (et le spectre de phase brouillé). Ainsi, nous avons pu évaluer indépendamment la contribution du signal de phase et celle du signal d'amplitude dans l'interférence sémantique de la vision périphérique. Dans cette expérience, les participants devaient toujours catégoriser le disque central tout en ignorant l'anneau périphérique.

2. Expérience 10 - Nature de l'interférence de la vision périphérique sur la catégorisation de scènes en vision centrale

2.1. Méthode

2.1.1. Participants

Vingt-deux étudiants de l'Université Grenoble Alpes (21 femmes ; âge moyen : 20,45 ± 1,71) ont participé à cette expérience. Nous avons fixé la taille de l'échantillon en nous basant sur les résultats obtenus dans l'**Expérience 9**. Pour cela, nous nous sommes basés sur l'effet d'interférence sémantique significatif pour la catégorisation de la scène cible centrale pour l'analyse des TRM avec une taille d'effet estimée par le $d_z = 1.296$, avec une puissance désirée

de 0.99 et un seuil alpha de 0.05. Grâce à cette analyse, nous avons obtenu une taille d'échantillon minimale de 22 participants. Tous les participants étaient droitiers avec une vue normale ou corrigée par le port de lunettes ou de lentilles de contact. Comme dans l'[Expérience 9](#), tous les participants ont donné leur consentement libre et éclairé et l'expérience a été approuvée par un comité d'éthique (CERGA).

2.1.2. Stimuli et procédure

Nous avons à nouveau utilisé les stimuli précédemment utilisés dans l'[Expérience 9](#). Dans cette expérience, nous avons manipulé les caractéristiques physiques disponibles dans la scène de l'Anneau périphérique. Elles étaient soit totalement préservées (condition Intact), soit préservées dans le spectre de phase et brouillées dans le spectre d'amplitude (condition Phase), soit préservées dans le spectre d'amplitude et brouillées dans le spectre de phase (condition Amplitude). Les stimuli Phase et Amplitude ont été créés de la façon suivante. Tout d'abord, nous avons calculé la transformée de Fourier de chaque scène pour obtenir leur spectre de phase et leur spectre d'amplitude. Nous avons créé une image de bruit blanc de même taille et nous avons également calculé sa transformée de Fourier afin d'obtenir son spectre de phase et son spectre d'amplitude. Pour créer les stimuli Phase, nous avons appliqué une transformée de Fourier inverse en utilisant le spectre de phase de la scène et le spectre d'amplitude du bruit blanc. A l'inverse, pour créer les stimuli Amplitude, nous avons appliqué une transformée de Fourier inverse en utilisant cette fois-ci le spectre d'amplitude de la scène et le spectre de phase du bruit blanc. Tous les stimuli ont ensuite été ramenés dans le domaine spatial. Une fois dans le domaine spatial, tous les stimuli ont été normalisés de façon à obtenir une luminance moyenne de 0,5 et un contraste RMS (*Root Mean Square*) de 0,2 pour des valeurs d'intensité des pixels comprises entre 0 et 1. Nous avons ensuite créé un Anneau périphérique pour les stimuli Phase et Amplitude en suivant la procédure décrite dans l'[Expérience 6](#).

Comme dans l'[Expérience 9](#), le Disque central et l'Anneau périphérique étaient présentés simultanément (Figure [4.7](#)). Ils pouvaient être soit sémantiquement congruents (e.g., une scène centrale d'intérieur assemblée à la même scène périphérique d'intérieur) ou sémantiquement incongruents (e.g., une scène centrale d'intérieur assemblée à une scène périphérique d'extérieur). Les informations contenues dans l'anneau périphérique étaient donc soit totalement préservées (condition Intact), soit préservées dans le spectre de phase et brouillées dans le spectre d'amplitude (condition Phase), soit préservées dans le spectre d'amplitude et brouillées dans le spectre de phase (condition Amplitude).

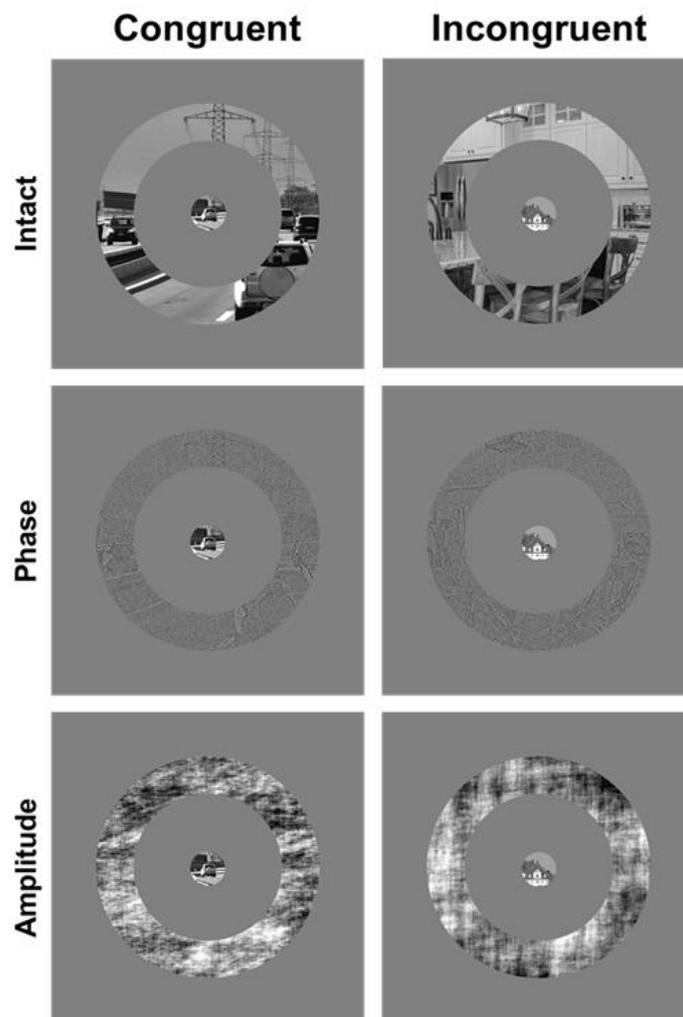


Figure 4.7 : Exemple de stimuli utilisés dans l'**Expérience 10**. Le Disque central et l'Anneau périphérique étaient présentés simultanément. Ils pouvaient être soit sémantiquement congruents (e.g., une scène centrale d'intérieur assemblée à la même scène périphérique d'intérieur), soit sémantiquement incongruents (e.g., une scène centrale d'intérieur assemblée à une scène périphérique d'extérieur). Les informations contenues dans l'anneau périphérique étaient soit totalement préservées (condition Intact), soit préservées dans le spectre de phase et brouillées dans le spectre d'amplitude (condition Phase), soit préservées dans le spectre d'amplitude et brouillées dans le spectre de phase (condition Amplitude).

Les stimuli ont été affichés à l'aide de la Psychtoolbox implémentée dans MATLAB R2019 (MathWorks, Natick, MA, USA) en utilisant le même écran de 30" que dans l'**Expérience 9** avec une distance à 70 cm afin de respecter la taille angulaire des stimuli. Tous les participants devaient catégoriser le Disque central et ignorer l'Anneau périphérique. Un essai commençait par un point de fixation central noir (sur un fond gris de luminance 0,5) présenté pendant 500 ms, suivi de la présentation du Disque central et de l'Anneau périphérique pendant 100 ms, puis d'un écran gris (de luminance 0,5) pendant 1900 ms au cours duquel les participants pouvaient répondre. La tâche des participants était similaire à celle de l'**Expérience 9** (i.e. catégorisation intérieur vs. extérieur). Ils avaient pour consigne de répondre le plus correctement et le plus rapidement possible et de fixer le centre de l'écran pendant toute la durée

de l'expérience. Les mouvements oculaires ont été enregistrés à l'aide de l'*eye-tracker* EyeLink 1000 (SR Research). Ils devaient appuyer sur la touche de réponse correspondante avec le majeur et l'index de la main droite. L'expérience comprenait 240 essais (80 anneaux intacts dont 40 congruents et 40 incongruents ; 80 anneaux de phase dont 40 congruents et 40 incongruents ; 80 anneaux d'amplitude dont 40 congruents et 40 incongruents) et durait environ 30 minutes, avec une pause au milieu. La présentation des essais était totalement aléatoire. Pour chaque essai, nous avons enregistré l'exactitude de la réponse et le temps de réponse (en millisecondes). Avant de commencer l'expérience, les participants réalisaient quelques essais d'entraînement (25 essais) en utilisant des stimuli qui n'étaient pas présentés dans l'expérience.

2.2. Hypothèses

Comme dans l'[Expérience 9](#), nous nous attendions à observer un effet d'interférence pour la condition Intact. De plus, nous avons cherché à déterminer quel type d'information périphérique contribuait à influencer la catégorisation de la vision centrale. Si les informations contenues dans le spectre de phase (i.e. la configuration spatiale de la scène) sont utilisées pour générer des prédictions, nous nous attendions à observer un effet d'interférence dans la condition Phase. Similairement, si les informations contenues dans le spectre d'amplitude (i.e., la distribution de l'information en fréquences spatiales et en orientations) sont utilisées pour générer des prédictions, nous devrions observer un effet d'interférence dans la condition Amplitude. Ainsi, si les deux informations physiques (spectre de phase et spectre d'amplitude) sont utilisées en vision périphérique pour générer des prédictions, nous devrions observer des effets d'interférence dans les deux conditions.

2.3. Analyse des résultats

Les données ont été analysées à l'aide de l'environnement R (R Core Team, 2019) et du paquet lme4 (Bates, Mächler, et al., 2015) pour effectuer les analyses sur les pourcentages d'erreur moyens (%ERm) et les temps de réponse corrects moyens (TRm ; [Tableau 4.2](#)). Pour analyser le %ERm, nous avons utilisé un modèle de régression logistique mixte. La réponse du participant était codée 1 lorsqu'elle était incorrecte et 0 lorsqu'elle était correcte. Pour analyser les TRm, nous avons utilisé un modèle linéaire mixtes pour la Congruence (congruent *vs.* incongruent) et les Informations physiques de la scène périphérique (Intact *vs.* Phase *vs.* Amplitude) dans un plan intra-sujet 2×3 . La Congruence, les Informations physiques et leur terme d'interaction ont été introduits dans chaque modèle comme effets fixes. Les intercepts pour les participants, ainsi que les pentes aléatoires par participants pour l'effet de la Congruence, des Informations physiques et de leur interaction ont été spécifiés comme effets aléatoires. Pour tester nos hypothèses, nous avons calculé si les effets d'interférence étaient significatifs dans chaque condition d'Informations physiques (Intact, Phase, Amplitude) pour les %ERm et les TRm.

Nous avons utilisé la méthode proposée par Bates, Kliegl et al. (2015) pour construire des modèles mixtes parcimonieux empêchant les problèmes de convergence. Le graphique d'*outliers* n'a pas révélé de déviations de l'homoscédasticité ou de la normalité des données. Les valeurs de p ont été obtenues avec un test de Wald pour les analyses des %ERm et par approximation de Satterthwaite pour les analyses des TRm avec le package lmerTest (Kuznetsova et al., 2017). Le seuil de significativité a été fixé à 0,017 après une correction de Bonferroni pour les trois tests non orthogonaux ($0,05/3 = 0,017$) pour les analyses des %ERm et des TRm. Les tailles d'effet ont été estimées à l'aide du d_z de Cohen (Lakens, 2013). Les essais dans lesquels les saccades étaient initiées à partir de plus de $1,7^\circ$ autour de la croix de fixation (i.e. le rayon du disque central) ont été supprimés des analyses. Cela nous a conduit à supprimer seulement 0,08 % des essais.

Tableau 4.2 : Résultats de l'**Expérience 10**. Pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Congruence sémantique (Congruent ; Incongruent) et des Informations physiques (Intact ; Phase ; Amplitude).

	Intact		Phase		Amplitude	
	%ERm	TRm	%ERm	TRm	%ERm	TRm
Congruent	4,43 ± 0,97	633 ± 14	3,64 ± 0,78	623 ± 14	5,34 ± 0,60	621 ± 12
Incongruent	8,18 ± 1,37	665 ± 14	3,86 ± 0,80	643 ± 15	4,20 ± 0,65	635 ± 13

Les analyses effectuées sur les %ERm n'ont pas révélé d'effet d'interférence ni pour les anneaux périphériques Intacts, $\beta = 0,70$, $z = 2,38$, $p = .017$ (Figure 4.8a), ni pour les anneaux périphériques Phase, $\beta = 0,09$, $z = 0,27$, $p = .784$, ni pour les anneaux périphériques Amplitude, $\beta = -0,25$, $z = -1$, $p = .317$.

Les analyses effectuées sur les TRm ont révélé un effet d'interférence significatif pour les anneaux périphériques Intacts, $\beta = 30,73$, $t(4981,13) = 4,14$, $p < .001$, $d_z = .883$ (Figure 4.8b), les participants étaient plus lents lorsque les deux scènes étaient sémantiquement incongruentes (665 ± 35 ms) que congruentes (633 ± 34 ms). Nous avons également observé un effet d'interférence significatif pour les anneaux périphériques Phase, $\beta = 19,05$, $t(4981,03) = 2,60$, $p = .009$, $d_z = .554$, les participants étaient plus lents lorsque les deux scènes étaient sémantiquement incongruentes (643 ± 37 ms) que congruentes (623 ± 33 ms). Cependant, l'effet d'interférence n'était pas significatif pour les anneaux périphériques Amplitude, $\beta = 14,18$, $t(4981,04) = 1,93$, $p = .054$, $d_z = .411$.

Nous avons ensuite effectué des analyses post-hoc de modèles mixtes sur les TRm pour tester si l'effet d'interférence était en fait plus important dans l'une des trois conditions d'informations physiques. Nous avons pour cela calculé comme effet fixe l'interaction entre la Congruence (congruent vs. incongruent) et les Informations physiques (Intact vs. Amplitude ou Phase vs. Amplitude ou Intact vs. Phase) sur les TRm. Cependant, ces analyses post-hoc n'ont montré ni d'effet d'interférence plus fort pour les anneaux périphériques Intacts que pour les anneaux périphériques Phase, $\beta = 11,90$, $t(3311,09) = 1,13$, $p = .257$, $d_z = .241$, ni d'effet d'interférence plus fort pour les anneaux périphériques Intacts que pour les anneaux périphériques Amplitude, $\beta = 16,43$, $t(3293,21) = 1,59$, $p = .113$, $d_z = .339$, et ni d'effet

d'interférence plus fort pour les anneaux périphériques Phase que pour les anneaux périphériques Amplitude, $\beta = 4,90$, $t(3337,08) = 0,47$, $p = .637$, $d_z = .100$.

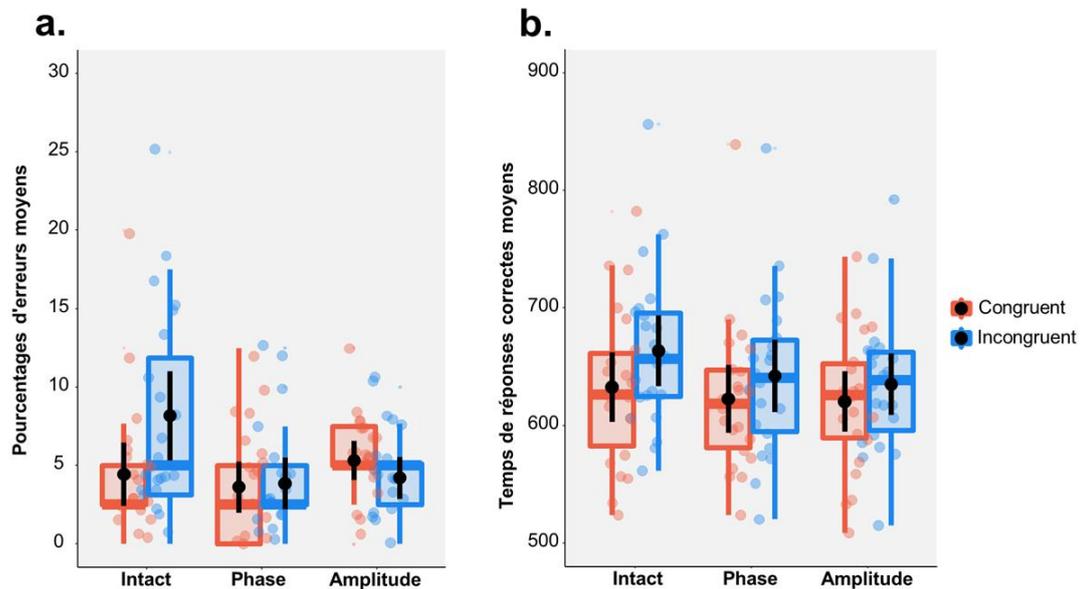


Figure 4.8 : Résultats de l'**Expérience 10** : (a) **Pourcentages d'erreurs moyens** et (b) **Temps de réponses correctes moyens** en fonction de la Congruence sémantique (Congruent ; Incongruent) et des Informations physiques (Intact ; Phase ; Amplitude). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

3. Discussion générale

L'objectif de ces deux expériences était donc de déterminer (1) si les informations en vision périphérique influencent la catégorisation des scènes en vision centrale et (2) quel type de caractéristiques physiques issues de la vision périphérique guide réellement le traitement en vision centrale. Pour ce cela, nous avons utilisé un paradigme d'interférence sémantique dans lequel nous avons présenté simultanément un disque central et un anneau périphérique qui contenaient des scènes pouvant appartenir à la même catégorie sémantique (condition congruent) ou pas (condition incongruent).

Dans l'**Expérience 9**, les participants devaient catégoriser comme une scène d'intérieur ou d'extérieur soit le disque central en ignorant l'anneau périphérique, soit l'anneau périphérique en ignorant le disque central. Les participants devaient fixer le centre de l'écran durant toute l'expérience. Les résultats ont montré un effet d'interférence sémantique indiquant que l'information à ignorer était quand même traitée et interférait avec le traitement de la scène à catégoriser. Comme dans l'étude de Lukavsky (2019), cet effet était présent à la fois pour la catégorisation du disque central et de l'anneau périphérique et ce, malgré le contrôle du facteur de magnification corticale que nous avons effectué afin que les deux stimuli activent la même

quantité de neurones dans le cortex visuel primaire. Ce qui nous importe plus pour la problématique de cette thèse, c'est que les résultats ont montré que malgré sa très faible résolution spatiale, l'information disponible en vision périphérique influence et guide le traitement de l'information que les participants fixent en vision centrale. Ce résultat pourrait être interprété dans le contexte des modèles prédictifs de la reconnaissance visuelle (Bar, 2003 ; Bar, 2007 ; Kauffmann, Ramanoël & Peyrin 2014 ; Roux-Sibilon, 2020). Selon ces modèles, une information de basse résolution spatiale serait suffisante et utile pour déclencher des mécanismes prédictifs corticaux qui guideraient ensuite une analyse visuelle plus détaillée. Roux-Sibilon et al. (2019) ont par exemple montré que la faible résolution spatiale des informations en vision périphérique pouvait être utilisée pour générer des prédictions qui guiderait le traitement des informations en vision centrale. De même, dans notre expérience, lorsque la scène en vision périphérique à ignorer appartient à la même catégorie sémantique que la scène en vision centrale, elle pourrait alors être utilisée pour générer des prédictions correctes sur la catégorie de la scène cible centrale, ce qui faciliterait sa catégorisation. En revanche, lorsque la scène en vision périphérique appartient à une catégorie différente de la scène en vision centrale, son traitement rapide entraînerait alors la génération de prédictions erronées, altérant la catégorisation de la scène en vision centrale et entraînant une baisse des performances de catégorisation.

Dans ce contexte, nous avons fait l'hypothèse que l'interférence sémantique de la vision périphérique serait plus forte que l'interférence sémantique de la vision centrale. Malheureusement, nous n'avons pas observé d'interaction significative entre la congruence et la position de la cible. A posteriori, notre hypothèse semble audacieuse au regard des contraintes méthodologiques de notre expérience et des habitudes visuelles au naturel. En effet, notre tâche imposait aux participants de fixer le centre de l'écran pour éviter de faire des mouvements oculaires tout au long de l'expérience. Cette consigne expérimentale induisait une contrainte attentionnelle à la tâche. Alors que notre attention est distribuée sur l'ensemble du champ visuel dans des conditions normales de vision, les participants étaient ici soumis à des contraintes attentionnelles focalisées. D'ailleurs, Brand et Johnson (2018) ont cherché à savoir si la catégorisation de la scène nécessitait une attention distribuée sur l'ensemble du champ visuel ou si une attention focalisée en vision centrale était suffisante. Ils présentaient une large scène sur laquelle ils superposaient deux rectangles, l'un au centre de l'écran et l'autre entourant la périphérie de la scène. La tâche consistait à catégoriser la scène comme naturelle ou artificielle, puis à indiquer l'orientation du rectangle central (attention focalisée) ou du rectangle périphérique (attention distribuée). Les résultats ont révélé de meilleures performances de catégorisation de la scène dans la condition d'attention distribuée que dans la condition d'attention focalisée, ce qui suggère que l'attention distribuée facilite la catégorisation de la scène à un niveau surordonné (ici, catégories naturelle *vs.* artificielle). En d'autres termes, on déploierait naturellement notre attention sur l'ensemble du champ visuel pour catégoriser une scène à un niveau surordonné comme dans notre expérience. Il est donc possible que notre tâche implique un biais attentionnel en vision centrale qui augmente le poids des interférences de la vision centrale sur la catégorisation en vision périphérique, et inversement, sous-estime le poids des interférences de la vision périphérique sur la catégorisation en vision centrale. De plus, les informations centrales sont intrinsèquement celles qui ont la plus haute résolution spatiale et

des études précédentes ont montré qu'il est plus difficile d'ignorer les HFS que les BFS en vision centrale (Kauffmann et al., 2017). D'ailleurs, indépendamment de l'effet d'interférence, nous avons observé que les participants étaient plus rapides et plus précis pour catégoriser le disque central que l'anneau périphérique, masquant potentiellement un effet d'interférence plus important de la vision périphérique.

Pour aller plus loin, nous avons cherché à déterminer quelles sont les caractéristiques physiques en vision périphérique qui influencent la catégorisation de scènes en vision centrale. Nous avons donc mené une expérience (**Expérience 10**) dans laquelle les participants devaient uniquement catégoriser le disque central tout en ignorant l'anneau périphérique et nous avons manipulé les caractéristiques physiques de l'anneau périphérique. La scène contenue dans l'anneau périphérique était soit intacte (i.e. contenant toutes les informations sémantiques), soit préservée au niveau de son spectre de phase (i.e. contenant seulement des informations sur la structure spatiale de la scène), soit préservée au niveau de son spectre d'amplitude (i.e. contenant les orientations et des fréquences spatiales de la scène). Si les informations du spectre de phase influencent le traitement de la vision centrale, nous nous attendions à un effet d'interférence dans la condition Phase. Similairement, si les informations du spectre d'amplitude influencent le traitement de la vision centrale, nous nous attendions à un effet d'interférence dans la condition Amplitude. Comme dans l'**Expérience 9**, nous avons observé l'effet d'interférence sémantique sur les temps de réponse lorsque la scène en vision périphérique était intacte. Contrairement à l'**Expérience 9**, cet effet n'a pas été observé sur les pourcentages d'erreurs. Il faut cependant noter que les pourcentages d'erreurs extrêmement bas (inférieurs à 8%) suggèrent un effet plafond. La catégorisation du disque central était peut-être trop facile pour détecter l'effet attendu sur les pourcentages d'erreurs. Plus important, nous avons également observé un effet d'interférence de la scène en vision périphérique lorsque l'information de phase était préservée, mais pas lorsque seule l'information d'amplitude était préservée. Les informations incongruentes contenues dans le spectre de phase, contrairement aux informations contenues dans le spectre d'amplitude, retardaient le temps de réponse autant que les informations contenues dans la scène intacte. De la même façon, Roux-Sibilon et al. (2019) ont observé que l'effet d'interférence d'un contexte d'une scène présenté en vision périphérique sur la catégorisation d'un objet présenté en vision centrale disparaissait lorsque le spectre d'amplitude du contexte de la scène était préservé, mais que la configuration spatiale était brouillée (brouillage de phase de la scène comme dans la présente expérience). Dans l'ensemble, ces résultats suggèrent que les caractéristiques sémantiques et spatiales contenues dans le spectre de phase d'une scène en vision périphérique sont primordiales pour guider le traitement de l'information en vision centrale. Il convient de noter que l'absence d'effet d'interférence pour les scènes périphériques conservant leur spectre d'amplitude n'implique pas nécessairement que les informations d'amplitude ne sont pas utilisées lors de la catégorisation des scènes. En effet, de nombreuses études ont montré que, bien que cette information seule puisse ne pas être suffisante (Loschky & Larson, 2008 ; Loschky et al., 2007), les régularités statistiques du spectre d'amplitude des scènes de différentes catégories peuvent être exploitées pour une catégorisation rapide des scènes (Field, 1987 ; Guyader et al., 2004 ; Oliva & Torralba, 2001 ; Torralba & Oliva, 2003). Cependant, nos résultats suggèrent que ces informations

seraient moins pertinentes que les informations de phase pour guider le traitement en vision centrale.

Pour conclure, ces deux expériences montrent une influence mutuelle entre les informations en vision centrale et en vision périphérique pour catégoriser rapidement des scènes. Malgré la faible résolution spatiale du contenu en vision périphérique, l'influence de l'information périphérique sur la catégorisation en vision centrale est aussi importante que l'influence de l'information détaillée de la vision centrale qui ne pourrait être ignorée. Bien que non pertinente pour la tâche, la faible résolution de la vision périphérique serait automatiquement traitée et interférerait avec le traitement de l'information en vision centrale, soutenant ainsi l'hypothèse du rôle prédictif de la vision périphérique pour guider l'analyse du contenu en vision centrale. Plus important encore, ces prédictions seraient basées sur la configuration spatiale de la scène contenue dans le spectre de phase. Malheureusement, la présente étude ne nous a pas permis de trancher entre un modèle d'accumulation de preuves et un modèle proactif concernant le rôle de la vision périphérique. Comme réalisé précédemment par Roux-Sibilon et al. (2019), il serait pertinent dans une expérience future de manipuler les temps pendant lesquels les informations en vision périphérique sont traitées avant l'apparition des informations en vision centrale. Un effet d'interférence croissant avec l'augmentation des SOA soutiendrait un modèle prédictif plutôt qu'un traitement simultané des informations présentes en vision centrale et en vision périphérique (i.e. modèle d'accumulation de preuves).

Les résultats des Expériences 9 et 10 ont été soumis dans la revue *Visual Cognition* (Annexe 2) sous le titre : “Influence of physical features from peripheral vision on scene categorization in central vision”.

Le travail mené dans ce chapitre soutient à nouveau le **modèle FP-Peri**. Le système visuel s'appuierait sur une première représentation grossière issue des BFS de la scène en vision centrale, mais surtout en vision périphérique, pour générer rapidement des prédictions utiles à la reconnaissance d'un stimulus visuel en vision centrale. Dans le **Chapitre 5** de cette thèse, notre objectif était toujours d'étudier le rôle de la vision périphérique dans la reconnaissance de scènes, mais avec une approche originale empruntée à la neuropsychologie. Nous avons étudié les conséquences d'une perte de la vision périphérique sur la reconnaissance de scènes (Figure 4.9) en menant une expérience comportementale auprès de patients atteints d'un glaucome primitif à angle ouvert (GPAO). En observant les modifications comportementales induites par perte de vision périphérique, nous avons pu déduire son rôle dans des conditions de vision normale.

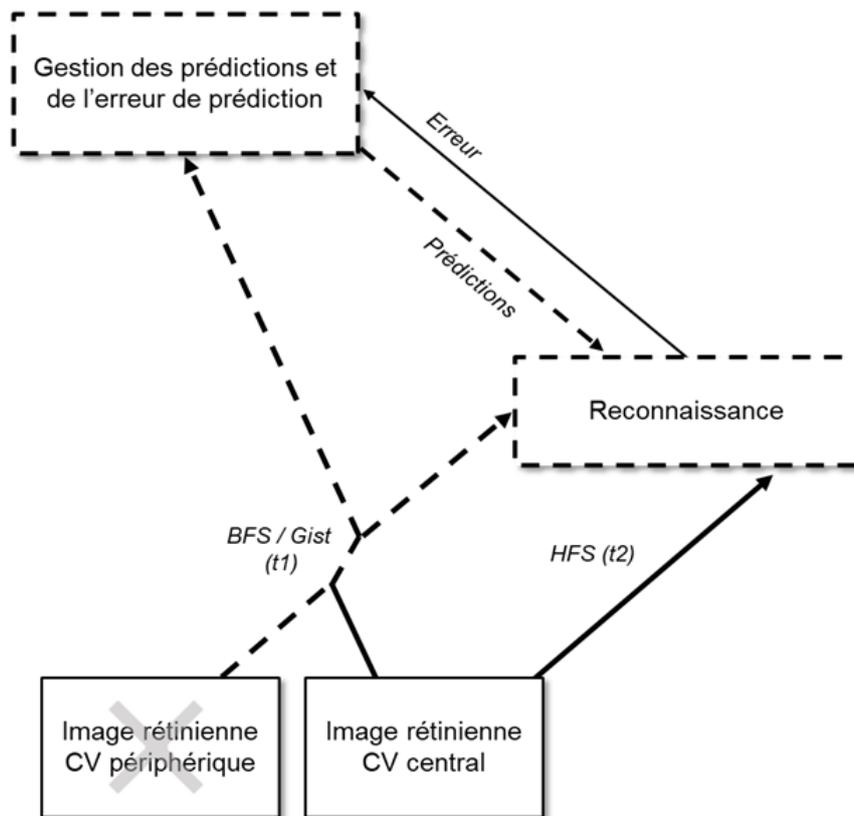


Figure 4.9 : Modèle **FP-Peri** pathologique. Une atteinte de la vision périphérique (représentée par une croix), comme dans le cas du glaucome, entraînerait des perturbations (représentées par les pointillés). CV = champ visuel.

Chapitre 5 - Etude du rôle de la vision périphérique à travers un modèle pathologique : le glaucome primitif à angle ouvert

1. Contexte théorique

1.1. Définition du glaucome primitif à angle ouvert (GPAO)

Le glaucome est une préoccupation majeure de santé publique. C'est une pathologie ophtalmologique chronique fréquente et son incidence augmente avec l'âge. Actuellement, plus de 50 millions de personnes sont concernées par le glaucome. Il s'agit de la deuxième cause de cécité dans le monde, juste après la cataracte, et de la première cause de cécité irréversible dans le monde (Quigley & Broman, 2006). Il existe deux types de glaucome : le glaucome primitif à angle fermé et le glaucome primitif à angle ouvert. Le glaucome primitif à angle ouvert (GPAO) est la forme la plus fréquente et représente 75% de la totalité des glaucomes. Il se caractérise par une destruction progressive des cellules ganglionnaires de la rétine et du nerf optique (Fechtner & Weinreb, 1994). Les principaux facteurs de risque sont l'âge et l'augmentation de la pression intraoculaire consécutive à une perturbation de l'écoulement de l'humeur aqueuse (liquide contenu dans le globe oculaire) à travers un filtre d'évacuation naturel (le trabéculum). Le GPAO entraîne une perte progressive de la vision affectant principalement la vision périphérique. Cependant, la perte de champ visuel est différente d'une vision tubulaire (Crabb, Smith, Glen, Burton & Garway-Heath, 2013).

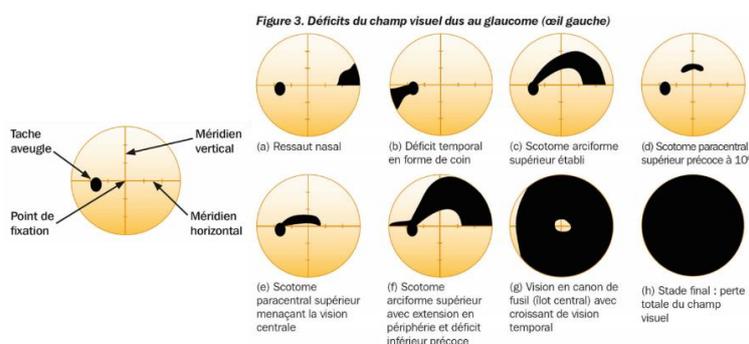


Figure 5.1 : Schématisation de la périmétrie statique automatisée des 30 degrés centraux d'un œil gauche. A gauche : Périmétrie normale. La tâche aveugle correspond à la tête du nerf optique (scotome physiologique). A droite (a, b, c, d, e, f, g, h) : Exemples de scotomes (en noir). Figure tirée de la revue SOC Vol.11 (2014).

Le GPAO est une maladie oculaire qui est insidieuse et peu symptomatique. Le diagnostic est souvent réalisé à un stade avancé, le nerf optique étant déjà très atteint. Dans la mesure où les lésions sont irréversibles, il est important de développer des tests visuels afin de détecter la maladie dès son installation. La perte du champ visuel est communément évaluée à l'aide d'une périmétrie statique automatisée (**Encart 2**). Les pertes du champ visuel dans le glaucome peuvent être de plusieurs types selon la localisation des cellules ganglionnaires lésées. En général, les scotomes (ou tâches aveugles dans le champ visuel ; Figure 5.1) concernent un hémichamp visuel (supérieur ou inférieur), mais ils peuvent également être périphériques ou paracentraux, situés en nasal ou en temporal. Autrement dit, la perte visuelle peut concerner la vision centrale, la vision périphérique ou seulement quelques parties du champ visuel.

L'importance de l'atteinte périmétrique peut être caractérisée grâce à la classification HPA (Hodapp, Parrish & Anderson, 1993). Cette classification se base principalement sur l'indice MD, mais aussi sur le nombre de points testés présentant une sensibilité anormale et sur la présence d'une sensibilité anormale sur les quatre points centraux (Tableau 5.1). L'examen du champ visuel ne permet pas de diagnostiquer la pathologie mais de rechercher et de cartographier des scotomes. Il est important de souligner qu'un scotome identifié par la périmétrie ne signifie pas une perte complète du champ visuel dans cette zone, mais une baisse de sensibilité à la luminance. D'ailleurs, les patients glaucomateux rapportent ne pas percevoir leur perte de vision comme des tâches noires masquant leur champ visuel (Crabb et al., 2013).

Tableau 5.1 : Classification HPA (Hodapp, Parrish & Anderson, 1993) de l'importance de l'atteinte périmétrique du glaucome.

Déficit précoce	Déficit modéré	Déficit sévère
MD ≤ -6dB	-6dB ≥ MD ≥ -12 dB	-12dB ≥ MD ≥ -20dB
Moins de 18 points avec p < 5% et moins de 10 points testés avec p < 1%	Moins de 37 points avec p < 5% et moins de 20 points testés avec p < 1%	Moins de 37 points avec p < 5% et moins de 20 points testés avec p < 1%
Absence de déficit supérieur à 15dB dans les 5 degrés centraux	Absence de déficit absolu (0dB) dans les 5 degrés centraux Pas plus d'un hémichamp ayant des points déficitaires supérieurs à 15 dB dans les 5 degrés centraux	Présence de déficit absolu (0dB) dans les 5 degrés centraux Présence de points dans les 5 degrés centraux avec des déficits supérieurs à 15dB dans les 2 hémichamps

Encart 2 : Evaluation de la perte du champ visuel à l'aide d'une périmétrie de Humphrey

La périmétrie statique automatisée Humphrey SITA-Standard 24-2 (Figure C) permet d'évaluer le seuil de sensibilité rétinienne au sein des 10°, 24° ou 30° centraux, selon le protocole utilisé. Dans cet examen, le patient est assis devant une coupole sur laquelle apparaissent des points lumineux espacés de 6 degrés dans l'espace de stimulation du protocole utilisé. Tout le long de l'examen, le patient doit fixer une cible au centre de la coupole, en même temps qu'apparaissent aléatoirement des points lumineux. Cet examen est généralement réalisé sur chaque œil séparément avec une correction optique de la vision de près. Le patient doit alors appuyer manuellement sur un bouton de réponse à chaque fois qu'il détecte un point lumineux. La luminance des points, qui est très faible lors des premiers essais, augmente jusqu'à ce qu'ils soient détectés par le patient.

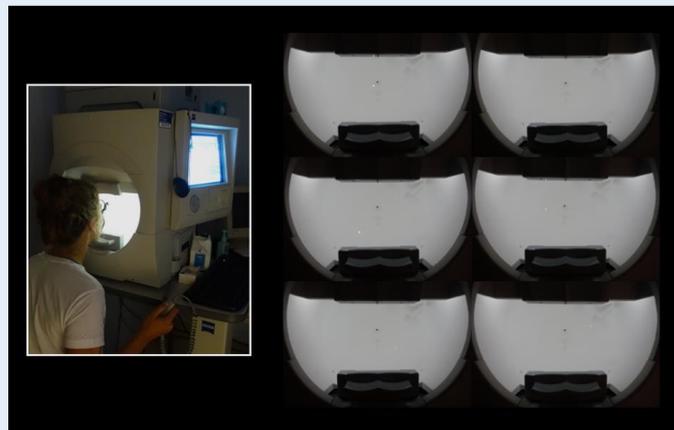


Figure C : Évaluation de la perte du champ visuel à l'aide d'une périmétrie statique automatisée Humphrey SITA-Standard 24-2. Le patient est installé à 33 cm de la coupole d'examen sur laquelle apparaissent des points lumineux espacés de 6 degrés dans l'espace de stimulation. Tout le long de l'examen, le patient doit fixer une cible au centre de la coupole, en même temps qu'apparaissent aléatoirement des points lumineux. Cet examen est généralement réalisé sur chaque œil séparément avec une correction optique de la vision de près. Le patient doit appuyer manuellement sur un bouton de réponse à chaque fois qu'il détecte un point lumineux. La luminance des points, qui est très faible lors des premiers essais, augmente jusqu'à ce qu'ils soient détectés par le patient.

Cette périmétrie permet de recueillir le seuil de détection de luminance pour chaque point rétinien (Figure D). On est ainsi en mesure de calculer un indice de déviation moyenne du champ visuel ou indice MD (Mean Deviation) qui correspond à la différence entre la sensibilité rétinienne normale pour l'âge et la sensibilité rétinienne du sujet testé. Lorsqu'il y a une baisse significative de sensibilité rétinienne, on parle alors d'atteinte périmétrique. Si la baisse n'est pas encore significative, on parle alors d'une atteinte pré-périmétrique.

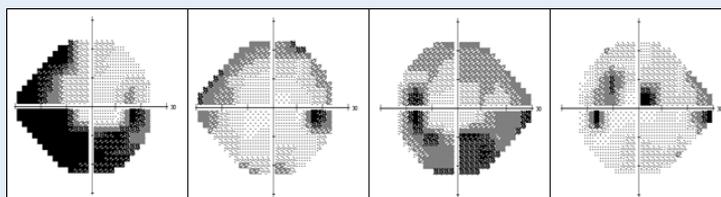


Figure D : Exemples de champs visuels de patients atteints d'un GPAO obtenus sur l'œil testé à l'aide d'une périmétrie statique automatisée (Humphrey SITA-Standard 24-2).

1.2. Troubles visuels cognitifs dans le glaucome

Le bilan ophtalmologique des patients évalue des aspects relativement simples du traitement visuel, comme c'est le cas par exemple avec la détection de points lumineux lors d'une périmétrie de Humphrey (**Encart 2**). Or, en raison d'un traitement rétinien appauvri, le glaucome affecte des aspects bien plus complexes du traitement visuel. En effet, les patients souffrant de glaucome signalent souvent une gêne dans leur vie quotidienne, même au début de la maladie. Les difficultés liées à l'éclairage, telles que l'éblouissement et la difficulté à s'adapter à différents niveaux de lumière sont régulièrement classées parmi les plaintes les plus fréquentes (Dubey et al., 2019 ; Lam et al., 2020 ; Nelson, Aspinall & O'Brien, 1999). Les patients glaucomeux bilatéraux déclarent également avoir des difficultés à accomplir un large éventail de tâches, comme des difficultés pour marcher (e.g., ils heurtent des objets ou peinent à monter des escaliers ; Nelson et al., 1999 ; Odden et al., 2020 ; Viswanathan et al., 1999), des difficultés pour lire (Lee et al., 1998) et des difficultés pour la conduite automobile (Bechetoille et al., 2008 ; Gutierrez et al., 1997), dans la recherche visuelle d'objets (Smith, Crabb & Garway-Heath, 2011 ; Smith et al., 2012) ou de visages (Glen & Crabb, 2015) et dans la réalisation de jeux de construction (Dive et al., 2016). Une étude récente menée auprès de ces patients met en évidence des difficultés dans la préhension d'objets qui seraient dues à des difficultés dans l'estimation visuelle de l'accessibilité (Lenoble et al., 2019). Plus précisément, les patients sous-estiment la distance de l'objet à saisir par rapport à des participants contrôles. Dans une autre étude dans laquelle les auteurs ont mesuré cette fois-ci les contraintes temporelles de la saisie d'objet, les résultats ont montré que les patients mettaient plus de temps à saisir un objet par rapport à des participants contrôles mais uniquement dans le cas où ils n'avaient pas eu le temps d'explorer leur environnement visuel et d'identifier la cible (Corveleyn et al., 2020). Ainsi, ces deux études démontrent que les patients atteints d'un glaucome présenteraient des troubles moteurs dans la préhension de l'objet en raison de troubles visuels liés aux contraintes de la tâche.

Plus récemment, quelques auteurs se sont intéressés aux capacités de catégorisation des patients glaucomeux en vision centrale (Lenoble, Lek & McKendrick, 2016 ; Roux-Sibilon et al., 2018 ; Schafer et al., 2018 ; Stievenard et al., 2020). En effet, jusqu'ici les capacités visuelles de cette population en vision centrale ont reçu peu d'attention, le glaucome étant généralement considéré comme impactant principalement les fonctions impliquant la vision périphérique. Par exemple, Lenoble et al. (2016) présentaient des petites photographies d'objets dans la vision centrale de patients glaucomeux. Pour tous les patients, cette partie du champ visuel était classifiée comme normale par la périmétrie statique automatisée. Les patients, ainsi que des participants contrôles du même âge avec une vision normale, devaient décider s'ils identifiaient un animal parmi des outils, ou un meuble parmi des légumes. Dans cette étude, les auteurs manipulaient le contraste de luminance des objets. Le contraste était soit maximum (100%), soit diminué de moitié (50%). Lorsque le contraste était maximum, les patients montraient des performances préservées par rapport aux participants contrôles. Par contre, lorsque le contraste était à 50%, les patients faisaient plus d'erreurs de catégorisation que les participants contrôles, et lorsque la catégorisation était correcte, elle était plus lente. Cette étude montre que les caractéristiques physiques des stimuli visuels - ici le faible niveau de contraste

- influencent fortement les capacités de reconnaissance visuelles des patients dans leur vision centrale (qui était leur vision résiduelle pour ces patients). Dans la continuité de cette étude, Roux-Sibilon et al. (2018) ont étudié les capacités de patients atteints d'un GPAO bilatéral pour catégoriser des petites images de scènes ou de visages en vision centrale. Dans cette étude, les patients étaient répartis suivant deux groupes sur la base de leur examen du champ visuel (périmétrie de Humphrey 24-2) : un groupe présentant à la fois une perte visuelle périphérique et centrale et un groupe présentant une perte visuelle seulement périphérique, la vision centrale étant épargnée (Figure 5.2a). Tous les patients, ainsi que des participants contrôles du même âge avec une vision normale, devaient réaliser deux tâches où des images à très faible contraste (2,5% et 10% de contraste ; Figure 5.2b) étaient présentées dans le champ visuel central (6° d'angle visuel) : une tâche de détection et une tâche de catégorisation. La tâche de catégorisation permettait d'évaluer les capacités de reconnaissance visuelle de haut niveau, alors que la tâche de détection permettait d'évaluer, sur les mêmes stimuli, les capacités visuelles de bas niveau, comme le fait la périmétrie statique automatisée. La différence des performances entre la détection et la catégorisation révélait ainsi le coût du traitement visuel de haut niveau. Par rapport aux participants contrôles, les résultats des patients dont la vision centrale était atteinte ont montré un déficit à la fois pour détecter et catégoriser toutes les images à faible contraste. Ce résultat est cohérent avec la sensibilité anormale de la rétine centrale telle qu'évaluée par la périmétrie. Cependant, le déficit était plus important pour la catégorisation que pour la détection. Les patients dont la vision centrale n'était pas atteinte montraient des performances similaires aux participants contrôles pour les tâches de détection et de catégorisation des visages. Par contre, alors que leurs performances de détection des scènes étaient préservées, ces patients présentaient un déficit pour catégoriser les scènes.

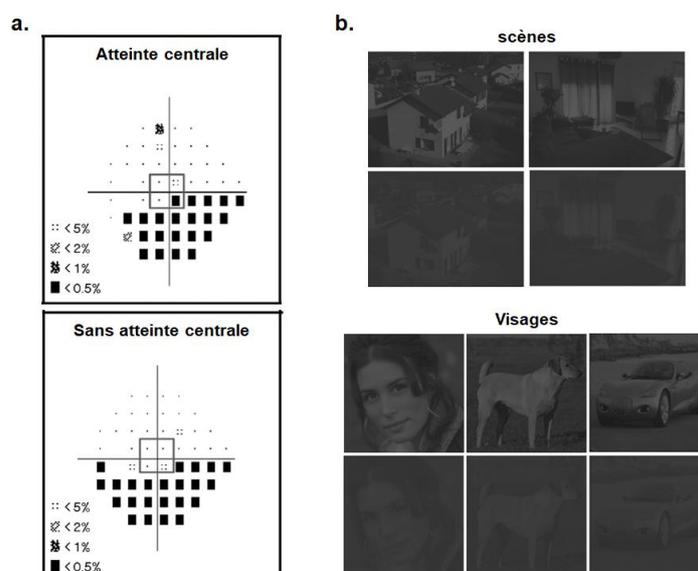


Figure 5.2 : (a) Périmétrie automatisée de de Humphrey SITA-Standard 24-2 de deux patients (un avec une atteinte centrale et un sans atteinte centrale) de l'étude de Roux-Sibilon et al. (2018). Un déficit dans le champ visuel central était défini par au moins un des quatre points centraux présentant une probabilité de 2% ou moins d'être normal. (b) Exemples de stimuli (des scènes et des visages) faiblement contrastés utilisés dans l'étude de Roux-Sibilon et al. (2018).

Ces deux études confortent les plaintes des patients glaucomateux liées aux conditions de luminosité. Les difficultés liées à l'éclairage telles que l'éblouissement et la difficulté à s'adapter à différents niveaux d'éclairage sont régulièrement classées parmi les plaintes les plus fréquentes (Crabb et al., 2013). Les patients expriment significativement plus de difficultés à voir dans l'obscurité (faible luminosité) et à la lumière du jour (forte luminosité) que les personnes non atteintes (Nelson, Aspinall & O'Brien, 1999). Dans ces deux situations, le contraste de luminance est faible. Plus surprenant, ces deux études suggèrent qu'une simple perte de la vision périphérique peut entraîner un déficit visuel de haut niveau - que nous qualifions de trouble visuel cognitif - dans la vision centrale de patients glaucomateux, alors que leur sensibilité rétinienne en vision centrale est considérée comme préservée (diagnostic normal à la périmétrie statique automatisée).

Nous ne connaissons toujours pas l'origine de ces troubles visuels cognitifs. C'est pourquoi il semble important de mieux comprendre les mécanismes de la perception de scènes en vision centrale chez les patients avec un glaucome. Nous avons pour cela poursuivi deux objectifs de recherche développés ci-dessous. D'un point de vue plus fondamental, notre objectif était d'étudier les conséquences d'une perte de la vision périphérique sur les traitements réalisés en vision centrale afin d'en déduire les mécanismes de la reconnaissance de scènes en vision centrale imputables à la vision périphérique.

1.3. Objectifs et hypothèses

- ***Etude des fonctions magno- et parvocellulaires à travers la manipulation des fréquences spatiales***

Notre premier objectif visait à étudier les fonctions magno- et parvocellulaires chez les patients GPAO. En effet, des analyses histologiques menées auprès de patients glaucomateux qui montrent que les neurones avec les corps cellulaires et les diamètres axonaux plus grands sont plus à risque d'être affectés au début du processus pathologique (Quigley, Dunkelberger & Green, 1988 ; pour une revue à ce sujet voir Wang et al., 2020). Ces caractéristiques étant propres aux cellules ganglionnaires magnocellulaires, ce résultat suggère donc une perte sélective de la fonction magnocellulaire au début de la maladie. L'hypothèse d'une atteinte sélective de la voie magnocellulaire est cohérente avec les déficits de patients pour catégoriser en vision centrale des stimuli faiblement contrastés (Lenoble et al., 2016 ; Roux-Sibilon et al., 2018). Toutefois, une étude psychophysique menée par McKendrick, Badcock et Morgan (2004) auprès de patients GPAO amène à des conclusions plus controversées, cette étude mettant en évidence un déficit des deux voies. Les auteurs ont utilisé un paradigme des *Steady-Pulsed pedestal* adapté de l'expérience de Pokorny et Smith (1997) qui permet de tester sélectivement l'une des deux voies (Figure 5.3).

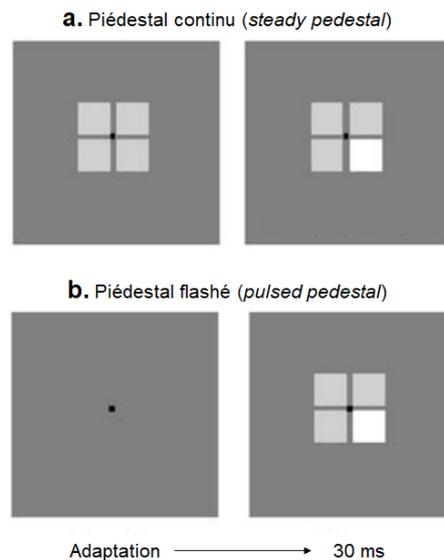


Figure 5.3 : Illustration du paradigme des *Steady-Pulsed pedestal* utilisé par McKendrick, Badcock et Morgan (2004). **(a)** Condition “Piédestal continu” et **(b)** condition “Piédestal flashé”. Figure adaptée et traduite de McKendrick, Badcock et Morgan (2004).

Dans une première condition de ce paradigme, condition qu’on appelle “piédestal continu” (ou *steady pedestal*), on présente aux participants quatre carrés (ou piédestal) sur un fond gris fortement contrasté, puis un seul de ces carrés change de luminosité. Le participant doit indiquer le carré dont la luminosité est différente. Il doit donc traiter des faibles différences de luminosité entre le carré cible et le piédestal. Cette condition est censée stimuler la voie magnocellulaire. Dans une seconde condition qu’on appelle “piédestal flashé” (ou *pulsed pedestal*), on présente un fond gris, puis quatre carrés (piédestal) sur ce fond gris dont un carré avec une luminosité différente. A nouveau, le participant doit indiquer le carré dont la luminosité est différente. Il peut dans cette condition comparer le carré cible au fond de l’écran qui présente un fort contraste. Cette condition est censée stimuler la voie parvocellulaire. Pour chaque condition, on mesure le seuil de luminosité du carré cible, c’est-à-dire la plus petite différence de luminosité entre le carré cible et le piédestal pour laquelle le participant arrive à détecter le carré cible. McKendrick et al. (2004) ont montré que pour les deux conditions expérimentales, les seuils des patients glaucomateux sont plus élevés que ceux des participants du même âge avec une vision normale. Cette dernière étude suggère qu’il y aurait une perte fonctionnelle de l’information transmise par les voies parvocellulaire et magnocellulaire dans le glaucome, et l’ampleur de ce déficit serait similaire dans les deux voies visuelles. Cependant, cette étude ne tient pas compte du stade de progression de la maladie, ni donc de l’hypothèse que seule la voie magnocellulaire serait affectée au stade précoce. De notre point de vue, une perte sélective de la fonction magnocellulaire en début de maladie est cohérente avec l’installation du scotome majoritairement en vision périphérique. En effet, puisque la rétine périphérique est composée en majorité de cellules magnocellulaires, une perte de stimulation en vision périphérique chez les patients devrait alors impacter la fonction magnocellulaire.

Dans cette partie expérimentale, nous proposons donc de tester les fonctions magno- et parvocellulaires dans la vision centrale de patients glaucomateux en manipulant le contenu en fréquences spatiales de scènes visuelles. Nous avons utilisé les scènes filtrées en BFS et HFS de l'**Expérience 1** de cette thèse (Figure 5.5). Les scènes étaient présentées en vision centrale et les patients devaient décider pour chaque scène si elle appartenait à la catégorie naturelle ou artificielle. Nous avons comparé les performances des patients à celles de participants avec une vision normale et appariés en âge. Nous avons évalué le traitement des fréquences spatiales en tenant compte de la progression de la maladie. Nous avons pour cela constitués deux groupes de patients : un groupe de patients au stade précoce de la maladie (GPAO Précoce) et un groupe de patients au stade avancé (GPAO Avancé). Nous nous attendions à observer un déficit du traitement des fréquences spatiales en vision centrale pour les patients par rapport aux participants contrôles. Sur la base des travaux postulant une atteinte sélective de la voie magnocellulaire au début du glaucome (Quigley, Dunkelberger & Green, 1988), nous nous attendions également à ce qu'au stade précoce de la maladie, le déficit du traitement des fréquences spatiales soit plus marqué pour les BFS que les HFS. L'avancée dans la maladie impacterait ensuite à la fois le traitement des BFS et HFS.

- ***Etude des mécanismes prédictifs***

Le second objectif visait à étudier les mécanismes prédictifs chez les patients GPAO. Dans le **Chapitre 4** de cette thèse, nous avons vu que l'information disponible en vision périphérique, même si de faible résolution spatiale, serait explicitement utilisée lors de la catégorisation de stimuli présentés en vision centrale. Ces interactions entre la vision centrale et périphérique peuvent être interprétées au regard de modèles du codage prédictif de l'information visuelle (Bar, 2003 ; Bar, 2007 ; Kauffmann, Ramanoël & Peyrin 2014) et en particulier le **modèle FP-Peri** (Roux-Sibilon, 2020) présenté au **Chapitre 4**. Des informations rudimentaires en BFS, extraites rapidement dans l'ensemble du champ visuel (en vision centrale et en vision périphérique) seraient rapidement utilisées par le cerveau pour déclencher des mécanismes prédictifs. Le système visuel s'appuierait majoritairement sur les prédictions issues de la vision périphérique pour guider l'analyse plus lente des informations présentes sur la fixation de l'œil (donc en vision centrale). Ce mécanisme serait d'autant plus impliqué que l'information présente en vision centrale serait de mauvaise qualité perceptive. Par exemple, dans l'étude de Roux-Sibilon et al. (2019 ; Expérience 2), nous avons manipulé le niveau de visibilité des objets présentés en vision centrale (7 niveaux allant du moins visible au plus visible ; Figure 5.4) et nous avons montré que le seuil de visibilité d'un objet était plus bas lorsque le contexte était congruent qu'incongruent. Si on extrapole ce résultat au quotidien, on reconnaîtrait plus facilement un objet dans l'obscurité si on dispose d'informations en vision périphérique.

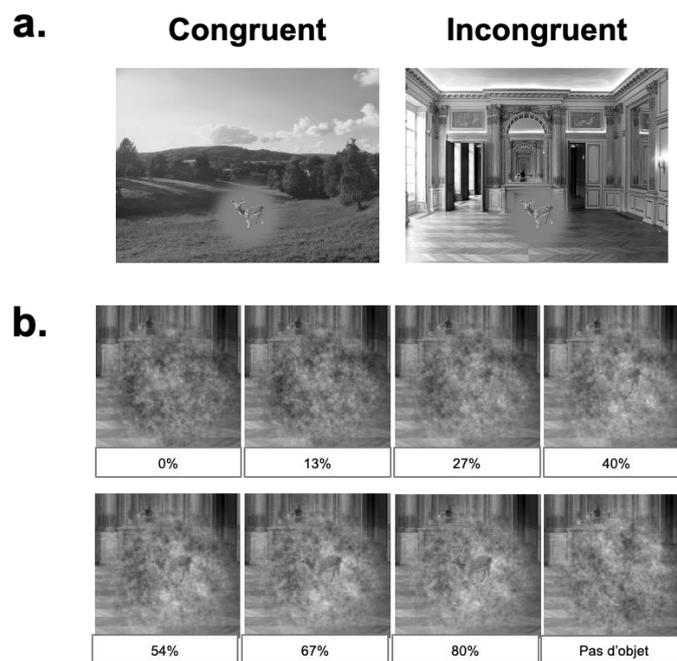


Figure 5.4 : Exemples de stimuli utilisés dans l'étude de Roux-Sibilon et al. (2019). **(a)** Un objet est présenté au centre d'une large scène. Le contexte de la scène périphérique est congruent (e.g., un animal dans une scène d'extérieur) ou incongruent avec l'objet (e.g., un animal dans une scène d'intérieur). **(b)** Dans l'expérience 2, la visibilité de l'objet central est réduite selon huit niveaux de visibilité différents.

Dans le cas du glaucome, notre hypothèse est qu'à cause de la perte progressive de stimulation de la rétine périphérique, ces patients ne bénéficieraient pas pleinement des processus prédictifs inhérents à la vision périphérique. Le second objectif de ce chapitre était donc de tester les mécanismes prédictifs engagés lors de la catégorisation de scènes en vision centrale, et ce à l'aide d'un paradigme d'interférence des BFS sur la catégorisation de scènes en HFS. Nous avons ici utilisé les images hybrides physiquement dissimilaires de l'**Expérience 1** (Figure 5.5). Nous avons préféré utiliser uniquement les images hybrides physiquement dissimilaires car nous pensons que la forte ambiguïté présente entre les deux scènes qui se superposent spatialement dans une image hybride allait particulièrement gêner la reconnaissance des patients atteints de GPAO. Les scènes étaient présentées en vision centrale et les patients devaient ignorer la scène en BFS et porter leur attention sur la scène en HFS. Ils devaient décider si la scène en HFS appartenait à la catégorie naturelle ou artificielle. Nous avons comparé les performances des patients à celles de participants avec une vision normale et appariés en âge. Ici aussi, nous avons pris en considération la progression de la maladie.

Si le système visuel génère des prédictions sur la base des informations disponibles en vision périphérique, la perte progressive de stimulation en vision périphérique chez les patients devrait alors impacter l'effet d'interférence sémantique des BFS sur la catégorisation des scènes en HFS pour les patients par rapport aux participants contrôles, et ce d'autant plus pour le groupe de patients avancé que précoce. Nous nous attendions donc à une interaction entre la

congruence sémantique entre les deux scènes d'une image hybride (congruent et incongruent) et le groupe (Contrôle, GPAO Précoce et GPAO Avancé).

Ces deux objectifs de recherche ont été testés au sein d'une même expérience comprenant deux sessions. La **Session 1** évaluait les fonctions magno- et parvocellulaires à travers la catégorisation de scènes filtrées en BFS et HFS. La **Session 2** évaluait les mécanismes prédictifs à travers la catégorisation de scènes hybrides. L'expérience a été menée au CHU de Lille avec l'étroite collaboration de Muriel Boucart (DR CNRS). Le recrutement des patients a été possible grâce à notre collaboration avec le Professeur Jean-François Rouland (chef de service du service d'ophtalmologie) et les passations expérimentales ont été réalisées par Elvia Ferry (interne en ophtalmologie).

2. Expérience 11 - Catégorisation de scènes filtrées en BFS et HFS et de scènes hybrides dans la vision centrale de patients atteints d'un glaucome

2.1. Méthode

2.1.1. Participants

Cinquante-trois participants ont participé à cette expérience (Tableau 5.2), dont 36 patients atteints de glaucome primitif à angle ouvert (GPAO) bilatéral (20 femmes ; âge moyen : $68,08 \pm 10,42$) et 17 participants sains appariés en âge (10 femmes ; âge moyen : $67,65 \pm 11,73$). Tous les participants ont réalisé la **Session 1** de l'Expérience. Nous avons exclu 1 participant du groupe Contrôle des analyses statistiques en raison de ces données aberrantes. Les *outliers* (i.e. participants dont les données diffèrent significativement des autres et faussent les résultats) ont été supprimées à l'aide du package *Outliers* utilisé sur Rstudio. Les analyses statistiques ont donc été conduites sur 52 participants (30 femmes ; âge moyen : $68,04 \pm 10,82$). Parmi ces participants, 28 ont participé à la **Session 2** de cette expérience, dont 18 patients atteints de GPAO bilatéral (7 femmes ; âge moyen $68,84 \pm 10,54$) et 9 participants sains appariés en âge (6 femmes ; âge moyen : $65,78 \pm 13,07$).

Tableau 5.2 : Données individuelles cliniques des 52 participants de l'Expérience 11. F = femme ; H = homme ; OD = oeil droit ; OG = oeil gauche ; X = données manquantes.

Groupe	Participant	Sexe	Age	Œil testé	Indice MD	Atteinte centrale	Acuité visuelle (Monoyer - Parinaud)	Sensibilité au contraste	Session 2
Contrôle	1	F	68	OD			10/10 - P2	1.95	Oui
	2	F	52	OD			10/10 - P2	1.80	Oui
	3	F	79	OG			10/10 - P2	1.80	Oui
	4	F	86	OD			9/10 - P2	1.95	Oui
	5	H	59	OG			10/10 - P2	1.95	Oui
	6	F	68	OG			10/10 - P2	1.80	Oui
	7	F	50	OD			10/10 - P2	2.10	Oui
	8	H	53	OG			10/10 - P2	2.10	Oui
	9	H	77	OD			10/10 - P2	2.10	Oui
	10	H	73	OD			8/10 - P2	1.80	
	11	F	81	OG			9/10 - P2	1.65	
	12	H	58	OG			10/10 - P2	2.10	
	13	F	66	OG			10/10 - P2	1.80	
	14	F	83	OG			9/10 - P2	1.80	
	15	H	78	OG			10/10 - P2	1.95	
	16	F	56	OD			10/10 - P2	1.95	
GPAO Précoce	1	H	58	OG	-2,97		10/10 - P2	1.95	Oui
	2	F	77	OG	-4,7		10/10 - P2	1.95	Oui
	3	H	50	OG	-2,89		12/10 - P2	2.10	Oui
	4	H	65	OD	-1,43		10/10 - P2	1.80	Oui
	5	F	56	OD	1,82		10/10 - P2	1.95	Oui
	6	F	77	OG	-3,25		10f/10 - P2	1.80	Oui
	7	H	60	OG	-3,77		10/10 - P2	1.65	Oui
	8	H	53	OD	-5,71		9/10 - P2	1.80	Oui
	9	H	75	OG	-2,64		9/10 - P2	1.95	Oui
	10	F	82	OD	-2,11		9/10 - P2	1.8	
	11	H	62	OD	-1,07		9/10 - P2	1.65	
	12	F	61	OG	-0,22		10/10 - P2	1.95	
	13	H	50	OG	-5,84		9/10 - P2	1.8	
	14	F	78	OD	-5,05		8/10 - P2	1.8	
	15	H	77	OG	-2,78		9/10 - P2	1.95	
	16	F	53	OD	-5,95		9/10 - P2	1.8	
	17	H	70	OG	-2,84		10/10 - P2	1.95	
	18	F	74	OD	-2,87		10/10 - P2	2.10	
	19	F	82	OD	-1,38		9/10 - P2	1.8	
GPAO Avancé	1	H	70	OG	-7,28		10/10 - P2	1.95	Oui
	2	F	68	OD	-24		8/10 - P2	1.80	Oui
	3	H	79	OD	-7,54		9/10 - P2	1.80	Oui
	4	H	82	OD	-8,03		8/10 - P2	1.65	Oui
	5	F	81	OG	-11,54		8/10 - P2	1.65	Oui
	6	F	61	OD	-6,28		10/10 - P2	1.80	Oui
	7	H	74	OG	-15,68	Oui	9/10 - P2	1.80	Oui
	8	H	56	OD	-23,34	Oui	10/10 - P2	1.95	Oui
	9	F	54	OD	-18,07	Oui	8/10 - P2	1.65	Oui
	10	F	74	OG	-6,4		9/10 - P2	1.80	Oui
	11	F	60	OG	-6,96		9/10 - P2	1.8	
	12	F	82	OD	-6,51		10/10 - P2	1.8	
	13	F	71	OD	-8,6		10/10 - P2	1.95	
	14	F	75	OG	-15,1		10/10 - P2	1.95	
	15	F	74	OD	-6,82		9/10 - P2	1.8	
	16	F	53	OG	-6,6	Oui	10/10 - P2	1.95	
	17	H	75	OD	-13,51		9/10 - P2	1.35	

Le diagnostic de GPAO a été établi dans le service d'ophtalmologie du CHU de Lille. Les patients ont participé à cette expérience comportementale au cours de leur examen clinique. Ils ont été recrutés lors des consultations de leur suivi ophtalmologique par l'interne en ophtalmologie dans le service d'ophtalmologie. Tous les patients avaient une acuité visuelle binoculaire supérieure à 8/10 sur l'échelle de Monoyer en vision de loin (2 mètres ; équivalent

à 0,1 logMAR) et supérieure à 2 sur l'échelle de Parinaud pour la vision de près. Le champ visuel monoculaire des patients était évalué pour chaque œil à l'aide d'un analyseur périmétrique Humphrey (Carl Zeiss, Meditec, Dublin, CA) avec une procédure SITA-Standard 24-2. Cette périmétrie évalue le seuil de sensibilité sur 54 points du champ visuel, séparés de 6 degrés d'angle, dans les 24 degrés centraux en supérieur, temporal et inférieur, et dans les 30 degrés centraux en nasal. Le champ visuel était réalisé à travers des lentilles avec une correction optique adaptée à la réfraction de chaque patient. Cette périmétrie permet de calculer un indice de déviation moyenne du champ visuel ou indice MD (Mean Deviation) qui correspond à la différence entre la sensibilité rétinienne normale pour l'âge et la sensibilité rétinienne du sujet testé. Les patients étaient testés sur un seul œil. Ils ont ensuite été classés en tenant compte de la sévérité du déficit du champ visuel en utilisant la classification HPA (Hodapp et al., 1993). Cette classification comprend 4 niveaux de sévérité : précoce ($MD \leq -6dB$), modéré ($-6dB \geq MD \geq -12 dB$), avancé ($-12dB \geq MD \geq -20dB$) et sévère ($MD \geq -20dB$). Dans cette expérience, nous avons décidé de classer les patients en 2 groupes : groupe précoce ($MD \leq -6dB$; $n = 19$) versus groupe avancé qui regroupe en fait les niveaux modéré, avancé et sévère de la classification de Hodapp ($MD \geq -6dB$; $n = 17$), afin d'obtenir des groupes équilibrés.

Durant l'étude, les patients ont été testés sur un seul œil. Dans le cas où les deux yeux correspondaient aux critères d'inclusion, l'œil testé a été choisi afin de correspondre à l'équilibre des groupes (GPAO Précoce ou GPAO Avancé) ou aux contraintes de la consultation ayant précédé l'inclusion (dilatation pupillaire ou réalisation de laser sur un œil). Les patients ne devaient pas présenter d'autres pathologies oculaires, ni de pathologies neurologiques, psychiatriques ou de traitements affectant le fonctionnement visuel et cérébral.

Les participants volontaires sains, appariés en âge, ont également été testé sur un seul œil. Ces participants ont également réalisé un examen du champ visuel à l'aide de la périmétrique Humphrey (SITA-Standard 24-2) sur chaque œil afin de s'assurer qu'il n'y avait pas de déficit de sensibilité rétinienne. Tous les participants contrôles avaient une acuité visuelle binoculaire supérieure à 8/10 sur l'échelle de Monoyer en vision de loin (2 mètres ; équivalent à 0,1 logMAR) et supérieure à 2 sur l'échelle de Parinaud pour la vision de près. Les participants contrôles ne devaient pas présenter d'autres pathologies oculaires, ni de pathologies neurologiques, psychiatriques ou de traitements affectant le fonctionnement visuel et cérébral.

Avant de participer à l'expérience, chaque participant a réalisé un examen ophtalmologique complet, utilisé pour répondre aux critères d'inclusion et d'exclusion et pour explorer les relations entre certaines mesures des fonctions visuelles et des performances lors des tâches expérimentales. Les tests ophtalmologiques (mesure de l'acuité, du champ visuel, OCT : *Optical Coherence Tomography*, fonction de sensibilité au contraste : test de Pelli-Robson, présence de cataracte, présence de DMLA) ont été réalisés dans le service d'ophtalmologie de l'hôpital Huriez par l'interne. L'existence d'antécédents neurologiques ou psychiatriques ont été vérifiée par l'interne par des questions aux patients et à leur famille ainsi que par une vérification sur le logiciel Sillage du CHU de Lille comportant des informations médicales sur les patients. Le test MMSE ont également été réalisés par l'interne le jour de la consultation. Les participants présentant des pathologies psychiatriques, neurologiques ou oculaires (atrophie maculaire, néovaisseaux maculaires, drusens, cataracte non opérée ou opérée mais avec complication) n'étaient pas inclus dans l'étude et aucun participant ne

présentait de déficits cognitifs (Mini Mental State Examination > 24/30). Enfin, tous les participants ont donné leur consentement libre et éclairé avant de participer à cette expérience qui a été approuvée par le Comité d'Éthique en sciences comportementales de l'Université de Lille.

2.1.2. Stimuli et procédure

Dans cette expérience, nous avons utilisé les scènes filtrées en BFS et HFS et les images hybrides de l'**Expérience 1** (Figure 5.5). Pour les images hybrides, nous avons décidé de sélectionner uniquement les paires de scènes qui avaient des propriétés physiques dissimilaires. Les deux scènes pouvaient être soit sémantiquement congruentes (condition Congruent ; par exemple, une scène naturelle en BFS superposée à une autre scène naturelle en HFS), soit sémantiquement incongruente (condition Incongruent ; par exemple, une scène naturelle en BFS superposée à une scène artificielle en HFS). Nous avons utilisé 160 scènes filtrées qui composent les images hybrides (80 en BFS et 80 en HFS) et 80 images hybrides physiquement dissimilaires (40 sémantiquement congruentes et 40 sémantiquement incongruentes).

L'expérience a été réalisée dans une pièce dédiée à la recherche au service d'ophtalmologie du CHU de Lille. Les participants étaient installés dans l'obscurité face à un écran d'ordinateur. Les stimuli ont été affichés à l'aide du Psychtoolbox (Brainard, 1997 ; Pelli, 1997) implémenté dans MATLAB R2019 (MathWorks, Natick, MA, USA) sur un écran de 21 pouces (46,5 × 26,2 cm), avec un affichage de résolution 1920 × 1080 pixels et un taux de rafraîchissement de 60 Hz. Les participants étaient installés à 61 cm de l'écran afin de respecter la taille angulaire de 6 × 6° d'angle visuel des stimuli.

Tous les participants réalisaient deux sessions expérimentales en vision monoculaire. Dans la première session expérimentale, nous avons présenté en vision centrale les scènes filtrées en BFS et HFS qui composaient les images hybrides. Dans la deuxième session expérimentale, nous avons présenté les images hybrides en vision centrale. Les participants devaient catégoriser les scènes filtrées en BFS et HFS dans la Session 1 et les scènes en HFS tout en ignorant les scènes en BFS dans la Session 2.

Pour chaque session expérimentale, un essai se déroulait comme suit : un point de fixation central (noir sur un fond gris de luminance 0,5) était présenté pendant 500 ms et suivi de la présentation du stimulus (Session 1 : scènes filtrées en BFS et en HFS ; Session 2 : images hybrides) sur un fond gris (luminance 0,5) pendant 100 ms et d'un écran gris (luminance 0,5) qui restait affiché au cours duquel les participants pouvaient répondre. Nous avons utilisé une réponse manuelle de type *go/no-go*. Les participants devaient appuyer sur la barre espace lorsqu'une scène filtrée (Session 1) ou la scène cible d'une scène hybride (Session 2) appartenait à une catégorie cible (essais *go* ; la catégorie 'naturelle' pour la moitié des participants de chaque groupe et la catégorie 'artificielle' pour l'autre moitié) et ils ne devaient rien faire lorsqu'elle appartenait à la catégorie distractive (essais *no-go*). L'utilisation d'une réponse manuelle de type *go/no-go* est assez courante dans les études menées auprès de patients ophtalmologiques (voir par exemple, Boucart et al., 2008 ; Roux-Sibilon et al., 2018). Cela permet de diminuer la charge cognitive de patients qui souffrent déjà de difficultés visuelles en

leur demandant de n'appuyer que sur une touche (ici, la barre espace) plutôt que d'associer deux touches à deux réponses différentes. Les participants avaient pour consigne de répondre le plus correctement et rapidement possible.

La Session 1 comprenait 160 essais (40 scènes en BFS de la catégorie 'naturelle', 40 scènes en BFS de la catégorie 'artificielle', 40 scènes en HFS de la catégorie 'naturelle' et 40 scènes en HFS de la catégorie 'artificielle'). La Session 2 comprenait 160 essais (Congruent : 40 images hybrides 'naturelle' et 40 images hybrides 'artificielle' ; Incongruent : 80 images hybrides). Pour chaque essai, nous avons enregistré l'exactitude de la réponse et le temps de réponse (en millisecondes). Avant de commencer chaque session expérimentale, les participants réalisaient quelques essais d'entraînement avec des stimuli non utilisés dans l'expérience afin de se familiariser avec la tâche et les stimuli.

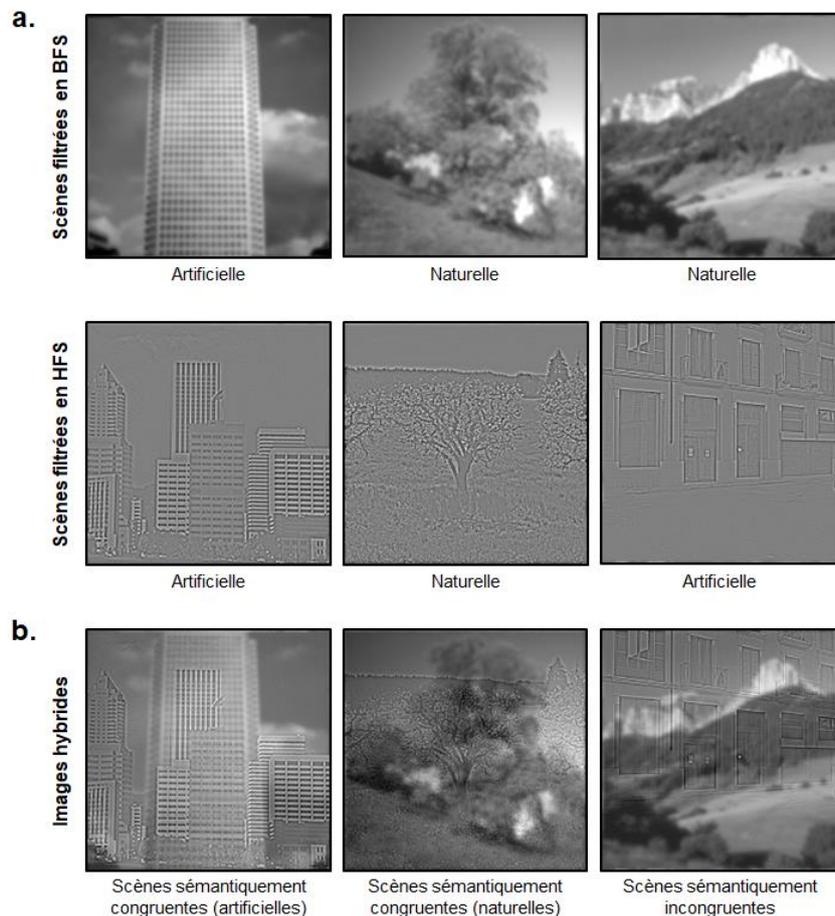


Figure 5.5 : Exemples de stimuli utilisés dans l'**Expérience 11** : **(a)** Scènes filtrées en BFS et en HFS ; **(b)** Images hybrides créées en superposant une scène filtrée en BFS et une autre scène filtrée en HFS. Les deux scènes sont physiquement dissimilaires. Elles pouvaient être soit sémantiquement congruentes (e.g., une scène naturelle en BFS superposée à une autre scène naturelle en HFS), soit sémantiquement incongruentes (e.g., une scène naturelle en BFS superposée à une scène artificielle en HFS).

2.2. Analyse des résultats

Pour chaque participant et pour chaque condition expérimentale de chaque session, nous avons calculé l'indice de sensibilité d' . Cet indice est utilisé par la théorie de la détection du signal afin d'estimer les capacités sensorielles des participants (Bonnet, 1986). Dans le cas de notre expérience, il nous renseigne sur la capacité des participants à identifier la catégorie d'une scène. Il combine le taux de détection correcte de la scène cible (par exemple, lorsqu'un participant appuyait sur la barre espace pour répondre qu'il avait vu une scène artificielle et qu'il s'agissait bien d'une scène artificielle) et le taux de fausses alarmes (par exemple, lorsque le participant appuyait alors qu'il s'agissait d'une scène naturelle). Dans des analyses additionnelles, nous avons également calculé les pourcentages d'erreurs moyens (%ERm) et les temps de réponses correctes moyens en millisecondes (TRm). Le seuil de significativité a été fixé à 0,05. La taille d'effet a été estimée par le calcul de l'éta-carré partiel (η^2).

2.2.1 Session 1 : Catégorisation de scènes filtrées en BFS et en HFS en vision centrale

Les données de 52 participants ont été analysées dans cette session expérimentale, dont 16 participants dans le groupe Contrôle (10 femmes ; âge moyen : $67,94 \pm 12,05$), 19 patients dans le groupe GPAO Précoce (9 femmes ; âge moyen : $66,32 \pm 10,87$) et 17 patients dans le groupe GPAO Avancé (11 femmes ; âge moyen : $69,94 \pm 9,73$). Les trois groupes ne différaient pas significativement en âge ($F(1,50) = 1,79, p = .189$).

Dans un premier temps, nous avons mené trois ANOVA de type 2×2 sur les indices de sensibilité d' , les %ERm et les TRm (Tableau 5.3) avec comme facteur intra-sujet la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et comme facteur inter-sujet le Groupe (Contrôle ; GPAO Précoce ; GPAO Avancé). Nous avons ensuite mené une analyse de corrélation sur ces mêmes valeurs afin de tester la relation entre l'indice MD des patients (indépendamment de leur groupe) et leurs performances obtenues pour catégoriser les scènes en BFS d'une part et les scènes en HFS d'autre part.

Tableau 5.3 : Résultats de la Session 1 de l'Expérience 11. Indices de sensibilité d' , pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et du Groupe (Contrôle ; GPAO Précoce ; GPAO Avancé).

	Contrôle			GPAO Précoce			GPAO Avancé		
	d'	%ERm	TRm	d'	%ERm	TRm	d'	%ERm	TRm
BFS	$3,77 \pm 0,27$	$5,08 \pm 2,24$	691 ± 84	$3,88 \pm 0,25$	$4,58 \pm 2,11$	622 ± 75	$3,13 \pm 0,25$	$10,76 \pm 2,11$	680 ± 81
HFS	$3,33 \pm 0,26$	$8,44 \pm 2,80$	733 ± 91	$3,59 \pm 0,25$	$5,90 \pm 2,64$	677 ± 82	$2,41 \pm 0,24$	$18,75 \pm 2,64$	777 ± 89

Les ANOVA menée sur les indices d' , les %ERm et les TRm ont toutes révélé un effet principal de la Fréquence spatiale (d' : $F(1,50) = 20,74, p < .001, \eta^2 = .29$; %ERm : $F(1,50) = 11,88, p = .001, \eta^2 = .19$; TRm : $F(1,50) = 12,56, p < .001, \eta^2 = .20$; Figure 5.6). Les participants avaient une meilleure sensibilité, faisaient moins d'erreurs et étaient plus rapides pour catégoriser les scènes en BFS ($d' : 3,59 \pm 0,15$; %ERm : $6,80 \pm 1,25\%$; TRm : 664 ± 46 ms) que pour catégoriser les scènes en HFS ($d' : 3,11 \pm 0,15$; %ERm : $11,03 \pm 1,55\%$; TRm : 729 ± 50 ms). Ce résultat est cohérent avec l'avantage du traitement des BFS sur les HFS présenté en introduction de cette thèse.

Plus important au regard de nos hypothèses de recherche, l'effet principal du Groupe était significatif pour les %ERm ($F(1,50) = 5,57, p = .007, \eta^2 = .18$) et pour les indices d' ($F(1,50) = 4,85, p = .012, \eta^2 = .16$), mais pas pour les TRm ($F(1,50) < 1$). La différence entre le groupe Contrôle et le groupe GPAO Avancé était significative pour les indices d' ($F(1,50) = 5,74, p = .022$; Contrôle : $3,55 \pm 0,25$; GPAO Avancé : $2,77 \pm 0,23$) et pour les %ERm ($F(1,50) = 6,49, p = .014$; Contrôle : $6,75 \pm 2,23$; GPAO Avancé : $14,75 \pm 2,15$). Similairement, la différence entre le groupe GPAO Précoce et le groupe GPAO Avancé était significative pour les indices d' ($F(1,50) = 8,63, p = .005$; GPAO Précoce : $3,74 \pm 0,23$; GPAO Avancé : $2,77 \pm 0,23$) et pour les %ERm ($F(1,50) = 9,75, p = .003$; GPAO Précoce : $5,24 \pm 2,15$; GPAO Avancé : $14,75 \pm 2,15$). Ces résultats suggèrent que le groupe GPAO Avancé avait une moins bonne sensibilité et faisait plus d'erreurs que le groupe Contrôle et que le groupe GPAO Précoce pour catégoriser des scènes, quelles que soient les fréquences spatiales de la scène (BFS ou HFS ; Figure 5.6). En revanche, le groupe GPAO Précoce ne différait pas du groupe Contrôle ($F_s(1,50) < 1$).

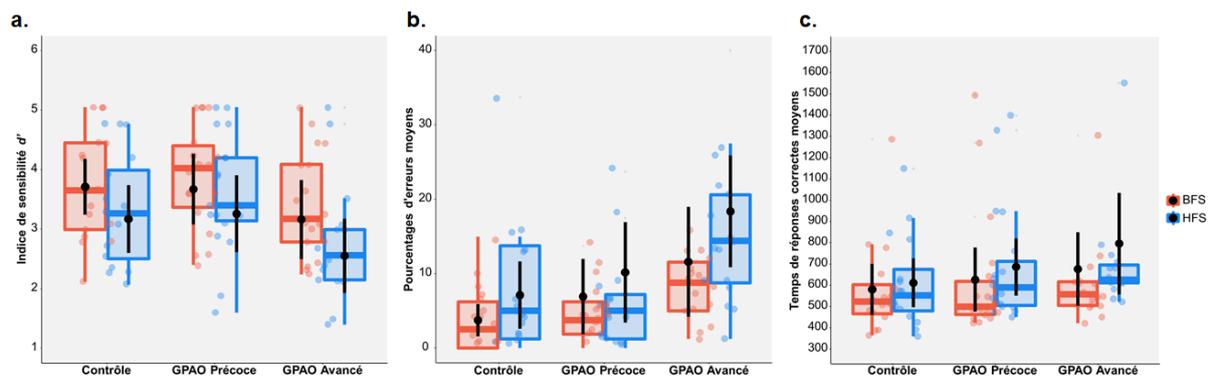


Figure 5.6 : Résultats de la Session 1 de l'Expérience 11 (catégorisation de scènes filtrées en BFS et en HFS en vision centrale) : **(a) Indices de sensibilité d'** , **(b) Pourcentages d'erreurs moyens** et **(c) Temps de réponses correctes moyens en ms** en fonction de la Fréquence spatiale de la scène à catégoriser (BFS ; HFS) et du Groupe (Contrôle ; GPAO Précoce ; GPAO Avancé). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

L'interaction entre la Fréquence spatiale et le Groupe n'était pas significative (d' : $F(1,50) = 1,52, p = .228$; %ERm : $F(1,50) = 2,69, p = .078$; TRm : $F(1,50) < 1$). Mais pour

répondre à nos hypothèses de recherche sur l'effet de la progression de la maladie, nous avons également testé cette interaction en comparant séparément chaque groupe de patients (groupe GPAO Précoce et groupe GPAO Avancé) au groupe Contrôle. En considérant le groupe GPAO Précoce en comparaison du groupe Contrôle, l'interaction Groupe \times Fréquence spatiale n'était pas significative ($F_s(1,50) < 1$). De la même façon, en considérant le groupe GPAO Avancé en comparaison du groupe Contrôle, elle n'était pas significative ($d' : F(1,50) = 1,14, p = .291$; %ERm : $F(1,50) = 2,33, p = .133$; TRm : $F(1,50) < 1$).

Par contre, en comparant les deux groupes de patients entre eux, l'interaction Groupe \times Fréquence spatiale était significative seulement pour les %ERm ($F(1,50) = 5,14, p = .028$; $d' : F(1,50) = 2,96, p = .091$; TRm : $F(1,50) < 1$). Les comparaisons planifiées réalisées sur les %ERm ont montré que le groupe GPAO Avancé faisait plus d'erreurs que le groupe GPAO Précoce aussi bien sur les BFS ($F(1,50) = 4,28, p = .0044$; GPAO Précoce : $4,58 \pm 2,11$; groupe GPAO Avancé : $10,76 \pm 2,11$) que sur les HFS ($F(1,50) = 11,83, p = .001$; GPAO Précoce : $5,90 \pm 2,64$; groupe GPAO Avancé : $18,75 \pm 2,64$). Toutefois, l'interaction montre que cette différence était plus importante pour la catégorisation des scènes en HFS qu'en BFS. Ces résultats suggèrent donc que l'avancée dans la maladie impacte le traitement des fréquences spatiales, ce déficit étant plus marqué pour le traitement des HFS. Enfin, le groupe GPAO Avancé faisait significativement plus d'erreurs pour catégoriser les HFS que les BFS ($F(1,50) = 14,76, p < .001$), alors que l'avantage des BFS n'était pas significatif pour le groupe GPAO Précoce.

Nous avons ensuite testé la relation entre l'indice MD des patients (GPAO Précoce et GPAO Avancé) et les performances de catégorisation de scènes en BFS et HFS en effectuant une analyse de corrélation entre l'indice MD et les performances pour chaque condition de Fréquence spatiale (BFS et HFS)⁶. La corrélation n'était pas significative, ni pour les BFS ($d' : r = -.18, p = .276$; %ERm : $r = -.16, p = .316$; TRm : $r = -.04, p = .811$), ni pour les HFS ($d' : r = -.24, p = .145$; %ERm : $r = -.23, p = .167$; TRm : $r = -.01, p = .954$).

2.2.2. Session 2 : Catégorisation d'images hybrides en vision centrale

Les données de 28 participants ont été analysées dans cette session expérimentale, dont 9 participants dans le groupe Contrôle (6 femmes ; âge moyen : $65,78 \pm 13,07$), 9 patients dans le groupe GPAO Précoce (3 femmes ; âge moyen : $63,44 \pm 10,55$) et 10 patients dans le groupe GPAO Avancé (5 femmes ; âge moyen : $69,90 \pm 10,08$). Les trois groupes ne différaient pas significativement en âge ($F(1,30) = 3,19, p = .055$).

⁶ Nous avons mené une analyse de corrélation supplémentaire entre l'acuité visuelle des patients (GPAO Précoce et GPAO Avancé) et les performances de catégorisation pour chaque condition de Fréquence spatiale (BFS ; HFS). Cette analyse n'a pas révélé de corrélation significative, ni pour les BFS ($d' : r = .05, p = .752$; %ERm : $r = .10, p = .551$; TRm : $r = -.137, p = .418$), ni pour les HFS ($d' : r = .09, p = .582$; %ERm : $r = .01, p = .999$; TRm : $r = -.14, p = .420$).

Nous avons tout d'abord mené trois ANOVA de type 2×2 sur les indices de sensibilité d' , les %ERm et les TRm⁷ (Tableau 5.4) avec comme facteur intra-sujet la Congruence sémantique de la scène à catégoriser (Congruent ; Incongruent) et comme facteur inter-sujet le Groupe (Contrôle ; GPAO Précoce ; GPAO Avancé). Nous avons ensuite mené une analyse de corrélation sur ces mêmes mesures afin de tester la relation entre l'indice MD des patients (indépendamment de leur groupe) et l'effet d'interférence sémantique (i.e. les performances à la condition Incongruent moins les performances à la condition Congruent).

Tableau 5.4 : Résultats de la Session 2 de l'Expérience 11. Indices de sensibilité d' , pourcentages d'erreurs moyens (%ERm), temps de réponses correctes moyens (en millisecondes) et erreurs-types en fonction de la Congruence sémantique (Congruent ; Incongruent) et du Groupe (Contrôle ; GPAO Précoce ; GPAO Avancé).

	Contrôle			GPAO Précoce			GPAO Avancé		
	d'	%ERm	TRm	d'	%ERm	TRm	d'	%ERm	TRm
Congruent	2,96 ± 0,30	12,22 ± 3,109	659 ± 50	3,06 ± 0,36	6,11 ± 3,09	574 ± 50	2,43 ± 0,33	16,50 ± 2,93	636 ± 45
Incongruent	1,05 ± 0,56	45,56 ± 9,08	823 ± 109	-0,46 ± 0,68	59,17 ± 9,08	779 ± 109	-0,38 ± 0,61	56,00 ± 8,61	808 ± 99

Les ANOVA menées sur les indices d' , les %ERm et les TRm ont toutes révélé un effet principal de la Congruence sémantique (d' : $F(1,30) = 54,69$, $p < .001$, $\eta^2 = .65$; %ERm : $F(1,30) = 56,64$, $p < .001$, $\eta^2 = .69$; TRm : $F(1,30) = 12,01$, $p = .002$, $\eta^2 = .32$; Figure 5.7). Les participants avaient en moyenne une moins bonne sensibilité, faisaient plus d'erreurs et étaient plus lents pour catégoriser les images hybrides incongruentes (d' : $0,07 \pm 0,36$; %ERm : $53,57 \pm 5,15\%$; TRm : 803 ± 61 ms) que congruentes (d' : $2,82 \pm 0,19$; %ERm : $11,61 \pm 1,75\%$; TRm : 623 ± 28). Ce résultat est cohérent avec un effet d'interférence sémantique. L'effet principal du Groupe n'était par contre pas significatif (d' : $F(1,30) = 1,95$, $p = .159$; %ERm : $F(1,30) < 1$; TRm : $F(1,30) < 1$) et ce facteur n'interagissait pas avec la Congruence sémantique (d' : $F(1,30) = 1,61$, $p = .217$; %ERm : $F(1,30) = 1,06$, $p = .362$; TRm : $F(1,30) < 1$).

Pour répondre à nos hypothèses de recherche, nous avons à nouveau comparé séparément chaque groupe de patients (groupe GPAO Précoce et groupe GPAO Avancé) au groupe Contrôle, puis les deux groupes de patients entre eux. Bien que les graphiques (Figure 5.7) montrent une augmentation de l'effet d'interférence pour les patients par rapport aux participants contrôles, l'interaction Groupe \times Congruence sémantique n'était pas significative que ce soit en considérant le groupe GPAO Précoce en comparaison du groupe Contrôle (d' : $F(1,30) = 3,12$, $p = .087$; %ERm : $F(1,30) = 2,02$, $p = .168$; TRm : $F(1,30) < 1$), ou en considérant le groupe GPAO Avancé en comparaison du groupe Contrôle ($F_s(1,30) < 1$). Enfin,

⁷ Les résultats sur les TRm sont à prendre avec précaution car les taux d'omissions (i.e. lorsqu'un participant n'appuie pas sur la barre espace alors qu'il doit appuyer) étaient très élevés pour certains participants, notamment dans la condition Incongruent. Nous avons quand même décidé de les présenter dans l'analyse des Sessions 1 et 2 pour les porter à la connaissance du lecteur. Qui plus est, aucun résultat n'étant significatif, leur présentation n'impacte pas nos interprétations. Nous ne les avons donc pas considérés dans nos interprétations.

en considérant la comparaison des deux groupes de patients, l'interaction Groupe \times Congruence sémantique n'était pas significative ($F_s(1,30) < 1$).

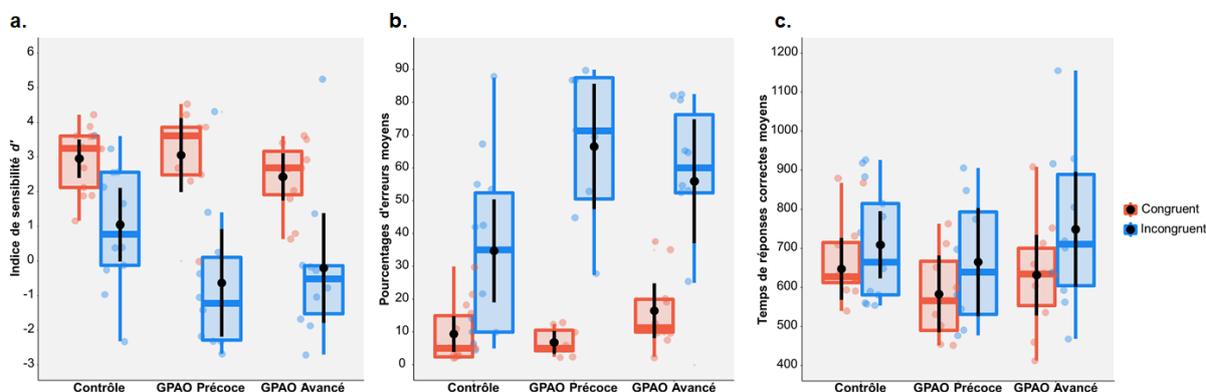


Figure 5.7 : Résultats de la Session 2 de l'**Expérience 11** (catégorisation d'images hybrides en vision centrale) : **(a) Indices de sensibilité d'** , **(b) Pourcentages d'erreurs moyens** et **(c) Temps de réponses correctes moyens en ms** en fonction de la Congruence sémantique (Congruent ; Incongruent) et du Groupe (Contrôle ; GPAO Précoce ; GPAO Avancé). Les points noirs et les barres d'erreurs indiquent l'erreur moyenne et l'erreur standard, respectivement. Les points de couleurs sont les observations individuelles.

Nous avons ensuite testé la relation entre l'indice MD des patients (GPAO Précoce et GPAO Avancé) et l'effet d'interférence sémantique en effectuant une analyse de corrélation entre l'indice MD et la différence de performances entre les conditions Incongruent et Congruent⁸. La corrélation était significative pour les %ERm ($r = -.45, p = .049$; Figure 5.8), indiquant que plus l'indice MD des patients était proche de 0 (et donc la maladie était précoce) et plus l'effet d'interférence sémantique était grand (Figure 5.8). En revanche, la corrélation n'était pas significative ni pour les indices d' ($r = -.35, p = .125$), ni pour les TRm ($r = -.17, p = .460$).

⁸ Nous avons mené une analyse de corrélation supplémentaire entre l'acuité visuelle des patients (GPAO Précoce et GPAO Avancé) et la différence de performances entre les conditions Incongruent et Congruent. Cette analyse n'a pas révélé de corrélation significative (d' : $r = -.35, p = .128$; %ERm : $r = .29, p = .216$; TRm : $r = .107, p = .653$).

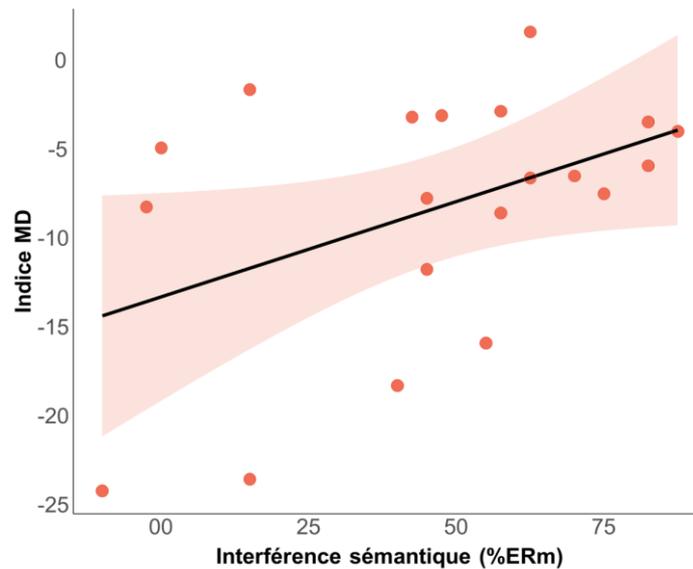


Figure 5.8 : Résultats de la Session 2 de l'**Expérience 11** (catégorisation d'images hybrides en vision centrale) : Relation entre l'indice MD des patients (GPAO Précoce et Avancé) et l'effet d'interférence sémantique (différence de performances entre les conditions Incongruent et Congruent) en pourcentages d'erreurs moyens. L'ombre représente l'intervalle de confiance à 95%. Les points de couleurs sont les observations individuelles.

3. Discussion générale

Cette expérience avait pour objectif (1) d'étudier les fonctions magno- et parvocellulaires dans la vision centrale de patients GPAO en manipulant le contenu en fréquences spatiales de scènes visuelles (**Session 1**) et (2) d'étudier les mécanismes prédictifs de ces patients engagés lors de la catégorisation de scènes en utilisant des images hybrides (**Session 2**). Dans cette expérience, nous avons considéré la progression de la maladie en constituant trois groupes : un groupe de patients atteints d'un GPAO au stade précoce, un groupe de patients atteint d'un GPAO au stade plus avancé et un groupe de participants contrôles sans atteinte périmétrique et appariés en âge aux deux groupes de patients.

Concernant tout d'abord les résultats des participants du groupe Contrôle, nous avons observé un avantage du traitement des BFS sur les HFS dans la Session 1 et un effet d'interférence sémantique des BFS sur les HFS dans la Session 2. Ces résultats reproduisent les effets classiques de la littérature que nous avons rapportés tout au long de cette thèse. En comparant de façon descriptive ces résultats obtenus auprès de personnes âgées (plus de 50 ans) à ceux obtenus dans l'**Expérience 1** de cette thèse, qui est identique à la présente expérience mais que nous avons menée auprès de personnes plus jeunes (moins de 30 ans), nous avons remarqué que les personnes âgées étaient globalement moins précises et plus lentes pour catégoriser les scènes en BFS et HFS (BFS : 5,08% pour le %ERm et 691 ms pour le TRm ; HFS : 8,44% pour le %ERm et 733 ms pour le TRm) que les personnes plus jeunes (BFS : 3,91% pour le %ERm et 637 ms pour le TRm ; HFS : 3,52% pour le %ERm et 642 ms pour le TRm). Cette baisse des performances semble affecter en particulier le traitement des HFS.

Plusieurs études comportementales ont mis en évidence une altération des fonctions visuelles au cours du vieillissement normal, en particulier de l'acuité visuelle (Elliott, Yang, & Whitaker, 1995 ; Gittings & Fozard, 1986 ; Ramanoël et al., 2015 ; Rubin et al., 1997). Dans une étude en IRMf, Ramanoël et al. (2015) ont observé que des participants âgés présentaient des difficultés pour catégoriser des scènes en HFS semblables à celles que nous avons utilisées dans notre expérience. Ce déficit lié au vieillissement normal s'accompagnait au niveau cortical d'une hypoactivité du cortex occipital par rapport à des participants jeunes. Bien que nous n'ayons pas pu tester statistiquement les différences de performances entre jeunes et âgés dans notre expérience, le seul vieillissement du système visuel pourrait rendre compte des différences de résultats entre la présente expérience et l'**Expérience 1**.

De plus, nous avons remarqué que l'effet d'interférence sur les %ERm était plus important pour les participants âgés (environ 33% d'erreurs supplémentaires dans la condition incongruente que congruente ; Congruent : $12,22 \pm 2,24$ % ; Incongruent : $45,56 \pm 2,80$ %) que les participants jeunes (environ 6% d'erreurs supplémentaires ; Congruent : $7,43 \pm 1,26$ % ; Incongruent : $12,96 \pm 1,94$ %). Cette différence notable pourrait résulter du déficit du traitement des HFS observé chez les âgés. Afin de compenser ce déficit, leur système visuel utiliseraient davantage les BFS disponibles afin de générer des prédictions sur la catégorie de la scène. Dans le cas où les deux scènes d'une image hybride sont sémantiquement incongruentes, ces prédictions erronées gêneraient alors fortement la catégorisation de la scène en HFS. Autrement dit, chez les personnes âgées, le système visuel augmenterait le poids de l'interférence des prédictions issues de l'analyse en BFS afin de compenser les déficits d'une analyse détaillée portée par les HFS. Ici aussi, le vieillissement du système visuel pourrait rendre compte des différences de résultats entre la présente expérience et l'**Expérience 1**.

En ce qui concerne à présent les résultats des patients GPAO, nous avons tout d'abord fait l'hypothèse d'un déficit du traitement des fréquences spatiales pour les patients par rapport aux participants contrôles. Toutefois, nous nous attendions à ce que ce déficit se manifeste par une altération spécifique du traitement des BFS en début de maladie. Les résultats n'ont pas mis en évidence de différence entre le groupe GPAO Précoce et le groupe Contrôle. Nous pensions que la perte de stimulation de la partie périphérique de la rétine chez les patients allait impacter la fonction magnocellulaire, et donc le traitement des BFS. Il est en fait également possible qu'au début de la maladie, la perte de la vision périphérique incite les patients à compenser leur déficit propre à la vision périphérique en mobilisant davantage les cellules ganglionnaires magnocellulaires disponibles en vision centrale, bien qu'en plus faible quantité. Grâce à cette stratégie de compensation, les patients au stade précoce de la maladie seraient capables de traiter aussi efficacement que les participants contrôles les informations en BFS en vision centrale. Ce mécanisme compensatoire disparaîtrait ensuite se manifestant par le déficit que nous avons observé chez le groupe GPAO Avancé. Alternativement, nous avons observé que les performances entre les participants du groupe Contrôle étaient extrêmement hétérogènes (par exemple, comme on peut le voir sur la Figure 5.6b, un participant du groupe contrôle pouvait présenter jusqu'à 30% d'erreurs dans la condition HFS sans toutefois être considéré comme *outliers*). Cette hétérogénéité pourrait simplement être induite par le vieillissement normal, mais augmenterait la difficulté à détecter les effets attendus.

En dépit de ce dernier biais, les performances des patients GPAO Avancé différaient de celles des participants contrôle aussi bien pour la catégorisation des scènes en BFS qu'en HFS. De plus, les performances des patients GPAO Avancé différaient également de celles des patients GPAO Précoce, cette différence étant plus marquée pour les HFS. Ainsi, l'avancée dans la maladie impacterait le traitement des fréquences spatiales et en particulier celui des HFS. Il est toutefois important de noter que quatre patients du groupe GPAO Avancé présentaient une atteinte centrale. Une atteinte du champ visuel central était définie par la présence d'au moins un point parmi les quatre points testés au centre du champ visuel, avec une probabilité inférieure ou égale à 2% d'être normal. Il est alors possible que le déficit du traitement des HFS au stade avancé de la maladie soit directement lié à l'atteinte de la rétine centrale d'où sont issues en majorité les cellules parvocellulaires utilisées pour traiter les HFS. Nous avons toutefois reconduit une ANOVA sur les %ERm en supprimant de l'analyse ces quatre patients de notre échantillon de 17 patients et les résultats ont à nouveau montré que les patients du groupe Avancé faisaient significativement plus d'erreurs que les participants du groupe Contrôle quel que soit les fréquences spatiales de la scène à catégoriser ($F(1,44) = 7,65$, $p = .008$). En résumé, les résultats de la Session 1 montrent qu'une perte de vision périphérique impacte le traitement des fréquences spatiales en vision centrale et suggère donc qu'une perte de la stimulation de la partie périphérique de la rétine impacte les fonctions magno- et parvocellulaires avec l'avancée dans la maladie. Ces altérations se manifesteraient alors dans l'ensemble du champ visuel.

Nous nous attendions également à ce que la perte de vision périphérique impacte les mécanismes prédictifs impliqués dans la reconnaissance de scènes. Nous avons étudié ces mécanismes en évaluant l'interférence sémantique d'une scène en BFS sur la catégorisation d'une scène en HFS dans des images hybrides. Nos résultats n'ont pas mis en évidence de différences entre les groupes. L'interaction Groupe \times Congruence sémantique n'était pas significative que ce soit en considérant le groupe GPAO Précoce en comparaison du groupe Contrôle, ou en considérant le groupe GPAO Avancé en comparaison du groupe Contrôle, bien que le graphique des %ERm (Figure 5.7b) montre une augmentation de l'effet d'interférence pour les patients par rapport aux participants contrôles. En effet, on peut voir que l'effet d'interférence est presque deux fois plus important pour les patients du groupe Précoce (environ 53% d'erreurs supplémentaires dans la condition incongruente que congruente) que pour les participants du groupe Contrôle (environ 33% d'erreurs supplémentaires). Encore une fois, on peut remarquer que l'hétérogénéité des performances des participants au sein du groupe Contrôle a pu augmenter la difficulté à détecter les effets attendus, d'autant plus que dans la Session 2, la taille des échantillons était très faible dans chaque groupe (Contrôle : $n = 9$; GPAO Précoce : $n = 9$; GPAO Avancé : $n = 10$). De plus, la répartition a priori des patients dans deux groupes GPAO Précoce et GPAO Avancé n'est peut-être pas idéale car le groupe de patients GPAO Avancé regroupe en réalité trois niveaux de sévérité de la classification HPA (Hodapp et al., 1993) : les niveaux modéré ($-6\text{dB} \geq \text{MD} \geq -12 \text{dB}$), avancé ($-12\text{dB} \geq \text{MD} \geq -20\text{dB}$) et sévère ($\text{MD} \geq -20\text{dB}$). Pour nous affranchir de ce biais méthodologique, nous avons analysé la corrélation entre l'indice MD des patients tout groupe confondu et leur effet d'interférence. Ces analyses complémentaires ont montré que plus l'indice MD était élevé (i.e. proche de 0), et donc la maladie précoce, plus l'effet d'interférence sémantique était important. Ainsi, ces résultats

suggèrent que les patients alloueraient plus de poids au traitement des BFS qu'au traitement des HFS au début de la maladie et que ce déséquilibre disparaîtrait avec l'avancée dans la maladie. Sur la base de précédents travaux utilisant le même paradigme expérimental (Kauffmann, Bourgin, Guyader & Peyrin, 2015 ; Kauffmann et al., 2017), nous avons interprété cet effet d'interférence sémantique comme la manifestation des mécanismes prédictifs qui interviennent dans la reconnaissance de scènes. Plus précisément, nous avons vu dans le modèle **FP-Péri** que le système visuel s'appuierait sur une première représentation grossière issue des BFS disponible dans l'ensemble du champ visuel, mais majoritairement en vision périphérique, pour générer rapidement des prédictions utiles à la reconnaissance d'un stimulus visuel en vision centrale. Dans le cas du glaucome, à cause de la perte progressive de la stimulation de la rétine périphérique, les patients ne pourraient plus utiliser de façon optimale ces mécanismes prédictifs. Nos résultats suggèrent donc que le glaucome induirait une altération progressive des mécanismes prédictifs issus de la vision périphérique. Ainsi, en montrant que la perte de vision périphérique peut moduler les effets d'interférence sémantique en vision centrale, notre expérience démontre le rôle primordial que joue la vision périphérique dans la génération de prédictions en conditions normales.

Enfin, bien que la comparaison statistique entre les deux groupes de patients et le groupe Contrôle ne soit pas significative dans la Session 2, on peut se poser la question de la forte augmentation de l'effet d'interférence pour les patients du groupe GPAO Précoce par rapport au groupe Contrôle. Une fois de plus, il est possible qu'au début de la maladie, l'installation de la perte de vision périphérique contraindrait les patients à très rapidement mobiliser davantage les ressources disponibles dans la vision centrale encore intacte et à renforcer l'utilisation des BFS qui y sont disponibles pour générer des prédictions. Ce mécanisme compensatoire disparaîtrait alors progressivement comme en témoigne la diminution de l'interférence sémantique avec l'avancée dans la maladie, sans doute du fait de l'apparition du déficit du traitement des fréquences spatiales. Ce mécanisme compensatoire est cohérent avec celui que nous avons proposé pour expliquer l'absence de différence entre le groupe Contrôle et le groupe GPAO Précoce dans la Session 1. Dans ce contexte, le traitement exacerbé des BFS serait le reflet d'un dysfonctionnement de la voie magnocellulaire au début de la maladie.

Dans les deux sessions de cette expérience, nous avons donc observé que la perte de vision périphérique induisait des modifications du traitement des informations présentées en vision centrale. Ces réorganisations fonctionnelles observées au niveau comportemental pourraient être la conséquence d'une réorganisation au niveau cortical chez les patients atteints d'un glaucome. Dans ce contexte, plusieurs auteurs suggèrent que le glaucome impacterait la cognition visuelle en raison d'une réorganisation cérébrale (Lenoble et al., 2016 ; Roux-Sibilon et al., 2018). De notre point de vue, la destruction progressive des cellules ganglionnaires rétiniennes dans le glaucome pourrait déclencher une dégénérescence trans-synaptique dans les noyaux géniculés latéraux et dans le cortex visuel (Gupta & Yücel, 2007 ; Weber et al., 2000, Yücel et al., 2001) qui provoquerait des changements structurels et fonctionnels dans des zones cérébrales de haut niveau, affectant la fonction visuelle dans son ensemble et donc, dans l'ensemble du champ visuel. Plusieurs données issues d'études en IRM confortent l'hypothèse d'importants changements corticaux anatomiques et fonctionnels liés à une dégénérescence neuronale dans le glaucome. Par exemple, grâce à la technique de morphométrie cérébrale (ou

technique VBM pour *voxel-based morphometry*), Boucard et al. (2009) ont montré que la densité de la matière grise des patients GPAO était réduite dans la partie médiane du cortex occipital antérieur par rapport à des participants contrôles. Cette région du cerveau correspondant approximativement aux projections du champ visuel périphérique, cette étude suggère que l'atteinte des cellules rétinienne périphériques entraîne directement des modifications anatomiques dans le cortex visuel. D'autres études utilisant cette technique ont révélé des anomalies dans de nombreuses autres aires cérébrales visuelles et non visuelles de patients glaucomateux (Chen et al., 2013 ; Frezzotti et al., 2014 ; Wang et al., 2016). Par ailleurs, des études utilisant cette fois-ci l'IRMf ont mis en évidence des altérations de l'activation du cortex visuel primaire chez ces patients (Duncan et al., 2007 ; Wang et al., 2016 ; Williams et al., 2013). Par exemple, Duncan et al. (2007) ont montré une altération du signal IRMf dans les régions du cortex visuel qui correspondent à la projection des parties du champ visuel atteintes des patients glaucomateux. Plus intéressant en ce qui concerne notre problématique de recherche, les mesures IRM de Frezzotti et al. (2014) ont montré de nombreuses anomalies morphologiques dans des aires visuelles intégratives qui supportent la reconnaissance d'objets ou de scènes. Ils ont notamment observé une atrophie dans des régions corticales impliquées dans la reconnaissance des objets (le *lateral occipital complex* ; Grill-Spector, Kourtzi & Kanwisher 2001) et des scènes (le gyrus parahippocampique englobant la *parahippocampal place area* ; Epstein & Kanwisher et al., 1998). Frezzotti et al. (2016) ont ensuite montré des altérations de la connectivité anatomique et fonctionnelle au sein des réseaux du cortex visuel, mais aussi des réseaux de la mémoire de travail et de l'attention. Ces deux études constitueraient, selon les auteurs, une preuve que le glaucome doit être considéré comme une maladie neurodégénérative généralisée. Une autre étude en IRMf s'est intéressée à la connectivité fonctionnelle des régions du cortex visuel (Sabbah et al., 2017), mais cette fois-ci avec des patients atteints de rétinopathie pigmentaire, une maladie rétinienne qui entraîne la destruction progressive des photorécepteurs causant une perte de la vision périphérique avec une vision dite en 'tunnel'. Les auteurs ont observé un résultat important concernant les régions du cortex occipital qui correspondent à la projection des parties du champ visuel préservés des patients, c'est-à-dire le champ visuel central (régions afférentées). Les auteurs ont observé une augmentation de la connectivité fonctionnelle entre les régions afférentées et les régions plus intégratives impliquées dans le traitement des scènes, de l'espace et de l'intégration multisensorielle (dans les aires *middle occipital gyrus*, *superior temporal sulcus* et *superior temporal gyrus*). Selon ces auteurs, la puissance des traitements serait alors exacerbée afin de compenser la perte des fonctions atteintes. Cette réorganisation cérébrale pourrait alors rendre compte de la capacité des patients à mettre en place des stratégies de compensation comme nous en avons fait l'hypothèse à plusieurs reprises pour le groupe GPAO Précoce.

Enfin, les altérations observées dans les aires visuelles intégratives pourraient expliquer pourquoi les patients de notre expérience, ainsi que ceux des études de Lenoble et al. (2016) et Roux-Sibilon et al. (2018) présentent des déficits dans leur vision centrale pourtant considérée comme préservée. Lors de sa thèse de doctorat, Roux-Sibilon (2020) a testé cette hypothèse en IRMf. Cette étude exploratoire qui n'a été menée qu'auprès de trois patients GPAO et quatre participants contrôles montrent toutefois que la *parahippocampal place area* (région spécifiquement impliquée dans la perception de scènes) était moins activée chez les patients

que chez les participants contrôles. Il faut savoir que cette région est plus sensible à des stimulations dans le champ visuel périphérique que central (Arcaro et al. 2009 ; Baldassano et al. 2016 ; Hasson et al. 2002 ; Levy et al. 2001 ; Malach et al. 2002). La perte de vision périphérique dans le glaucome pourrait ainsi tout particulièrement perturber le fonctionnement de la PPA avec, en conséquence, des répercussions sur la perception de scènes dans l'ensemble du champ visuel.

Pour conclure, dans ce chapitre, nous avons eu l'opportunité de travailler auprès de patients avec une perte de sensibilité rétinienne en vision périphérique. A cause de leur perte, nous n'avons pas pu tester directement les mécanismes prédictifs dans cette partie du champ visuel comme nous l'avons fait dans les expériences de cette thèse menée auprès de participants sains. Cependant, avec une approche expérimentale empruntée à la neuropsychologie, nous avons pu étudier les conséquences d'une perte de vision périphérique sur la catégorisation de scènes en vision centrale afin d'en déduire le rôle de la vision périphérique dans des conditions de vision normale. Nos résultats apportent de nouveaux arguments empiriques en faveur du modèle **FP-Peri** que nous allons présenter en détail dans la Discussion générale de cette thèse qui suit.

Chapitre 6 – Discussion générale

Le système visuel disposerait d'une première représentation globale et rudimentaire d'une scène visuelle grâce au traitement rapide des basses fréquences spatiales (BFS) du signal visuel. Cette représentation serait très utile pour catégoriser une scène dans le cas où une réponse rapide - bien que peu précise - est nécessaire, mais elle servirait également à générer des prédictions afin d'anticiper et guider un traitement plus minutieux de la scène. Étant donné que les BFS sont majoritairement extraites par la rétine périphérique, nous nous sommes interrogés tout au long de cette thèse sur le rôle de la vision périphérique pour catégoriser rapidement une scène, mais aussi sur son rôle dans le processus de génération de prédictions. Ceci nous a nécessairement amené à confronter les traitements réalisés en vision centrale et vision périphérique lors de la catégorisation de scènes à travers plusieurs expériences comportementales. Chaque expérience, ainsi que leurs limites, ayant été largement discutées au sein de chaque chapitre expérimental, nous proposons dans ce dernier chapitre de discuter spécifiquement de l'apport de ce travail pour le modèle **FP-Péri** présenté en chapitre théorique, et nous aborderons les perspectives de recherche.

1. Implications théoriques des résultats

1.1. Apport des résultats aux modèles existants

Le modèle **FP** (Alexia Roux-Sibilon, 2020) postule que le système visuel s'appuie sur une première représentation grossière issue des BFS d'une scène visuelle pour générer rapidement des prédictions et guider ensuite le traitement des HFS. Ce modèle a été formalisé uniquement sur la base d'expériences utilisant des stimuli présentés en vision centrale, le modèle **FP-Péri** intègre par extension les caractéristiques de la vision périphérique. Ce dernier modèle postule que les informations du champ visuel central et celles du champ visuel périphérique sont traitées de manière indépendante et en parallèle. Il comprend alors un système prédictif qui reçoit des BFS provenant toujours de la rétine centrale, mais surtout de la rétine périphérique. Les prédictions seraient alors majoritairement générées sur la base des informations présentes en vision périphérique. Elles guideraient alors le traitement ultérieur des informations plus détaillées issues des HFS (majoritairement en vision centrale).

Ce modèle est pertinent au regard des recherches qui ont mis en évidence l'importance de la vision périphérique, voir même sa supériorité sur la vision centrale, lors de la catégorisation rapide de scènes (Boucart et al., 2013 ; Geuzebroek & van den Berg, 2018 ; Larson & Loschky, 2009 ; Loschky et al., 2019). C'est d'ailleurs ce que montrent les résultats de nos propres expériences (**Trouilloud et al., 2020**) présentés dans le **Chapitre 3**. Dans ce chapitre, nous avons étudié comment différentes régions du champ visuel contribuent respectivement à la reconnaissance rapide de scènes. Nous avons pour cela utilisé des stimuli

révélant différentes parties du champ visuel à travers des anneaux circulaires de différentes excentricités rétiniennes. Nos résultats ont montré que les performances de catégorisation s'amélioraient à mesure que l'information visuelle périphérique de la scène était dévoilée, avec une zone critique (entre 7,10° et 10,70° d'excentricité visuelle). Nous avons également observé que l'analyse visuelle d'une scène était plus efficace lorsqu'elle suivait une séquence de traitement *coarse/peripheral-to-fine/central*. Par analogie avec l'hypothèse d'une analyse *coarse-to-fine* des fréquences spatiales, l'analyse rapide des informations disponibles en vision périphérique pourrait permettre une première catégorisation qui serait validée ou non par les informations en HFS disponibles en vision centrale.

Si les expériences du **Chapitre 3** nous ont permis de démontrer le rôle primordial de la vision périphérique pour catégoriser rapidement une scène visuelle, c'est à travers les expériences des **Chapitres 4 et 5** que nous avons spécifiquement étudié son rôle prédictif pour guider l'analyse de l'information présente au point de fixation visuelle. Dans le **Chapitre 4**, nos expériences nous ont permis de démontrer que la catégorisation d'une scène en vision centrale est influencée par la présence d'informations sémantiques congruentes ou incongruentes en vision périphérique, bien que ces dernières soient de faible résolution spatiale et, de prime abord, non pertinentes pour une tâche à réaliser en vision centrale. Nous avons interprété ces résultats dans le contexte des modèles prédictifs de la reconnaissance visuelle (Bar, 2003 ; Bar, 2007 ; Kauffmann, Ramanoël & Peyrin 2014 ; Roux-Sibilon, 2020) et en particulier le modèle **FP-Péri**. La scène en vision périphérique pourtant à ignorer serait automatiquement utilisée pour générer des prédictions sur la catégorie de la scène en vision centrale. Ces prédictions seraient correctes dans le cas où les scènes en vision périphérique et en vision centrale appartiennent à la même catégorie sémantique, ce qui faciliterait la catégorisation. En revanche, elles seraient erronées dans le cas où les scènes en vision périphérique et en vision centrale n'appartiennent pas à la même catégorie sémantique, ce qui entraverait la catégorisation. Par ailleurs, les travaux auxquels j'ai participé au cours de ma thèse ont montré l'importance des informations sémantiques, mais aussi des informations physiques des BFS disponibles en vision périphérique pour générer ces prédictions (**Peyrin et al., 2021** ; **Roux-Sibilon et al., 2019**). Cependant, ces travaux ne permettaient pas de déterminer la nature des caractéristiques physiques utilisées dans la génération de prédiction. En effet, dans le domaine de Fourier, une image peut être caractérisé selon deux types d'informations physiques : le spectre d'amplitude, qui fournit la distribution du contraste de luminance en fonction des orientations et des fréquences spatiales, et le spectre de phase, qui fournit la position relative de la luminance des fréquences spatiales dans l'espace de l'image. Dans le **Chapitre 4**, nous avons montré que c'est la position relative des BFS (information de phase) disponibles en vision périphérique qui serait tout particulièrement utilisée pour générer des prédictions.

Dans le **Chapitre 5**, nous avons étudié les conséquences d'une perte de la vision périphérique sur la catégorisation de scènes sur les traitements réalisés en vision centrale afin d'en déduire les mécanismes imputables à la vision périphérique en conditions normales. Nous avons mené une expérience comportementale auprès de patients atteints d'un glaucome primitif à angle ouvert. Nous avons observé qu'en vision centrale, l'influence des BFS sur la catégorisation de scènes en HFS diminuait avec l'avancée dans la maladie. Ainsi, la seule perte de vision périphérique influençait les traitements réalisés en vision centrale. Nous avons

interprété ce résultat comme étant une nouvelle fois la manifestation des mécanismes prédictifs qui interviennent dans la reconnaissance de scènes. Plus précisément, nous avons défendu l'hypothèse qu'à cause de la perte progressive de stimulation de la rétine périphérique, les patients n'utiliseraient pas de façon optimale les mécanismes prédictifs inhérents à la vision périphérique qui viennent guider les traitements réalisés en vision centrale. Ainsi, en accord avec le modèle **FP-Péri**, nos résultats que nous avons obtenus auprès de patients nous ont permis de conforter le rôle primordial que jouerait la vision périphérique dans la génération de prédictions en conditions normales.

Pour finir, le laboratoire GIPSA-lab de l'Université Grenoble Alpes développe depuis plus de 20 ans un modèle de vision artificiel qui s'inspire du fonctionnement du système visuel pour simuler la perception visuelle chez l'humain. Ce modèle qui extrait les caractéristiques visuelles de bas niveau dans l'image d'une scène telles que les fréquences spatiales et les orientations (Guyader et al., 2004 ; Hérault et al., 1997 ; Huo Phuoc, Guyader & Guérin-Dugué, 2010) se montre particulièrement efficace pour décrire la scène en termes de catégorie et de zones de saillance (c'est à dire, des régions les plus susceptibles d'attirer le regard de l'observateur). Dans une version plus récente (Vlad-Debusschere, Guyader & Guérin-Dugué, 2015), ce modèle intègre un filtrage spatialement variant afin de modéliser la diminution d'acuité visuelle avec l'excentricité rétinienne. En collaboration avec ce laboratoire, notre équipe de recherche au LPNC mène des études psychophysiques auprès de sujets humains (e.g., pour une tâche de catégorisation de scènes, voir Guyader et al., 2004) afin de valider si les simulations informatiques du modèle sont effectivement capables de prédire le comportement de sujets humains. En partant d'une démarche inverse, nos résultats comportementaux pourraient peut-être orienter des modifications du modèle artificiel existant, notamment en ce qui concerne (1) la supériorité d'une partie critique de la vision périphérique sur la vision centrale pour catégoriser des scènes et (2) les interactions entre la vision centrale et la vision périphérique.

1.2. Extension du modèle FP-Péri

Bien que menées chronologiquement au début de ma thèse, ce sont les expériences du **Chapitre 2** qui nous permettent au final de proposer une extension au modèle **FP-Péri**. Dans ce chapitre, nous avons présenté des images hybrides en vision périphérique. Dans une condition expérimentale, les participants devaient catégoriser la scène en HFS tout en ignorant la scène en BFS et nous avons observé que le traitement des HFS (des détails) était largement influencé par la présence d'informations en BFS congruentes ou incongruentes. Le traitement rapide des informations présentes en vision périphérique pourrait être utilisé pour générer des prédictions sur les informations visuelles plus détaillées qui sont également présentes en vision périphérique (Figure [6.1](#)). Au-delà de nos expériences, nous pensons que ces prédictions pourraient être utilisées pour guider les mouvements des yeux afin d'amener la vision centrale sur des détails pertinents lors d'une prochaine fixation oculaire.

Nous proposons dans la Figure [6.1](#) une extension au modèle **FP-Péri** qui inclut un mécanisme de guidage des mouvements oculaires initié par traitement rapide des BFS en vision

périphérique (analyse visuelle dite *coarse*). Dans cette figure, l'image rétinienne (Figure 6.1a) représente une image de la vie quotidienne (une espace de travail informatique) telle qu'elle devrait être perçue par le système visuel humain en termes de résolution spatiale : la résolution spatiale est la plus haute au centre (HFS) et diminue en périphérie avec l'excentricité rétinienne (BFS). Prenons alors l'exemple où nous cherchons une paire de lunettes alors que notre regard fixe le clavier (au centre de l'image). Les informations visuelles en périphérie (les lunettes à gauche, la souris et la lampe à droite) sont alors perçues comme des formes globales et floues. Dans ce cas, le traitement rapide des BFS en vision périphérique (**t1**) va permettre d'extraire suffisamment d'informations pour générer des prédictions sur les petits objets et détails présents en vision périphérique. Ces prédictions vont alors servir à déclencher une saccade oculaire (**t3**) vers les objets qui nous intéressent (la paire de lunettes ; Figure 6.1b) pour une analyse plus détaillée (analyse visuelle dite *fine*).

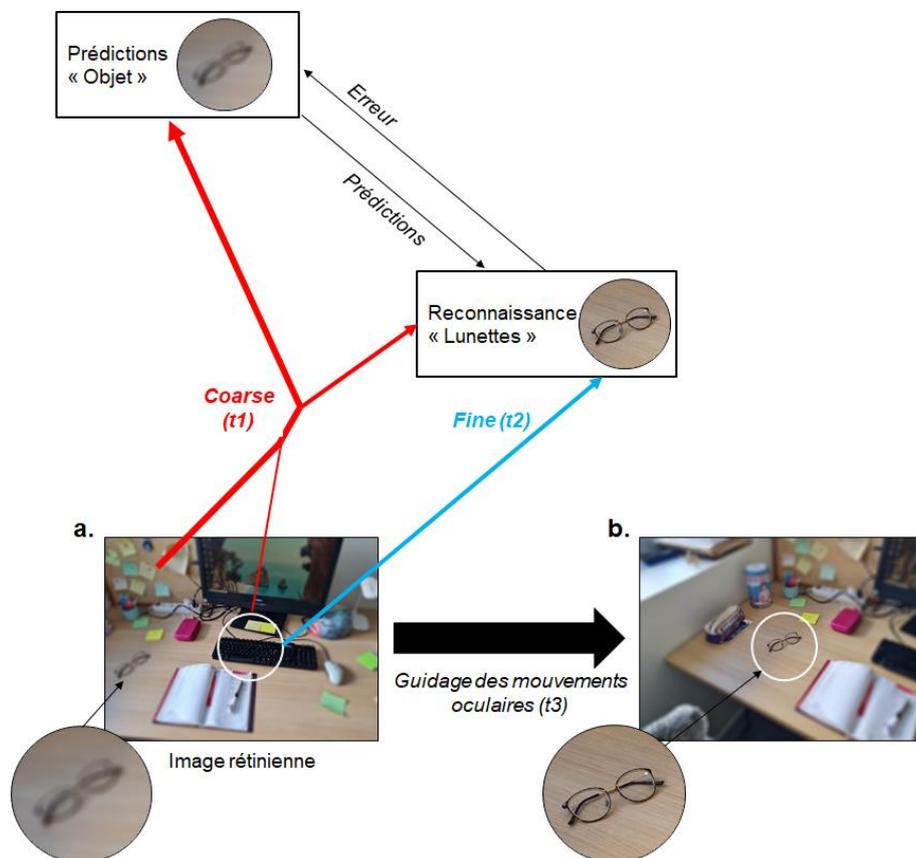


Figure 6.1 : Extension du modèle **FP-Péri**. **(a)** La fixation oculaire est sur le clavier et la paire de lunettes est perçue comme une forme globale en vision périphérique. **(b)** La fixation oculaire s'est déplacée sur la paire de lunettes grâce à un mécanisme de guidage de la vision périphérique. Les lunettes sont alors perçues comme de façon détaillée en vision centrale. Le cercle blanc représente approximativement la vision centrale.

Des expériences futures pourront nous aider à préciser les résultats des expériences du **Chapitre 2** et notre hypothèse d'un mécanisme de guidage des mouvements oculaires par les prédictions faites en vision périphérique. Notre objectif serait d'étudier spécifiquement de

quelle façon les informations d'une scène disponibles en vision périphérique influencent les mouvements oculaires en direction d'un objet d'intérêt en vision périphérique. Nous proposons de nous inspirer du paradigme de choix saccadique que nous utilisons depuis plusieurs années au LPNC pour étudier, en collaboration avec le GIPSA-lab (UGA), la perception des visages (Guyader et al., 2017 ; Kauffmann et al., 2019 ; Kauffmann et al., 2021 ; Entzman et al., sous presse). Notre idée est de présenter deux objets simultanément en vision périphérique, un dans le champ visuel gauche et l'autre dans le champ visuel droit. Les participants auraient pour consigne de réaliser une saccade le plus correctement et rapidement possible vers l'objet appartenant à une catégorie cible (par exemple, la catégorie des animaux). Afin de tester notre hypothèse de recherche, nous manipulerons alors la congruence entre la catégorie de l'objet cible et celle de la scène contextuelle présente en vision périphérique. L'objet cible pourrait être soit sémantiquement congruent avec la scène (e.g., un animal dans une scène d'extérieur ou un meuble dans une scène d'extérieur ; Figure 6.2a), soit sémantiquement incongruent (e.g., un meuble dans une scène d'extérieur ou un animal dans une scène d'intérieur ; Figure 6.2b). Etant donné que notre objectif s'adresse spécifiquement à l'influence d'une information contextuelle en vision périphérique sur les mouvements oculaires, nous proposons de masquer les informations présentes sur la fixation centrale de façon à ce qu'au début de chaque essai, le participant ne soit pas influencé par des informations présentes en vision centrale avant d'initier une saccade. Si les informations en BFS disponibles en vision périphérique sont utilisées pour générer des prédictions qui guident les mouvements oculaires sur les détails en vision périphérique, alors nous devrions observer des latences de saccades plus courtes et moins d'erreurs de saccades dans la condition congruente qu'incongruente.



Figure 6.2 : Exemple de paradigme expérimental de choix saccadique. Deux objets sont présentés simultanément, un dans le champ visuel gauche et l'autre dans le champ visuel droit. **(a)** L'objet cible (un animal en haut, un meuble en bas) est congruent avec la scène périphérique (une scène d'extérieur en haut, une scène d'intérieur en bas). **(b)** L'objet cible (un meuble en haut, un animal en bas) est incongruent avec la scène périphérique (une scène d'extérieur en haut, une scène d'intérieur en bas).

De façon cohérente avec notre modèle, plusieurs études suggèrent que les informations présentes en vision périphérique seraient utilisées pour générer des prédictions sur les informations visuelles plus détaillées présentes à la prochaine fixation oculaire. Par exemple, Herwig et Schneider (2014) ont montré que l'*input* pré-saccadique périphérique est utilisé pour générer des prédictions sur l'apparence de l'*input* post-saccadique en vision fovéale. Les auteurs présentaient aux participants deux objets en vision périphérique (un réseau sinusoïdal triangulaire ou circulaire) et ils devaient initier une saccade vers l'un d'eux afin que le nombre de saccades vers les deux objets soit égal. Pendant l'exécution de la saccade, la fréquence spatiale de l'objet en vision périphérique pouvait rester intacte (condition normale) ou changer (condition permutée) selon l'objet visé de façon à ce que les participants apprennent implicitement à faire des associations entre l'apparence de chaque objet entre la vision périphérique et la vision fovéale. Dans une phase de test ultérieure, l'un des deux objets en vision périphérique était présenté et les participants devaient initier une saccade vers cet objet. L'objet disparaissait pendant la saccade et était remplacé par un nouvel objet dont la fréquence spatiale était aléatoire. Les participants devaient ajuster la fréquence spatiale de l'objet qui se trouvait désormais en fovéa afin de correspondre à la fréquence spatiale attendue de l'objet qu'ils avaient vu au préalable en vision périphérique avant d'initier leur saccade. Ils ajustaient la fréquence spatiale en appuyant sur les touches fléchées vers le haut et vers le bas du clavier afin de respectivement augmenter ou diminuer la fréquence spatiale. Les résultats ont montré que les réponses des participants étaient influencées par la phase d'entraînement implicite. Les participants s'attendaient à ce que les objets périphériques ayant la même fréquence spatiale devaient avoir une fréquence spatiale différente lorsqu'ils étaient en fovéa, en fonction de l'association apprise pendant la phase d'acquisition. Cela indique donc que les associations entre l'apparence pré et post-saccadique des stimuli visuels sont utilisées pour prédire les caractéristiques des objets au cours des saccades.

Par ailleurs, plusieurs études ont montré que les informations périphériques permettent d'initier un mouvement saccadique vers un endroit pertinent de la scène (Cajar et al., 2020 ; Nuthmann, 2014). Par exemple, Cajar et al. (2020) ont utilisé un paradigme du contrôle du regard (*gaze-contingent paradigm*) dans lequel l'expérience affichée sur l'écran d'ordinateur s'adapte continuellement à la position du regard du participant. Cette étude avait pour objectif d'étudier comment les fréquences spatiales pouvaient affecter la recherche visuelle d'objets dans une scène. Pour cela, les auteurs ont présenté des scènes dans lesquelles ils ont filtré en BFS ou HFS les informations en vision centrale ou en vision périphérique (Figure 6.3). Les participants devaient chercher un objet cible dans la scène et décider le plus rapidement possible si l'objet était présent ou absent de la scène. Pour les taux de réussite, les résultats ont montré que la précision de la recherche visuelle diminuait considérablement uniquement lorsque la vision centrale était filtrée en BFS. Le filtrage (en BFS ou HFS) de la périphérie n'affectait pas la précision. Pour la rapidité, les résultats ont montré que les temps de localisation de la cible augmentaient lorsque la périphérie était filtrée en BFS ou en HFS, tandis que les temps de vérification de l'identité de la cible augmentaient lorsque le centre était filtré en BFS ou en HFS, surtout avec un filtrage en BFS. Ces résultats suggèrent le rôle primordial de la vision périphérique pour localiser des objets et le rôle primordial de la vision centrale pour identifier des objets.

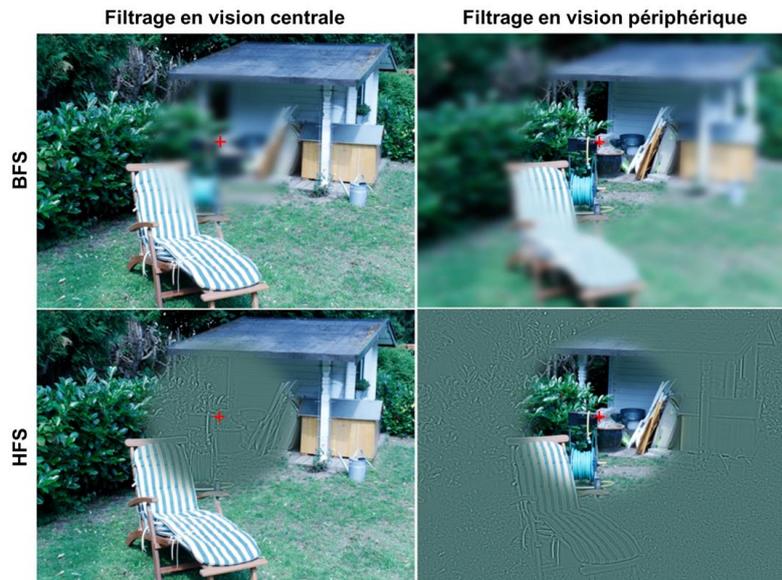


Figure 6.3 : Exemple de stimuli utilisé dans l'étud de Cajar et al. (2020). A gauche, le centre des scènes était filtré soit en BFS, soit en HFS. A droite, la périphérie des scènes était filtrée soit en BFS, soit en HFS. Figure adaptée de Cajar et al. (2020).

Dans une autre étude, Nuthmann (2014) a également utilisé le paradigme du contrôle du regard (*gaze-contingent paradigm*) afin d'étudier l'importance de la vision centrale et périphérique lors de la recherche d'un objet dans une scène. Les auteurs restreignaient le champ visuel à l'aide de masques flous de différentes excentricités, de sorte que seule la partie centrale était nette (i.e. en floutant soit la vision périphérique, soit la vision périphérique et parafovéale) ou seule la partie périphérique était nette (i.e. en floutant soit la vision fovéale, soit la vision fovéale et parafovéale). L'emplacement des masques centraux et périphériques bougeait continuellement en fonction de la position du regard. Dans chaque condition (vision centrale nette et vision périphérique nette), les participants devaient réaliser une saccade le plus rapidement possible sur un objet cible et une fois l'objet trouvé, le participant devait le fixer et appuyer sur un bouton. Les résultats ont montré que lorsque les participants recherchaient l'objet avec uniquement des informations nettes en vision fovéale, leurs précisions et leurs temps de recherche étaient significativement réduits par rapport à la condition de contrôle où toutes les informations de la scène étaient nettes. Cependant, lorsque les informations en vision périphérique étaient nettes et que les informations fovéales étaient floues, leurs performances de recherche étaient aussi bonnes que lorsque tout le contenu de l'image était net. L'auteur a ensuite analysé les taux de réussites et les temps de réponses des participants afin d'identifier quelle partie du champ visuel était utile pour (1) localiser et (2) identifier l'objet. Lorsque la partie centrale était floue, les participants n'avaient pas de difficulté à localiser l'objet, mais ils étaient plus lents à l'identifier. Lorsque la partie périphérique était floue, les résultats étaient inversés. Les participants avaient besoin de plus de temps pour localiser l'objet, mais n'étaient pas gênés pour l'identifier. Ces résultats soutiennent donc l'utilité de la vision périphérique par rapport à la vision centrale pour générer des saccades vers une partie pertinente de la scène et localiser les objets d'intérêts.

Dans l'ensemble, ces différentes études de recherche visuelle sont plus écologiques que nos expériences car elles incluent la possibilité d'effectuer des mouvements oculaires mais elles n'en restent pas moins éloignées de la réalité. En effet, dans notre vie quotidienne nous ne percevons pas des images en deux dimensions comme c'est le cas dans ces expériences, mais nous les percevons en trois dimensions et nous pouvons interagir avec en faisant par exemple des mouvements de la tête et des yeux. Dans ce contexte, David et al. (2020) ont mené une étude en réalité virtuelle dans laquelle ils ont également utilisé un paradigme de contrôle du regard. L'objectif principal était de répliquer les résultats des travaux qui ont étudié les rôles de la vision centrale et de la vision périphérique lors de la recherche visuelle d'objets mais dans des situations plus proches de la réalité. Les participants voyaient des scènes d'intérieur (salon, chambre, salle de bain, cuisine et bureau) dans leur casque de réalité virtuelle et ils devaient chercher un objet cible (e.g., une corbeille à fruits). Pendant la recherche visuelle de la cible, une partie du champ visuel était masqué, soit la partie centrale (en masquant la vision périphérique), soit la partie périphérique (en masquant la vision centrale) à l'aide de "taches aveugles". Les auteurs ont mesuré les rotations de la tête, les mouvements oculaires et les temps de fixation du regard. Les résultats ont montré que la perte de vision périphérique était davantage préjudiciable que la perte de la vision centrale sur toutes les mesures. Ce résultat soutient une nouvelle fois le rôle crucial de la vision périphérique pour localiser un objet dans le champ visuel.

2. Perspectives de recherche

2.1. Limites et perspectives théoriques

L'ensemble de nos résultats peuvent donc être interprété au regard des modèles de codage prédictif de l'information visuelle visuelle (Bar, 2003 ; Bar, 2007 ; Kauffmann, Ramanoël & Peyrin 2014) et en particulier le modèle **FP-Péri** (Roux-Sibilon, 2020). Cependant, à ce stade de réflexion nous pouvons nous interroger sur deux points concernant ce modèle. Premièrement, peut-on vraiment considérer le modèle **FP-Péri** comme un modèle purement fréquentiel ? Deuxièmement, peut-on considérer le modèle **FP-Péri** comme un modèle par défaut, qui serait flexible dans certains cas ?

Concernant notre première question, le modèle **FP-Péri** est par définition un modèle fréquentiel, c'est-à-dire qu'il suppose que c'est spécifiquement le contenu en BFS (majoritairement disponible en vision périphérique) qui contribue au déclenchement des mécanismes prédictifs qui guiderait ensuite le traitement des HFS en vision centrale. Toutefois, dans la plupart de nos expériences nous avons choisi de ne pas filtrer le contenu en fréquences spatiales de nos stimuli pour deux raisons. D'une part, la manipulation des HFS en vision périphérique nous semblait contre-nature avec les propriétés du système visuel et d'autre part, nous avons considéré que la vision périphérique agit comme un filtre passe-bas naturel, c'est-à-dire que l'information en HFS disparaît progressivement en vision périphérique tandis que l'information en BFS est préservée. De plus, éviter une telle manipulation permettait de limiter la dénaturation de nos stimuli. L'absence de filtrage de nos stimuli ne nous garantit donc pas

que c'est spécifiquement et uniquement le contenu en BFS qui contribue au déclenchement des mécanismes prédictifs. En effet, la perception d'un stimulus en vision périphérique n'est pas seulement contrainte par sa faible résolution spatiale mais aussi par un phénomène d'encombrement appelé le *crowding*. Ce phénomène implique que la perception d'un objet en vision périphérique est détériorée en présence d'éléments adjacents. Un objet isolé et facilement identifiable peut alors devenir totalement méconnaissable en présence d'autres éléments qui l'entourent en vision périphérique. Le *crowding* se définit selon la loi de Bouma qui suppose que les éléments adjacents interfèrent avec la reconnaissance d'un objet en vision périphérique lorsqu'ils sont situés à une distance égale ou inférieure à $e/2$ de l'objet, e étant la valeur d'excentricité de l'objet par rapport à la fixation centrale (Bouma 1970). Ainsi, plus l'objet est excentré et plus il subira l'interférence d'éléments adjacents, même s'ils sont éloignés de lui. Ce phénomène impacte donc fortement l'apparence d'un stimulus en vision périphérique et il peut se manifester pour des objets (Kimchi & Pirkner, 2015), des lettres (Grainger, Tydgate, & Isselé 2010), des patches de Gabor (Levi et al., 2002) et même des visages. Par exemple, Louie et al. (2007) ont présenté à des participants deux visages en vision périphérique, à droite et à gauche d'un point de fixation central, dont l'un des deux visages était une cible à reconnaître (vu au préalable). Ces deux visages pouvaient être entourés de visages distracteurs à l'endroit ou retournés, ou d'aucun autres visages (Figure 6.4). Les participants devaient indiquer si le visage cible était situé à gauche ou à droite du point de fixation ou s'il n'était pas présent. Les résultats ont montré que les performances de reconnaissance étaient altérées lorsque le visage cible était en présence d'autres visages par rapport à lorsqu'il était présenté isolément. Cette interférence était plus grande pour les visages à l'endroit que pour les visages retournés. Ces résultats n'ont pas été retrouvés en présentation fovéale, ce qui suggère que le *crowding* est spécifique à la vision périphérique et peut se produire pour des représentations de haut niveau, comme la perception de visages.

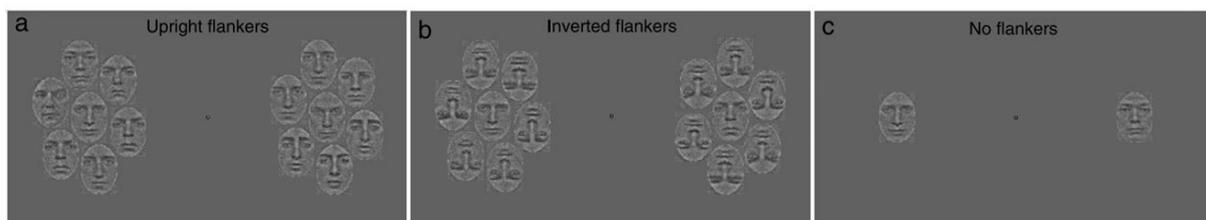


Figure 6.4 : Exemple de stimuli de visages utilisés pour tester le phénomène de *crowding* dans l'étude de Louie et al. (2007). Les participants devaient indiquer si le visage cible (vu au préalable) était situé à gauche ou à droite du point de fixation ou s'il n'était pas présent. **(a)** Les visages distracteurs sont à l'endroit, **(b)** les visages distracteurs sont à l'envers, et **(c)** les visages cibles sont présentés isolément. Figure adaptée de Louie et al. (2007).

Ainsi, la reconnaissance visuelle en vision périphérique reposerait sur l'interaction de plusieurs informations, comme la configuration spatiale des formes globales contenues dans le signal BFS et le *crowding*. Certains modèles visuels de reconnaissance visuelle ne considèrent d'ailleurs même pas les fréquences spatiales. Par exemple, dans le **Chapitre 1** de cette thèse, nous avons vu la Théorie de la hiérarchie inversée (*reverse hierarchy theory*) proposée par

Hochstein et Ahissar (2002). Dans cette théorie, les auteurs postulent que les connexions corticales sont bidirectionnelles et à la base de deux modes d'analyses visuels complémentaires *vision at a glance* et *vision with scrutiny*. Le premier mode *vision at a glance* correspondrait à la reconnaissance du *gist* de la scène, c'est-à-dire, aux traitements ascendants permettant d'aboutir à une première représentation de l'essentiel d'une scène visuelle mais sans précision. Le deuxième mode *vision with scrutiny* correspondrait au traitement de l'information détaillée. Bien que nous puissions aisément faire le rapprochement de ces deux modes de traitements avec celui des BFS et des HFS respectivement, les auteurs ont abordé la reconnaissance visuelle avec une approche plus générale. Dans ce contexte, il nous semble important dans le modèle **FP-Péri** de **considérer au même niveau l'ensemble des caractéristiques visuelles servant à la reconnaissance de l'essentiel de la scène, comme le signal en fréquences spatiales et l'arrangement des informations en vision périphérique (le crowding)**. Comme présenté dans la figure 6.1, nous avons décidé de dissocier dans l'extension du modèle **FP-Péri** les informations grossières/*coarse* (portée par les BFS) en vision périphérique et les informations détaillées/*fine* (portée par les HFS) en vision centrale.

Concernant notre deuxième question, peut-on considérer le modèle **FP-Péri** comme un modèle par défaut qui serait flexible dans certains cas ? Cette question a émergé suite aux résultats de l'expérience du **Chapitre 5** menée auprès de patients atteints d'un glaucome dans laquelle nous avons observé que les prédictions générées sur la base des BFS sont exacerbées en début de maladie et s'atténuent au stade avancé de la maladie. L'effet d'interférence très grand des BFS sur les HFS suggère en fait que les patients se basaient essentiellement sur la scène en BFS de l'image hybride pour répondre. Bien que les mécanismes prédictifs issus de la vision périphérique soient altérés chez ces patients, nous constatons que la vision centrale à elle seule, semble être en mesure de générer des prédictions sur la base d'une scène en BFS afin de guider la catégorisation d'une scène en HFS. Nous avons interprété ce résultat comme étant la conséquence de mécanismes compensatoires dans lesquels la perte de vision périphérique contraindrait les patients à très rapidement mobiliser les ressources disponibles dans la vision centrale encore intacte et à renforcer l'utilisation des BFS qui y sont disponibles pour générer des prédictions. Ce mécanisme compensatoire disparaîtrait alors progressivement avec l'avancée dans la maladie. En ce sens, nous pouvons postuler que le modèle **FP-Péri** correspondrait à un modèle "par défaut" qui serait flexible dans le cas où le système visuel serait contraint de s'adapter à la perte des fonctions de la vision périphérique. La vision centrale prendrait alors le relais pour assurer les fonctions initialement assurées par la vision périphérique. Nous pouvons d'ailleurs faire le rapprochement avec l'analyse *coarse-to-fine* qui correspond à une analyse par défaut, qui serait flexible dans le cas où la tâche nécessiterait d'allouer plus de poids au traitement des HFS (voir **Chapitre 1**).

Dans ce contexte, nous pouvons postuler que le système visuel suivrait par défaut une analyse *peripheral/coarse-to-central/fine*, dans laquelle les prédictions seraient déclenchées grâce à une analyse *coarse* en vision périphérique. Ces prédictions guideraient ensuite les mouvements oculaires afin d'amener la vision centrale sur des informations de la vision périphérique qui bénéficieraient alors d'une analyse fine *fine* (Figure 6.1). Dans le cas où le système visuel serait contraint de se réorganiser pour s'adapter aux changements (par exemple dans le glaucome), les prédictions seraient alors déclenchées grâce à une analyse *coarse* également possible en vision centrale, extraites grâce aux cellules ganglionnaires

magnocellulaires qui y sont disponibles en minorité. Cependant, nous pouvons nous interroger sur la capacité du système visuel à réorganiser les mécanismes de guidage des mouvements oculaires pour les patients atteints d'un glaucome. Smith, Crabb et Garway-Heath (2011) ont par exemple montré que les saccades de patients atteints d'un glaucome étaient moins nombreuses et plus courtes que celles des participants contrôles pour localiser un objet cible dans des scènes visuelles de la vie quotidienne. Cette capacité de réorganisation du système visuel est l'élément fondamental pour mettre en place de nouvelles rééducations dans des pathologies telle que le glaucome.

2.2. Perspectives cliniques

Bien que le système visuel soit en capacité de se réorganiser, les déficits engendrés par le glaucome ont des répercussions importantes dans la vie quotidienne des patients. Ces déficits se traduiraient par un coût cognitif plus important et épuisant pour les patients. En effet, les patients souffrant de glaucome signalent de nombreuses difficultés auxquelles ils doivent faire face dans leur vie quotidienne, même au début de la maladie. Nous avons vu par exemple dans le **Chapitre 5** un large éventail de difficultés exprimées par les patients, par exemple, pour marcher (Nelson et al., 1999 ; Odden et al., 2020 ; Viswanathan et al., 1999), pour lire (Lee et al., 1998), pour la conduite automobile (Bechetoille et al., 2008 ; Gutierrez et al., 1997), pour recherche visuelle d'objets (Smith, Crabb & Garway-Heath, 2011 ; Smith et al., 2012) ou de visages (Glen & Crabb, 2015). Une récente étude montre même que les patients âgés atteints d'un glaucome chutent plus de fois que des personnes âgées avec une vision normale (Jian-Yu et al., 2021). Ils signalent également sept fois plus de sensations de fatigue et de douleurs oculaires que des participants sains (Lee et al., 1998).

Un enjeu important de la recherche est donc de comprendre les mécanismes qui sous-tendent ces déficits afin de pouvoir proposer différentes méthodes de rééducation. Cet enjeu est d'autant plus important puisqu'il n'existe actuellement pas de traitement curatif de ces maladies. Plusieurs travaux ont comme point commun d'utiliser une méthode de rééducation visuelle basée sur l'apprentissage perceptif afin d'induire une plasticité spontanée dans les régions corticales répondant à la vision centrale et/ou périphérique. L'apprentissage perceptif a été utilisé de nombreuses fois dans certains troubles visuels (notamment dans l'amblyopie, la presbytie et la myopie) afin d'améliorer les performances à la suite d'un entraînement dans une tâche perceptive, mais rarement dans les cas de troubles complexes comme le glaucome ou la DMLA. Par exemple, Sabel et Gudlin. (2014) ont spécifiquement rééduqué la détection de contraste chez des patients atteints d'un glaucome. En effet, nous avons vu dans le **Chapitre 5** que les patients présentaient des déficits pour catégoriser, ou simplement détecter des images à faible niveau de contraste dans leur vision résiduelle (Lenoble, Lek & McKendrick, 2016 ; Roux-Sibilon et al., 2018). Dans cette étude, les patients devaient réaliser pendant trois mois à raison d'une heure par jour un entraînement visuel basé sur l'apprentissage perceptif. Cet entraînement consistait à demander aux patients de détecter des points de différentes intensités lumineuses présentés uniquement dans leur vision résiduelle. Les points lumineux étaient similaires à ceux présentés lors d'un examen périmétrique. Par rapport à un groupe placebo qui

devait déterminer l'orientation d'une barre présentée en vision centrale (sans manipuler la luminance), les auteurs ont observé que cet entraînement améliorerait la détection du contraste des patients. Autrement dit, le seuil de détection de luminance diminuait après l'entraînement. Cette étude suggère une plasticité du système visuel en vision résiduelle, même à un âge avancé. Selon les auteurs, la perte de vision dans le glaucome ne serait peut-être pas permanente mais partiellement réversible grâce à la réorganisation corticale. La vision centrale résiduelle des patients pourrait alors être surinvestie grâce à l'apprentissage perceptif.

Plus récemment, l'équipe de Benoît Cottureau du Centre de Recherche Cerveau et Cognition (CerCo) à Toulouse ont proposé une nouvelle méthode de rééducation à des patients atteints de DMLA (Maniglia et al., 2018) afin d'induire une réorganisation cérébrale susceptible d'améliorer les performances dans leur vision résiduelle (la vision périphérique). La perte de vision centrale dans cette maladie entraîne le développement d'un nouveau point de fixation excentrique qui remplace la fovéa, appelé le locus rétinien préféré ou PRL (*preferred retinal locus*). Dans leur étude, les patients réalisaient pendant deux mois un entraînement visuel basé sur la détection de contraste au niveau de leur PRL afin d'induire une plasticité spontanée dans les régions corticales répondant à la vision centrale. Par rapport aux participants contrôles, les résultats ont montré une réduction des seuils de contraste de luminance suite à l'apprentissage perceptif. En d'autres termes, plus les patients réalisaient l'entraînement, plus leur détection du contraste s'améliorait dans leur vision résiduelle. Cette étude suggère une nouvelle fois qu'un apprentissage perceptif peut induire une plasticité du système visuel.

Dans l'ensemble, ces études montrent qu'une partie des capacités visuelles peut être restaurée en activant davantage la vision résiduelle grâce à la plasticité cérébrale induite par un entraînement régulier. Ces études montrent également que le système visuel est capable de se réorganiser tout le long de la vie d'un individu, même à un âge avancé. En effet, le système visuel s'adapte continuellement grâce à ces fonctions intrinsèques mais il possède également la capacité de s'adapter en tenant compte des caractéristiques extrinsèques de l'environnement. Cette capacité d'adaptation est renforcée par des outils comme l'apprentissage perceptif. Concernant notre problématique de recherche, l'apprentissage perceptif pourrait être utilisé afin d'améliorer les mécanismes prédictifs des patients atteints d'un glaucome. Comme nous l'avons proposé dans le **Chapitre 5**, les patients au début de la maladie surutiliseraient les BFS qui sont disponibles en vision centrale pour générer des prédictions. Les patients seraient donc capables de mettre en place des stratégies de compensation en surinvestissant leur vision résiduelle (la vision centrale). Cependant, nous avons observé que cette stratégie disparaîtrait avec l'avancée dans la maladie. Ainsi, un apprentissage perceptif qui entraînerait les mécanismes prédictifs en vision centrale serait particulièrement pertinent pour ces patients.

Annexe 1



ELSEVIER

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

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

Audrey Trouilloud^{a,1}, Louise Kauffmann^{a,b,1}, Alexia Roux-Sibilon^a, Pauline Rossel^a,
Muriel Boucart^c, Martial Mermillod^a, Carole Peyrin^{a,*}

^a Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

^b Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

^c SCALab, University of Lille, Centre National de la Recherche Scientifique, Lille, France

ARTICLE INFO

Keywords:

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

ABSTRACT

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

1. Introduction

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

Kauffmann, Ramanoël, & Peyrin, 2014).

Interestingly, many studies have shown a crucial role of spatial frequencies for scene perception (Bar, 2003, 2007; Bar et al., 2006; Hegdé, 2008; Kauffmann et al., 2014; Schyns & Oliva, 1994). According to these studies, the visual analysis of the scene begins with the parallel extraction of different basic features at different spatial frequencies and follows a predominant coarse-to-fine sequence of processing. LSF containing the coarse information (e.g., the global shape and spatial layout of the scene) are rapidly and predominantly sent through the fast magnocellular pathways to the occipital cortex. LSF then access high-order cortical areas of the dorsal stream (such as parietal and frontal areas) and of the ventral stream (occipito-temporal areas), in order to activate the more probable interpretations of the visual input. This rapid coarse analysis is then used to guide the later processing of HSF, containing the fine information (e.g., the edges and the object detail)

* Corresponding author at: Laboratoire de Psychologie et NeuroCognition (LPNC), CNRS UMR 5105, Université Grenoble Alpes, BSHM – 1251 Av Centrale CS40700, 38058 Grenoble Cedex 9, France.

E-mail address: carole.peyrin@univ-grenoble-alpes.fr (C. Peyrin).

¹ These authors contributed equally.

<https://doi.org/10.1016/j.visres.2020.02.008>

Received 22 July 2019; Received in revised form 12 December 2019; Accepted 17 February 2020

0042-6989/ © 2020 Elsevier Ltd. All rights reserved.

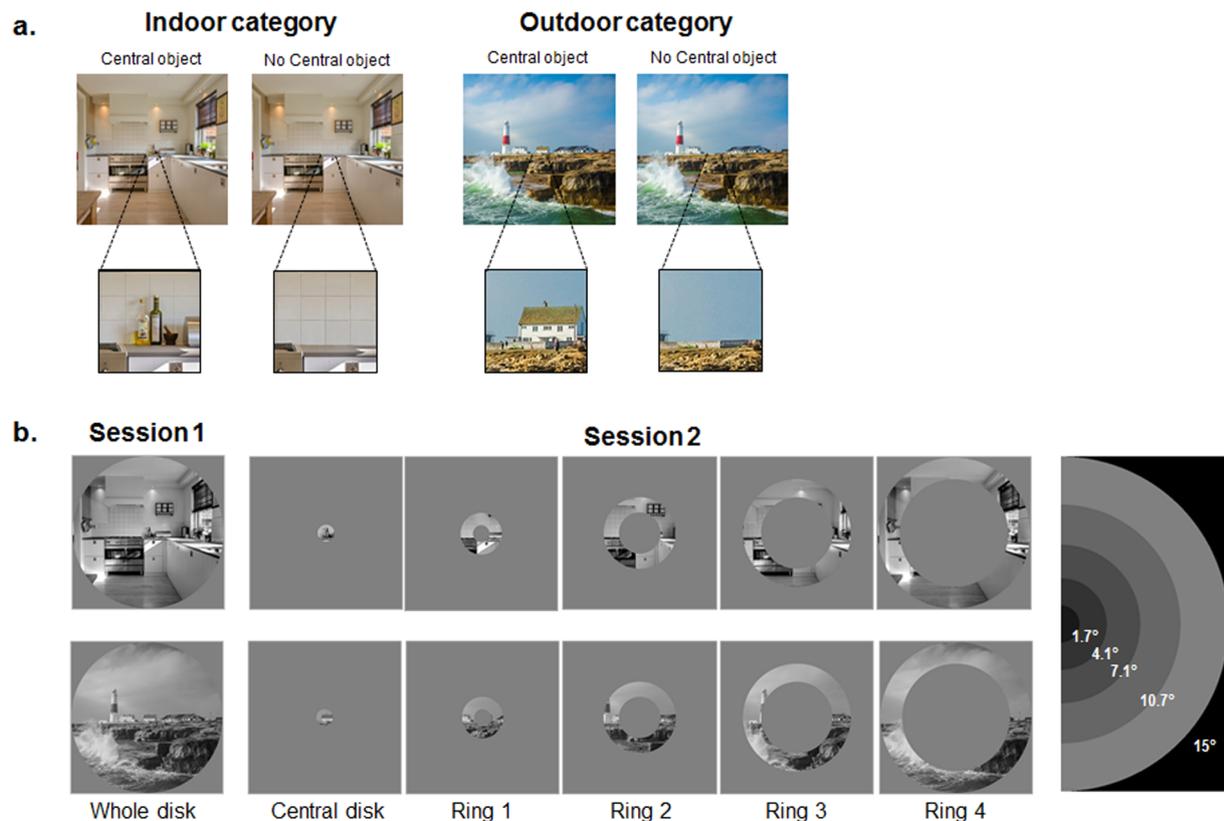


Fig. 1. (a) Example of images used in Experiment 1: an indoor and an outdoor scene of 1500×1500 pixels presented with or without a relevant object in central vision. (b) Examples of stimuli used in Session 1 (Whole disk) and Session 2 (Central disk, Ring 1, Ring 2, Ring 3, and Ring 4). The right image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

and conveyed more slowly by the parvocellular pathways to the ventral stream (Bar, 2003; Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Kveraga, Ghuman, & Bar, 2007; Peyrin et al., 2010; Trapp & Bar, 2015). Recent studies directly examined whether a coarse-to-fine sequence was advantageous for a rapid scene categorization (Kauffmann et al., 2015; Kauffmann et al., 2015). In these studies, the authors presented dynamic sequences composed of six filtered images of the same scene, assembled from LSF to HFS as in a coarse-to-fine (CtF) sequence, or from HSF to LSF as in a fine-to-coarse (FtC) sequence. Participants were faster to categorize CtF sequences than FtC sequences as indoor or outdoor scenes. This result confirms that the processing of LSF before HSF is advantageous for a rapid scene categorization.

Given that LSF signals would mainly come from the peripheral visual field, while HSF signals would mainly come from the central visual field, these past studies lead us to wonder about the relative contribution of peripheral vs. central processing for scene categorization. In agreement with a coarse-to-fine sequence of spatial frequency processing, does scene categorization also follow a peripheral-to-central processing? In support to this hypothesis, Boucart, Moroni, Thibaut, Szaffarczyk, and Greene (2013) have shown that scene categorization is still robust when scenes are presented at very large visual eccentricities (e.g., 70°). Importantly, psychophysical studies have highlighted the advantage of peripheral vision relative to central vision for scene categorization (Geuzebroek & van den Berg, 2018; Larson & Loschky, 2009; Loschky, Szaffarczyk, Beugnet, Young, & Boucart, 2019). Larson and Loschky (2009) originally used a “Window” and “Scotoma” paradigm. In the Window condition, participants viewed a scene through a circular window centered on the fovea. In the Scotoma condition, a circular area was hiding the central part of the scene and only the peripheral part was shown. In a control condition, the scene was presented entirely. Results showed that categorization performances were closer to the maximal performances (obtained in control condition) in

the Scotoma condition (peripheral vision) than in the Window condition (central vision). These results therefore suggested that the low resolution of peripheral vision is more useful than the high resolution of central vision to categorize a scene (see however, Larson, Freeman, Ringer, & Loschky, 2014). The importance of peripheral vision for natural scene perception has also been observed in pathologies. Roux-Sibilon et al. (2018) performed a study with open-angle glaucomatous patients in order to directly investigate the impact of a peripheral vision loss on scene categorization. This pathology is due to a progressive destruction of the ganglion cells and the optic nerve which mainly affect the peripheral retina. In this study, glaucomatous patients and healthy controls had to categorize scenes presented foveally. The results showed that patients had worse categorization performances than healthy participants, even though the stimuli appeared in their intact central visual field region. The integrity of peripheral retina therefore seems to be essential for scene categorization.

Overall, studies on spatial frequencies and central vs. peripheral visual field processing are consistent with the idea that the visual analysis of scenes follows a coarse/peripheral-to-fine/central sequence. The aim of the present study was to directly test this hypothesis using dynamic sequences to impose a Peripheral-to-Central (PtC) or Central-to-Peripheral (CtP) analysis of visual information. These sequences, inspired by the CtF and FtC sequences used by Kauffmann et al. (2015), were built from five versions of the same scene, revealing different parts of the visual field through circular rings at different eccentricities and assembled from the periphery to the center (PtC sequence) or from the center to the periphery (CtP sequence). In Experiment 1, we tested the categorization of each ring composing the dynamic scenes in the same way as the Larson and Loschky (2009) experiment. Participants had to categorize rings as indoor or outdoor scenes. We expected an advantage of peripheral on central vision on categorization performance. In this experiment, we also manipulated the content available in the central

part of the scene. Indeed, it is possible that the advantage of peripheral vision observed by Larson and Loschky (2009) during scene categorization resulted from an absence of relevant/diagnostic information in the central portion of the scene. Manipulating the central semantic content of the scene thus allowed us to determine whether the contribution of central vision depends on the presence of relevant semantic content to categorize scenes but also whether the presence of such content is enough to suppress the advantage of peripheral vision. In our experiment, the central part of the scenes thus either contained an object semantically related to the scene category (e.g., kitchen utensils for an indoor scene, or a house for an outdoor scene; Fig. 1) or no semantically relevant object (the object was removed from the picture; Fig. 1). Critically, Larson et al. (2014) revealed that the relative contribution of the central and peripheral visual field changes over time. Actually, the authors observed that during the first 100 ms of processing, there was a superiority of central vision, the relative contribution of peripheral vision increasing thereafter. They interpreted their results according to a zoom-out hypothesis of covert attention. Attention is first focused on central vision and then rapidly expands outward. Experiment 1 was also designed to explore the effect of temporal constraints on the processing of central versus peripheral visual information by manipulating the presentation time of stimuli. Stimuli were presented either 100 ms or 33 ms. Based on the hypothesis of a predominant coarse-to-fine processing during scene categorization (Schyns & Oliva, 1994), we expected that the coarse information available in peripheral vision would be even more considered for very short presentation times. Alternatively, according to the zoom-out hypothesis (Larson et al., 2014), information in central vision should be prioritized for short presentation times.

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

2. Experiment 1

2.1. Method

2.1.1. Participants

Thirty-two undergraduate students of Psychology from University Grenoble Alpes (28 women, mean age: 20.63 ± 2.68) participated in the experiment. They were divided into two groups ($N = 16$ each) for which stimuli were presented for either 33 or 100 ms. The sample size for each group was chosen based on a power analysis with estimated effect size of 1.061 (effect size of the interaction between the Eccentricity and the Central framing of the scene in Experiment 1) to achieve power of 0.99 at an alpha level of 0.05. All were right handed, with normal or corrected-to-normal vision. The study was approved by the ethics committee of the University Grenoble Alpes (CER-Grenoble Alpes, COMUE University Grenoble Alpes, IRB00010290). All participants involved in the study gave their informed oral and written consent.

2.1.2. Stimuli

The stimuli were constructed from 40 color photographs from the Pixabay website (<https://pixabay.com/fr/>), a photo sharing site under

CC0 (Creative Commons Zero) license. Half of the photographs represented indoor scenes (e.g., living room, kitchen, bathroom), and the other half represented outdoor scenes (e.g., cityscape, mountain, beach). In the context of rapid scene categorization, the choice of these categories was motivated by the fact that categorization at the superordinate-level has been found to precede basic-level distinctions during visual recognition (see for example Rousselet, Joubert, & Fabre-Thorpe, 2005; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Loschky & Larson, 2010; Kadar & Ben-Shahar, 2012).

For each scene, we constructed two square images of 1500×1500 pixels (Fig. 1a) centered on an object¹ fitting in a square of 100×100 pixels. For example, for indoor scenes, the object could be a household appliance, dishes or kitchen utensils, a book, a clock, a lamp, a radio, a candle, a vase or a pair of spectacles. For outdoor scenes, the object could be a car, a boat, a house, a building, a tree, a plant, a mountain or a bridge. Scenes without a central object were built by removing objects in the central area of the image (~ 200 pixels wide) using Adobe Photoshop CS6 (Fig. 1a). The tools used were Content Aware, Spot Healing Brush and Clone-Stamp. The Content Aware tool identifies similar details near the area selected and automatically replaces the removed object with details of the close environment. The Spot Healing Brush tool was used to complete the previous tool and clean more precisely the automatic correction defects. The Clone-Stamp tool is another technique used to remove an object. It allows to specifically select an area in a scene and to duplicate it on the object we want to remove. The size of the duplicated area and the area opacity can be defined in advance. In order to create a scene looking as natural as possible we have recreated shadows and light instead of the object with the high and low density tools. All images were converted to 256 Gy levels by averaging the values from the three color channels at each pixel. We also equalized the stimuli luminance and contrast. All stimuli were normalized to obtain an average luminance of 0.5 and a RMS contrast (Root Mean Square) of 0.2 for pixel intensity values between 0 and 1 (i.e. an average luminance of 128 and an RMS contrast of 51 on a scale of 256 Gy levels). Thus, we obtained 80 images (40 indoor scenes and 40 outdoor scenes) of size 1500×1500 pixels. Stimuli can be downloaded from <https://osf.io/ke6np/>.

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

pixels) and 10.70° of visual angle (535 pixels). Finally, for R4, adjacent to the third ring, the inner and outer edges were respectively fixed at 10.70° visual angle (535 pixels) and 15° visual angle (750 pixels). Thus, the Whole disk was the sum or superposition of all other stimuli. Stimuli were presented on a gray background of 0.5 average luminance.

2.1.3. Procedure

Stimuli were displayed using E-Prime software (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) on a 30" monitor (DELL ULTRA-SHARP), with a resolution of 2560×1600 pixels and a refreshing rate of 60 Hz. The viewing distance was set at 70 cm in order to respect the stimuli angular size of 30° visual angle. Participants were divided into two groups ($N = 16$ in each group) for which stimuli were presented for either 33 ms or 100 ms. In each group, all participants performed two sessions. In the first control session, only the Whole disk stimuli were displayed. This session allowed us to assess if participants were able to accurately categorize the scenes, but also to test the effect of the presence of a relevant object in central vision when categorizing a large scene. In the second session, participants were presented with parts of the scenes (CD, R1, R2, R3, and R4).

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

2.2. Results

Results are shown in Fig. 2. In each session, two repeated measures ANOVAs were performed on mean error rates (mER, in %) and mean correct response times (mRT, in ms). In order to reduce an effect of extreme RT values, RTs were trimmed for each participant's correct response of each condition. We removed responses less than and greater than three times the interquartile interval (0.15% of the trials for Session 1 and 0.49% of the trials for Session 2). ANOVAs were performed using Statistica 13.3 software (Statsoft, Tulsa, USA). Effect sizes were estimated by calculating the partial eta-squared (η^2). The alpha level of tests was set at 0.05.

In Session 1 (Whole disk), mER was very low (under 5%). The ANOVAs included the Exposure Duration of each stimulus (100 ms and 33 ms) as between-subject factor, the presence of a Central object (Object and No object) and the Category of scenes (Outdoor and Indoor) as within subject factors. There was no effect of the Central object, neither for mER (Mean \pm Standard deviation; Object: $4.53 \pm 0.96\%$; No object: $4.53 \pm 1.16\%$; $F(1,29) < 1$, $p = 1$), nor for mRT (Object: 615 ± 20 ms; No object: 622 ± 23 ms; $F(1,30) = 1.22$, $p = .277$). The main effect of the Exposure duration was not significant, neither for mER ($F(1,30) < 1$, $p = .903$), nor for mRT ($F(1,30) < 1$, $p = .486$)

and this factor did not interact with the Central object (mER: $F(1,30) < 1$, $p = .844$; mRT: $F(1,30) < 1$, $p = .862$). Furthermore, the Category of scenes did not interact with the presence of a Central object, neither for mER ($F(1,30) < 1$, $p = .461$), nor mRT ($F(1,30) = 2.34$, $p = .136$). These results suggest that the presence of a central object semantically related to the scene category does not improve categorization performances, whatever the scenes exposure duration, although this interpretation can be limited by a potential ceiling effect.

In Session 2, the ANOVAs included the Exposure Duration of each stimulus (33 ms and 100 ms) as between-subject factor, the Eccentricity of visual information (CD, R1, R2, R3, R4), the presence of a Central object (Object and No object) and the Category of scenes (Outdoor and Indoor) as within subject factors. The ANOVA performed on mER revealed a main effect of Eccentricity ($F(4,120) = 65.79$, $p < .001$, $\eta^2 = 0.69$). The linear and quadratic contrast tests performed between the different eccentricities were significant (Linear contrast: $F(1,30) = 84.09$, $p < .001$; Quadratic contrast: $F(1,30) = 114.24$, $p < .001$; CD: $15.74 \pm 2.13\%$; R1: $5.51 \pm 1.32\%$; R2: $5.00 \pm 1.47\%$; R3: $5.00 \pm 1.044\%$; R4: $6.42 \pm 1.89\%$). We subsequently performed planned comparisons between each stimulus in order to identify which ring leads to the best performance. These analysis showed a significant difference between CD and R1 ($F(1,30) = 146.13$, $p < .001$), but no significant difference between R1 and R2 ($F(1,30) = 0.51$, $p = .479$), neither for R2 and R3 ($F(1,30) < 1$, $p = .997$) or R3 and R4 ($F(1,30) = 3.268$, $p = .081$). These results suggest that errors exponentially decreased as eccentricity increased. Importantly, the main effect of the Central object was significant ($F(1,30) = 19.45$, $p < .001$, $\eta^2 = 0.39$; Object: $6.49 \pm 2.20\%$; No object: $8.59 \pm 2.27\%$) and the presence of the Central object of the scene interacted with Eccentricity ($F(4,120) = 25.04$, $p < .001$, $\eta^2 = 0.45$). The linear and quadratic contrast tests performed between the different eccentricities were significant both when there was a central object (Linear contrast: $F(1,30) = 17.60$, $p < .001$; Quadratic contrast: $F(1,30) = 16.16$, $p < .001$; CD: $10.39 \pm 1.35\%$; R1: $5.78 \pm 1.08\%$; R2: $5.39 \pm 1.19\%$; R3: $4.61 \pm 1.12\%$; R4: $6.62 \pm 1.65\%$) and when there was no central object (Linear contrast: $F(1,30) = 79.68$, $p < .001$; Quadratic contrast: $F(1,30) = 124.77$, $p < .001$; CD: $21.09 \pm 2.19\%$; R1: $5.23 \pm 1.05\%$; R2: $4.61 \pm 1.10\%$; R3: $5.39 \pm 1.23\%$; R4: $6.60 \pm 1.41\%$). We then performed planned comparisons between each stimulus in order to identify which ring leads to the best performance. When there was a central object, these analyses showed a significant difference between CD and R1 ($F(1,30) = 23.28$, $p < .001$), but no significant difference between R1 and R2 ($F(1,30) < 1$), neither for R2 and R3 ($F(1,30) = 1.30$, $p = .264$) or R3 and R4 ($F(1,30) = 2.44$, $p = .129$). When there was no central object, these analyses showed a significant difference between CD and R1 ($F(1,30) = 163.05$, $p < .001$), but no significant difference between R1 and R2 ($F(1,30) < 1$), neither for R2 and R3 ($F(1,30) < 1$) or R3 and R4 ($F(1,30) = 1.21$, $p = .279$). Thus, errors exponentially decreased as eccentricity increased irrespective of the presence of a central object in the scene. Critically, for CD, participants made less errors for categorizing a scene that contained an object in central vision ($10.39 \pm 1.35\%$; $F(1,30) = 37.32$, $p < .001$), compared to a scene that did not contain one ($21.09 \pm 2.19\%$). Yet, despite the presence of a central object that significantly improved the categorization performances, participants still made fewer errors when scene information was available in peripheral vision than in central vision. The main effect of the Exposure duration was not significant ($F(1,30) = 1.33$, $p = .258$) and this factor did not interact neither with the Eccentricity ($F(4,120) = 1.42$, $p = .231$), nor with the Eccentricity*Central object interaction ($F(4,120) < 1$, $p = .542$). Finally, the main effect of the Category was significant ($F(1,30) = 18.85$, $p < .001$, $\eta^2 = 0.39$). Participants made less errors when categorizing outdoor scenes ($5.93 \pm 1.20\%$) than indoor scenes

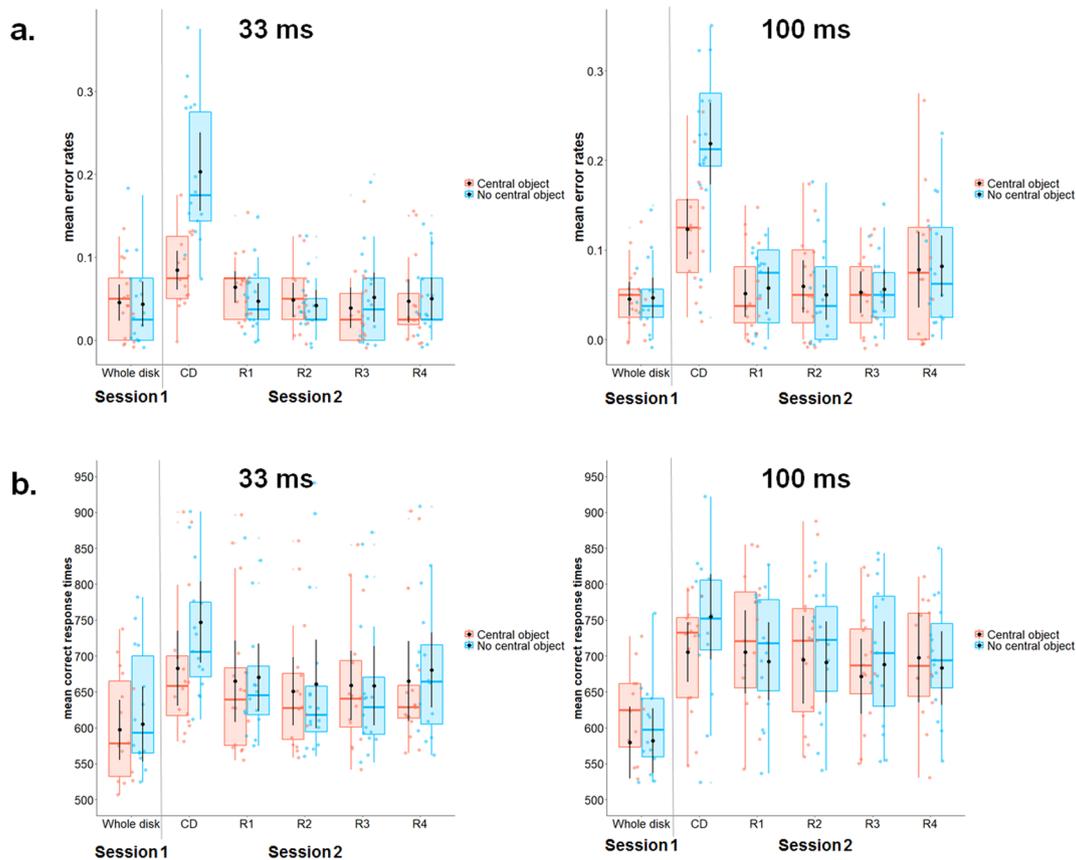


Fig. 2. (a) Mean error rates in percentage and (b) mean correct response times in milliseconds in Experiment 1 during the categorization of the Whole Disk in Session 1 and each ring in Session 2 (CD, R1, R2, R3, R4), according to the exposure duration (33 ms and 100 ms) and the presence of a central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

($9.14 \pm 2.79\%$). But the category did not interact with the others manipulated factors (Eccentricity \times Category: $F(4,120) < 1$; Eccentricity \times Exposure duration \times Category: $F(4,120) < 1$; Eccentricity \times Central object \times Category: $F(4,120) = 1.27$; Eccentricity \times Central object \times Exposure duration \times Category $F(1,24) < 1$).

We also compared performances of each eccentricities of Session 2 (CD, R1, R2, R3, R4) to performances obtained in Session 1 (Whole disk). These comparisons showed that categorization performance of R4 and CD was significantly different to the categorization of the Whole disk (R4: $t(32) = -2.41, p = .022$; CD: $t(32) = -12.60, p < .001$). Performances on other eccentricities (R3, R2 and R1) were not significantly different ($ts(32) < -1.62, ps > 0.114$). These results suggest that scene categorization is disadvantaged when only the central visual information or the most peripheral visual information is available.

Concerning now the ANOVA performed on mRT, it revealed a main effect of Eccentricity ($F(4,120) = 29.32, p < .001, \eta p^2 = 0.49$). Similarly to the mER analysis, a linear and quadratic contrast test performed between the different eccentricities revealed both a linear and quadratic decrease of mRT as eccentricity increased (Linear contrast: $F(1,30) = 61.47, p < .001$; Quadratic contrast: $F(1,30) = 37.41, p < .001$; CD: 718 ± 32 ms; R1: 679 ± 33 ms; R2: 672 ± 35 ms; R3: 665 ± 33 ms; R4: 679 ± 34 ms). We performed planned comparisons between each stimulus in order to identify which ring leads to the best performance. These analysis showed a significant difference between CD and R1 ($F(1,30) = 55.86, p < .001$) and R3 and R4 ($F(1,30) = 5.13, p = .031$), but no significant difference between R1 and R2 ($F(1,30) = 2.03, p = .164$) or R2 and R3 ($F(1,30) = 2.17, p = .151$). These results suggest that mRT follows a U-shape curve. As for the mER analysis, the main effect of the Central object was

significant ($F(1,30) = 12.71, p = .001, \eta p^2 = 0.30$). Participants categorized a scene that contained an object in central vision faster (677 ± 52 ms) than a scene that did not contain one (688 ± 52 ms). The presence of a Central object interacted with Eccentricity ($F(4,120) = 10.32, p < .001, \eta p^2 = 0.26$). The linearity and quadratic contrast tests performed between the different eccentricities were significant both when there was a central object (Linear contrast: $F(1,30) = 13.71, p < .001$; Quadratic contrast: $F(1,30) = 11.05, p = .002$; CD: 694 ± 21 ms; R1: 682 ± 26 ms; R2: 671 ± 25 ms; R3: 661 ± 22 ms; R4: 679 ± 26 ms) and when there was no central object (Linear contrast: $F(1,30) = 71.37, p < .001$; Quadratic contrast: $F(1,30) = 43.01, p < .001$; CD: 742 ± 25 ms; R1: 675 ± 22 ms; R2: 672 ± 26 ms; R3: 669 ± 25 ms; R4: 680 ± 23 ms). We then performed planned comparisons between each stimulus in order to identify which ring leads to the best performance. When there was a central object, these analysis showed a significant difference between R3 and R4 ($F(1,30) = 4.32, p = .05$), but no significant difference between CD and R1 ($F(1,30) = 2.42, p = .130$), neither for R1 and R2 ($F(1,30) = 2.10, p = .157$) or R2 and R3 ($F(1,30) = 2.28, p = .141$). When there was no central object, these analysis showed a significant difference between CD and R1 ($F(1,30) = 60.80, p < .001$), but no significant difference between R1 and R2 ($F(1,30) < 1$), neither for R2 and R3 ($F(1,30) < 1$) or R3 and R4 ($F(1,30) = 1.87, p = .181$). Critically, for CD, planned comparisons showed that participants categorized the scenes that contained an object in central vision (694 ± 21 ms; $F(1,30) = 37.32, p < .001$) faster than the scenes that did not contain one (742 ± 25 ms). The main effect of the Exposure duration was not significant ($F(1,30) = 0.62, p = .438$) and this factor did not interact neither with the Eccentricity ($F(4,120) = 1.37, p = .248$), nor with the Eccentricity*Central object interaction (F

(4,120) = 2.16, $p = .078$).

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

As for the mER analysis, compared performances of each eccentricities of Session 2 (CD, R1, R2, R3, R4) to performances obtained in Session 1 (Whole disk). These comparisons showed that categorization performance of all eccentricities (R4, R3, R2, R1, CD) were significantly different from the categorization of the Whole disk both when there was no object and when there was one ($t(32) > -4.17, ps < 0.001$). Thus, a partial presentation of a scene systematically lengthened response times.

2.3. Discussion

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

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

the results of Larson and Loschky (2009) by demonstrating superiority of peripheral vision over central vision regardless of their respective informational content.

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

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

3. Experiment 2

In this experiment, we created dynamic sequences composed of five stimuli assembled from the periphery to the center (PtC) or from the center to the periphery (CtP), so that the whole scene was revealed successively through the five stimuli. As in Experiment 1, stimuli activated approximately the same cortical surface on V1 and we manipulated the content available in the central part of the scene. The sequence either contained - in the Central Disk - an object semantically related to the scene category (e.g., kitchen utensils for an indoor scene, or a house for an outdoor scene), or no semantically relevant object (the object was removed). We also manipulated the Exposure duration of the sequence. Sequences lasted either 165 ms (each ring within the sequence

was presented for 33 ms as in the shortest exposure duration condition of Experiment 1) or 500 ms (each ring was presented for 100 ms as in the longest exposure duration condition of Experiment 1). We expected that PtC sequences would be categorized more rapidly and more accurately than CtP sequences in both exposure duration condition, and this, despite the presence of an object in central vision. In addition, based on the hypothesis of the rapid processing of LSF during scene categorization, we expected that the coarse information available in peripheral vision would be more considered when sequences lasted 165 ms, whereas the slower analysis of details in central vision would take over when sequences lasted 500 ms.

3.1. Method

3.1.1. Participants

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

3.1.2. Stimuli and procedure

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

Stimuli were displayed using E-Prime software (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) using the same 30" monitor as in Experiment 1 with a viewing distance of 70 cm. All participants underwent one session and they were divided into two groups ($N = 16$ in each group) for which CtP and PtC sequences were presented for either 165 ms or 500 ms. The CtP and PtC sequences were randomly presented to the participants. A trial began with a central black fixation point presented for 500 ms on a 0.5 luminance background, followed by a

sequence of five stimuli. Inter-sequences interval was a grey screen of 1900 ms during which participants could respond. Participants' task was similar to Experiment 1 (i.e. indoor vs. outdoor categorization). The session included 160 trials (i.e. 20 stimuli \times 2 Central framing conditions \times 2 Categories \times 2 Sequence orders). The experiment lasted about 7 min for the 165 ms Exposure duration condition, and about 8 min for the 500 ms Exposure duration condition, including breaks every 100 trials. For each trial, response accuracy and response time (in ms) were recorded. Before each experimental session, participants performed a training session (20 trials) using other stimuli.

3.2. Results

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

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

The ANOVA performed on mRT revealed a main effect of the Sequence ($F(1,30) = 65.88, p < .001, \eta^2 = 0.69$). Participants were faster for categorizing PtC sequences (613 ± 29 ms) than CtP sequences (654 ± 30 ms). The main effect of the Exposure duration was not significant ($F(1,30) < 1$), but this factor interacted with the Sequence ($F(1,30) = 14.52, p < .001$). Planned comparisons revealed that participants categorized PtC sequences faster than CtP sequences both when they were presented for 165 ms (PtC: 634 ± 41 ms; CtP:

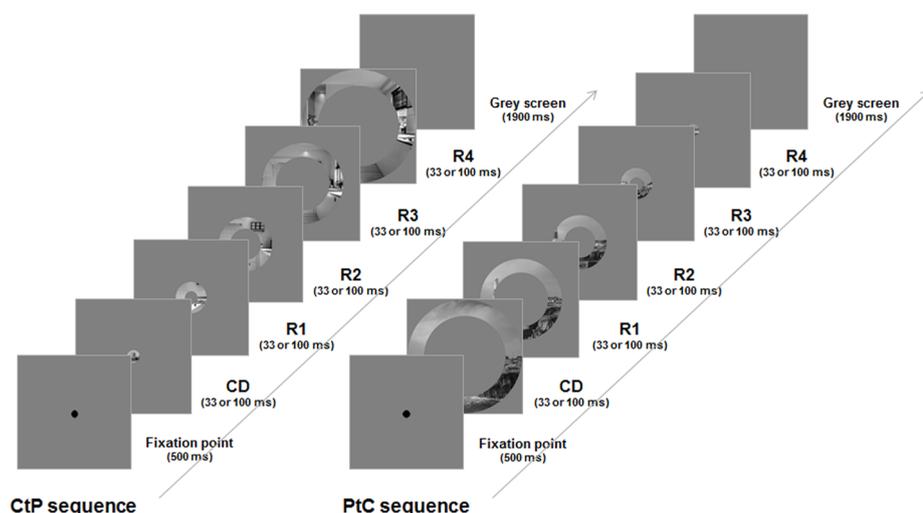


Fig. 3. Trial schematic in Experiment 3. Central disk (CD), Ring 1 (R1), Ring 2 (R2), Ring 3 (R3) and Ring 4 (R4) are assembled from the center to the periphery (CtP sequences) or from the periphery to the center (PtC sequences). The sequences lasted either 160 ms (each stimulus was presented for 33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

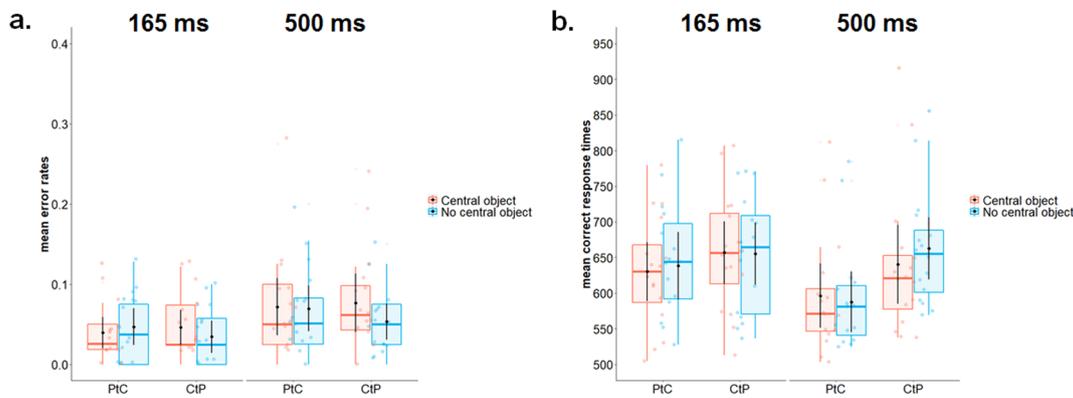


Fig. 4. (a) Mean error rates in percentage and (b) mean correct response times in milliseconds during the categorization of the PtC and CtP sequences in each exposure duration condition (165 ms, 500 ms), according to the presence of a central object or not. Black dots and error bars indicate mean and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

656 ± 43 ms; $F(1,30) = 9.27, p < .005$) and for 500 ms (PtC: 592 ± 41 ms; CtP: 652 ± 43 ms; $F(1,30) = 71.12, p < .001$), the difference between the two sequences being greater for exposure duration of 500 ms than 165 ms. The main effect of the Central object was not significant ($F(1,30) = 2.03, p < .164$), and this factor did not interact with the Sequence ($F(1,30) = 2.08, p = .162$). However, the analysis revealed a significant interaction between the Sequence, the Exposure duration and the Central object ($F(1,30) = 7.75, p = .009, \eta^2 = 0.20$). We thus tested the interaction between the Sequence and the Exposure duration for each condition of Central object separately. This interaction was significant when there was no central object ($F(1,30) = 19.44, p < .001$), but not when there was a central object ($F(1,30) = 2.13, p = .154$). More precisely, when there was a central object, participants categorized PtC sequences faster than CtP sequences irrespective of the exposure duration of sequences (for 500 ms: PtC: 597 ± 29 ms, CtP: 640 ± 33 ms; for 165 ms: PtC: 630 ± 29 ms, CtP: 657 ± 33 ms). However, when there was no central object participants categorized faster a PtC than a CtP sequence only when sequences lasted 500 ms (for 500 ms: PtC: 588 ± 30 ms, CtP: 663 ± 29 ms, $F(1,30) = 63.37, p < .001$; for 165 ms: PtC: 638 ± 30 ms, CtP: 655 ± 29 ms, $F(1,30) = 2.98, p = .095$). These results suggest that the PtC advantage was reduced when there was no central object in the sequences lasting 165 ms.

Furthermore, there was no main effect of Category ($F(1,30) < 1$), but the interaction between the Category and the Sequence was significant ($F(1,30) = 4.85, p = .035, \eta^2 = 0.14$). Planned comparisons showed that the difference between CtP and PtC sequences was significant both for Indoor category (CtP: 657 ± 23 ms; PtC: 610 ± 21 ms; $F(1,30) = 53.52, p < .001$) and Outdoor category (CtP: 651 ± 21 ms; PtC: 617 ± 21 ms; $F(1,30) = 46.84, p < .001$), the PtC advantage being greater for the outdoor scenes. Finally, the Category did not interact with the others manipulated factors (Exposure duration × Category: $F(1,30) < 1$; Sequence × Exposure duration × Category: $F(1,30) < 1$; Sequence × Central framing × Category: $F(1,30) < 1$; Sequence × Central framing × Exposure duration × Category: $F(1,30) = 1.77, p = .193$).

3.3. Discussion

Experiment 2 was specifically designed to test the advantage of a rapid analysis of coarse information available in peripheral vision, which could allow a first categorization, prior to its validation through the slower detailed analysis in central vision. For this, we used dynamic scene sequences composed of the central disk and the four rings assembled from the central disk to the more peripheral ring to impose a Central-to-Peripheral analysis (CtP sequences) or assembled from the more peripheral ring to the central disk to impose a Peripheral-to-Central analysis of visual information (PtC sequences). In this

experiment, we also tested whether the superiority of peripheral vision was enhanced by a short exposure duration. Sequences were presented for either 165 ms or 500 ms, so that each stimulus that composed the sequence was presented for either 33 ms or 100 ms, respectively.

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

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

In this control experiment, we manipulated the eccentricity of the stimuli to categorize and the presence of a mask sequence following the stimulus. In a non masking condition, participants had to categorize the central disk (CD) and the more peripheral ring (R4) of a scene. In a masking condition, they had to categorize a sequence composed of five stimuli assembled from the center to the periphery in which the CD

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

4. Experiment 3

4.1. Method

4.1.1. Participants

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

4.1.2. Stimuli and procedure

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

Stimuli were displayed using on E-Prime software (E-Prime Psychology Software Tools Inc., Pittsburgh, USA) using the same 30" monitor as in Experiment 1 with a viewing distance of 70 cm. All participants performed one session and they were divided into two groups ($N = 16$ in each group) depending on the presentation of stimuli. Central stimuli, Peripheral stimuli, Central-to-Noise sequences and Peripheral-to-Noise sequences were randomly presented to the participants. A trial began with a central black fixation point presented for 500 ms on a 0.5 luminance background, followed an individual stimulus or a sequence. Inter-stimuli interval was a grey screen of 1900 ms during which participants could respond. Participants' task

was similar to Experiment 1 (i.e. indoor vs. outdoor categorization). The session included 320 trials (i.e. 20 stimuli \times 2 Eccentricities \times 2 Masking \times 2 Categories). The experiment lasted about 14 min for the short Exposure duration condition, and about 15 min for the long Exposure duration condition, including breaks every 80 trials. For each trial, response accuracy and response time (in ms) were recorded. Before each experimental session, participants underwent a training session (20 trials) using other stimuli.

4.1.3. Results

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

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

The ANOVA performed on mER showed no significant main effect of the Eccentricity ($F(1,29) < 1$; R4: $8.55 \pm 1.89\%$; CD: $9.01 \pm 2.18\%$), but the main effect of the Mask was significant ($F(1,29) = 12.67$, $p = .001$, $\eta^2 = 0.30$). Participants made less errors for categorizing a stimulus without a mask following it ($7.22 \pm 1.70\%$) than a stimulus with a mask ($10.34 \pm 2.33\%$). This result suggests a visual backward masking effect. However, the Mask did not interact with the Eccentricity ($F(1,29) < 1$), suggesting that the visual backward masking was not greater for a peripheral stimulus surrounding a central stimulus than the reverse. The Exposure duration

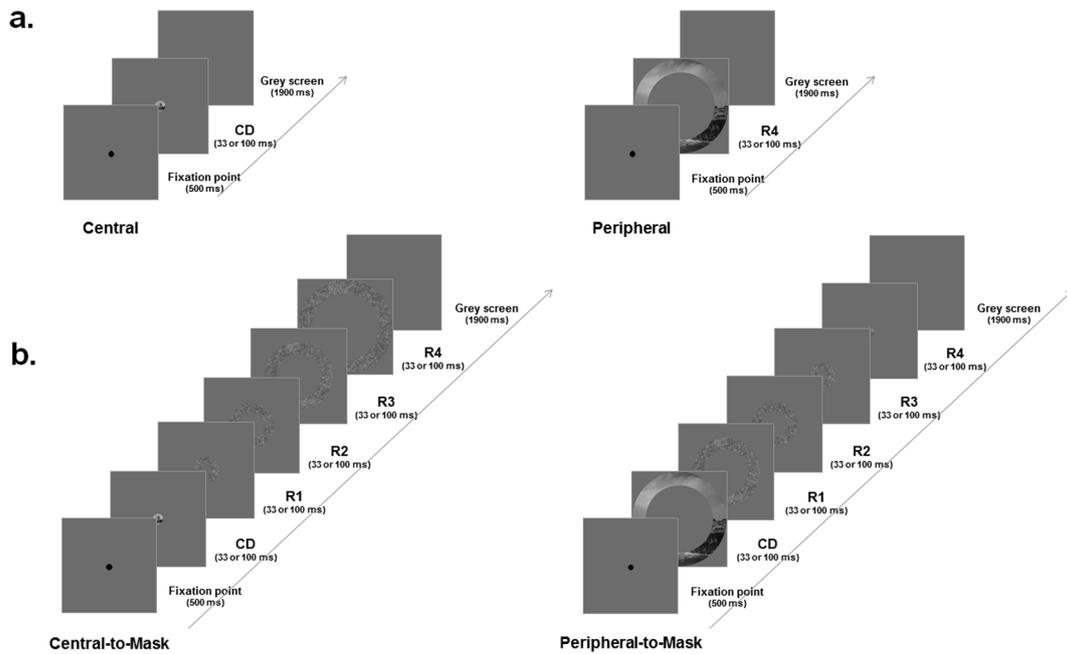


Fig. 5. Trial schematic of Experiment 3. (a) Participant had to categorize a Central disk (CD) or a Ring 4 (R4) of a scene presented either for 33 or 100 ms. (b) Participant had to categorize a Central-to-Mask sequence composed of five stimuli assembled from the center to the periphery in which the CD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

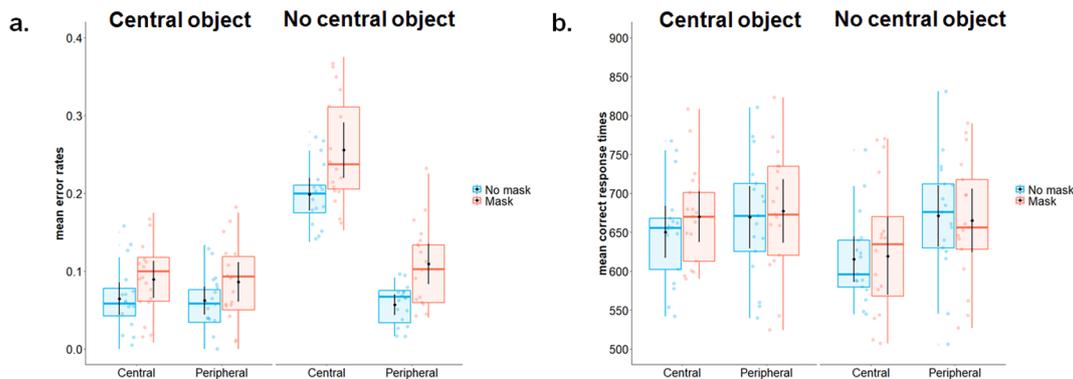


Fig. 6. (a) Mean error rates in percentage and (b) mean correct response times in milliseconds during the categorization of the central (Central disk) and peripheral (Ring 4) stimuli followed or not by a sequence of masks, according to the presence of a central object or not. Black dots and error bars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

did not interact with the Eccentricity and the Mask ($F(1,29) < 1$). Finally, there was neither a main effect of Category ($F(1,29) < 1$), nor interaction between the Category, the Eccentricity and the Mask ($F(1,29) < 1$).

The ANOVA performed on mRT showed no significant main effect of the Eccentricity ($F(1,29) = 1.95, p = .173$; R4: 675 ± 22 ms; CD: 662 ± 17 ms), but the main effect of the Mask was significant ($F(1,29) = 45.68, p < .001, \eta^2 = 0.61$). Participants were faster for categorizing a stimulus without a mask following it (702 ± 32 ms) than a stimulus with a mask (744 ± 36 ms). Again, this result suggests a backward masking effect. As the mER analysis, the Mask did not interact with the Eccentricity ($F(1,29) = 2.73, p = .109$), suggesting that the visual backward masking was not greater for a peripheral stimulus surrounding a central stimulus than the reverse. The Exposure duration did not interact with the Eccentricity and the Mask ($F(1,29) < 1$). Furthermore, the main effect of Category was significant ($F(1,29) = 4.30, p = .047, \eta^2 = 0.13$; Indoor: 730 ± 41 ms; Outdoor:

717 ± 33 ms), and this factor interacted with Eccentricity ($F(1,29) = 5.88, p = .022$). Planned comparisons showed that participants categorized faster the CD than the R4 for indoor scenes (R4: 743 ± 31 ms; CD: 717 ± 21 ms; $F(1,29) = 4.83, p = .036$) but not for outdoor scenes ($F(1,29) < 1$). Finally, the Category did not interact with the Eccentricity and the Mask ($F(1,29) = 2.47, p = .127$).

5. General discussion

The present study investigated the advantage of the peripheral vision on the central vision during rapid scene categorization. Given that spatial resolution decreases with retinal eccentricity, peripheral visual information would be mainly encoded in LSF while central visual information would be mainly encoded in HSF. Based on the hypothesis of a predominant coarse-to-fine processing of spatial frequencies, we wondered if scene analysis follows a coarse/peripheral-to-fine/central processing. Three experiments were conducted to address these issues.

We used large scene from which we built one central disk and four circular rings of different eccentricities. In Experiment 1, participants had to categorize individual stimuli as indoor or outdoor scenes. In Experiment 2, they had to categorized dynamic sequences composed of the central disk and the four rings assembled from peripheral to central vision (PtC sequences) or assembled from central to peripheral vision (CtP sequences). Experiment 3 was designed to assess if our results could be biased by a metacontrast masking effect. In all experiments, we manipulated the semantic content available in the central part of the scene. The Central Disk either contained an object semantically related to the scene category (e.g., kitchen utensils for an indoor scene, or a house for an outdoor scene), or no semantically relevant object. These experiments yielded two key findings: (1) categorization performances improved as scene information was presented in peripheral vision (Experiment 1), (2) better performances were observed during categorization for PtC than CtP sequences when a relevant object was present in the Central disk (Experiment 2). However, the presence of a central object semantically related to the category of the scene significantly improved the categorization performances of the central disk, as well as of the CtP sequences presented for 500 ms.

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

It should be noted that we constructed our stimuli so that the surface of visual information revealed by each ring took into account the cortical magnification factor. Indeed, retinotopic projections to the primary visual cortex (V1) are deformed. A large number of V1 cells are dedicated to processing visual information from the central retina while a smaller number of areas are dedicated to more peripheral vision (Engel et al., 1994; Engel, Glover, & Wandell, 1997; Daniel & Whitteridge, 1961; Virsu, Näsänen, & Osmoviita, 1987; Virsu & Rovamo, 1979). Therefore, visual information from the central visual field is over-represented at the cortical level. The foveal portion of the retina, which represents only 1% of the total surface of the retina, is represented on 50% of V1. This cortical magnification contributes to explain the decrease in visual acuity with retinal eccentricity (Duncan & Boynton, 2003). In order to account for this cortical magnification, the width and surface of rings in our experiments increased with eccentricity and the most peripheral ring (Ring 4) contained more quantity of visual information potentially more useful for categorization than the central ring. Importantly, Geuzebroek and van den Berg (2018) showed that peripheral vision would have an advantage over central vision in

scene categorization precisely because it covers a larger surface of visual field allowing to reveal more scene content. In their experiment, authors presented scenes to categorize using a Window-Scotoma paradigm. Scenes were presented according five conditions. In a central viewing condition C1, a 5° radius window was placed on the scene and participants saw only the scene center. In a central viewing condition C2, the whole scene was reduced in a 5° radius window. Thus, stimuli C1 and C2 activated the same surface of V1, but stimulus C2 contained more visual information than stimulus C1. In a peripheral viewing condition P1, the scene was presented within a ring with inner and outer edges at 5° and 40° respectively. In a peripheral viewing condition P2, the scene was presented within a ring with inner and outer edges at 12° and 40° respectively. Finally, in a CO1 condition, stimulus P2 was reduced taking into account the cortical magnification factor, so that the scene was presented with the inner and outer edges at 3.6° and 12°. Thus, P2 and CO1 stimuli activated the same V1 surface and contained the same quantity of visual information. The results showed that performances were lower in the conditions where only central information was available (C1 and C2) compared to the conditions where only peripheral information was available (P1 and P2), thus confirming the previous results of Larson and Loschky (2009), as well the results of our Experiment 1. More interestingly, performances were worse in the condition where only the central information was available (C1) compared to the condition where all scene information was presented in central vision (C2). However, for conditions P2 and CO1, performances were similar despite the eccentricity change in scene presentation. These results suggests that scene categorization processes are independent of scene eccentricity, provided that central vision and peripheral vision both activate the same number of cells in V1 and that the same quantity of information is available in central and peripheral vision, although this is never the case under natural conditions. Overall, these results indicate that retinal eccentricity plays a minimal role in scene perception and peripheral vision has an advantage compared to central vision only when it covers a larger visual field, allowing presenting a larger scene content. In the present study, we chose to respect the natural functioning properties of the visual system by manipulating the width of stimuli according to the cortical magnification factor. Furthermore, we believe it would have been irrelevant to present the same width of visible scene at each eccentricity. For example, we could have applied the central disk width (radius of 1° of visual angle) to the peripheral rings. But this experimental control would result in exceeding the visual system capabilities in peripheral vision, which is tuned to extract LSF components, and may have strongly hindered categorization. In contrast, if we had applied the width of Ring 4 (7.74° of visual angle) to the Central disk, it would have included a portion of peripheral visual field. Therefore, controlling the width of rings for each eccentricity would not respect natural vision properties and may bias the strategies used by the visual system to categorize scenes.

Experiment 2 was specifically designed to test the advantage of a rapid analysis of coarse information available in peripheral vision, which could allow a first categorization, prior to its validation through the slower detailed analysis in central vision. For this, we used dynamic scene sequences composed of the central disk and the four rings assembled from the central disk to the more peripheral ring to impose a Central-to-Peripheral analysis (CtP sequences) or assembled from the more peripheral ring to the central disk to impose a Peripheral-to-Central analysis of visual information (PtC sequences). Sequences were presented for either 165 ms or 500 ms, so that each stimulus that composed the sequence was presented for either 33 ms or 100 ms, respectively. Participants categorized PtC sequences faster than CtP sequences, but the exposure duration of sequences significantly influenced this difference. The PtC advantage was greater for the longest than shortest exposure duration of the sequences. Importantly, this interaction was actually due to an unexpected effect when no object was present in central vision for the shortest exposure duration condition. Indeed, we observed a significant interaction between the type of

sequences, their exposure duration and the presence of a central object. When the central part of the scenes contained an object semantically related to the scene category, participants categorized PtC sequences faster than CtP sequences irrespective of the exposure duration of sequences. Surprisingly, when there was no object, a PtC advantage was only observed for the longest exposure duration. In comparison to the condition in which an object was present, reaction times increased for categorizing the PtC sequences, as if the absence of a relevant information in central vision following the peripheral processing was detrimental to rapid scene categorization (sequences lasted 165 ms). This result could be interpreted in the context of a predominant coarse-to-fine processing during scene categorization (Kauffmann et al., 2014; Schyns & Oliva, 1994). According to this hypothesis, a rapid extraction of predominantly coarse information in the peripheral visual field should provide the global shape and structure of the scene used for a rapid initial perceptual categorization. This perceptual categorization should be refined, confirmed, or infirmed by the processing of finer information available in the central visual field. In the present experiment, this periphery-based rapid categorization would be then confirmed by the presence of a relevant object in the central part of the scene. When no relevant information was present in central vision, the visual system would need more visual information, delaying the categorization time. The stronger PtC advantage for the longest sequences irrespective of the presence of a central object suggests that the fine information from central vision would not even be used. It is possible that for sequences of longer duration (500 ms), the categorization could have been only based on the first ring, presented for 100 ms. Results of Experiment 1, as well as previous studies on scene recognition (Johnson & Olshausen, 2003; Thorpe, Fize, & Marlot, 1996; Vanrullen & Thorpe, 2001), confirms the ability of the visual system to categorize a scene presented for just 100 ms.

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

Another alternative account for a stronger PtC advantage for longer exposure duration may be linked to eye-movements. Indeed, for the longest sequences, it is likely that participants fixated each stimulus sequentially. Unfortunately, we could not record eye movements during this experiment in order to control that participants maintained their gaze on the central fixation throughout the sequence presentation. Therefore, it cannot be totally excluded that participants made eye movement when sequence lasted 500 ms. We however believe it is unlikely that such behavior occurred and explains the PtC advantage. Given that participants did not know in advance the eccentricity of the first stimulus (either the central disk or Ring 4), it would have been

inefficient without impairing scene categorization. Participants could also initiate a saccade from one stimulus to the next stimulus in the sequence, but this strategy would have been demanding. Indeed, saccadic eye movements are usually initiated within 100–150 ms (Fisher & Weber, 1993). If participants made a saccade during the presentation of one stimulus (presented during 100 ms), this puts severe constraints on the time left to process visual information in the next stimulus.

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

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

6. Footnote

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

CRedit authorship contribution statement

Audrey Trouilloud: Methodology, Formal analysis, Investigation, Writing - original draft. **Louise Kauffmann:** Conceptualization, Methodology, Writing - original draft. **Alexia Roux-Sibilon:** Methodology, Formal analysis. **Pauline Rossel:** Methodology, Formal analysis, Investigation. **Muriel Boucart:** Writing - original draft. **Martial Mermillod:** Writing - original draft. **Carole Peyrin:** Conceptualization, Methodology, Writing - original draft, Supervision, Project administration.

Acknowledgment

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

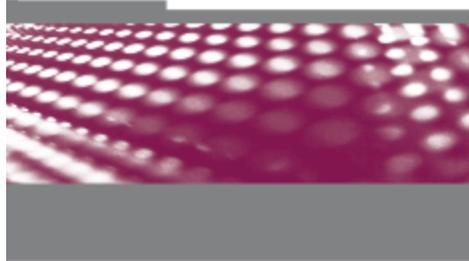
References

- Alpern, M. (1952). Metacontrast; historical introduction. *American Journal of Optometry and Archives of American Academy of Optometry*, 43, 648–657.
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of cognitive neuroscience*, 15(4), 600–609.
- Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in cognitive sciences*, 11(7), 280–289.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., ... Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2), 449–454.
- Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, 86, 35–42.
- Breitmeyer, B. G., Kafaligönül, H., Ögmen, H., Mardon, L., Todd, S., & Ziegler, R. (2006). Meta- and paracontrast reveal differences between contour- and brightness-processing mechanisms. *Vision Research*, 46(17), 2645–2658.
- Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, 62(8), 1572–1595.
- Collin, C. A. (2006). Spatial-frequency thresholds for object categorisation at basic and subordinate levels. *Perception*, 35(1), 41–52.
- Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Perception & Psychophysics*, 67(2), 354–364.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of comparative neurology*, 292(4), 497–523.
- Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology*, 159(2), 203–221.
- Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38(4), 659–671.
- Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral cortex*, 7, 181–192.
- Engel, S. A., Rumelhart, D., Wandell, B. A., Lee, A., Glover, G. H., Chichilnisky, E.-J., & Shadlen, M. (1994). fMRI of human visual cortex. *Nature*, 369, 525.
- Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in cognitive sciences*, 4(9), 345–352.
- Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, 16(3), 553–567.
- Geuzebroek, A. C., & van den Berg, A. V. (2018). Eccentricity scale independence for scene perception in the first tens of milliseconds. *Journal of Vision*, 18(9) 9-9.
- Hegd , J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, 84(4), 405–439.
- Johnson, J. S., & Olshausen, B. A. (2003). Time-course of neural signatures of object recognition. *Journal of Vision*, 3(7) 4-4.
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, 47(26), 3286–3297.
- Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of Vision*, 12(13) 16-16.
- Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, 107, 49–57.
- Kauffmann, L., Ramano l, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, 8, 37.
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, 65(2), 145–168.
- Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 471.
- Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, 9(10) 6-6.
- Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, 18(4), 513–536.
- Loschky, L. C., McConkie, G. W., Yang, J., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*, 12(6), 1057–1092.
- Loschky, L. C., Szaffarczyk, S., Beugnet, C., Young, M. E., & Boucart, M. (2019). The contributions of central and peripheral vision to scene-gist recognition with a 180° visual field. *Journal of Vision*, 19(5) 15-15.
- Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: Evidence from neuro-computational modeling. *Brain Cognition*, 57, 151–157.
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., ... Vuilleumier, P. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Journal of cognitive neuroscience*, 22(12), 2768–2780.
- Quattoni, A., & Torralba, A. (2009). Recognizing indoor scenes. *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 413–420). IEEE.
- Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, 12(6), 852–877.
- Roux-Sibilon, A., Rutg , F., Aptel, F., Attye, A., Guyader, N., Boucart, M., ... Peyrin, C. (2018). Scene and human face recognition in the central vision of patients with glaucoma. *PLoS ONE*, 13(2), e0193465.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological science*, 5(4), 195–200.
- Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, 69(3), 243–265.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520.
- Torralba, A., Murphy, K. P., Freeman, W. T., & Rubin, M. A. (2003). Context-based vision system for place and object recognition. *In Proc ICCV*.
- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, 1339(1), 190–198.
- Vanmarcke, S., Caldere, F., & Wagemans, J. (2016). The time-course of ultrarapid categorization: the influence of scene congruency and top-down processing. *i-perception*, 7(5), 1–23.
- Vanmarcke, S., & Wagemans, J. (2016). Individual differences in spatial frequency processing in scene perception: The influence of autism-related traits. *Visual Cognition*, 24(2), 115–131.
- Vanrullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, 13(4), 454–461.
- Virsu, V., N s nen, R., & Osmoviita, K. (1987). Cortical magnification and peripheral vision. *JOSA A*, 4(8), 1568–1578.
- Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research*, 37(3), 475–494.
- Wu, J., Yan, T., Zhang, Z., Jin, F., & Guo, Q. (2012). Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging. *Human Brain Mapping*, 33(7), 1727–1740.

Annexe 2

Volume 21 – Issue 1 – January 2013

Visual Cognition


 Routledge
Taylor & Francis Group

Influence of physical features from peripheral vision on scene categorization in central vision

Journal:	<i>Visual Cognition</i>
Manuscript ID	Draft
Manuscript Type:	Full Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Trouilloud, Audrey; Laboratoire de Psychologie et NeuroCognition, Rossel, Pauline; Laboratoire de Psychologie et NeuroCognition Faurite, Cynthia; Laboratoire de Psychologie et NeuroCognition Roux-Sibilon, Alexia; UCL Psychological Sciences Research Institute, Kauffmann, Louise; Laboratoire de Psychologie et NeuroCognition Peyrin, Carole; Laboratoire de Psychologie et NeuroCognition,
Keywords:	Scene perception, Peripheral vision, Central vision, Semantic interference, Spatial configuration
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Figure1.tif Figure2.tif Figure3.tif Figure4.tif	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
2
3 **Influence of physical features from peripheral vision on scene**
4 **categorization in central vision**
5
6
7
8
9
10

11 Audrey Trouilloud¹, Pauline Rossel¹, Cynthia Faurite¹, Alexia Roux-Sibilon², Louise
12 Kauffmann¹ and Carole Peyrin¹
13
14

15
16 ¹ Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

17 ² Psychological Sciences Research Institute (IPSY), UC Louvain, Louvain-la-Neuve, Belgium
18
19
20
21
22

23 Short title: Scene perception in peripheral vision
24
25
26
27

28 ***Corresponding Author:**

29 Carole Peyrin, PhD

30 Laboratoire de Psychologie et NeuroCognition (LPNC)

31 CNRS UMR 5105 - Université Grenoble Alpes

32 Bâtiment Michel Dubois - 1251 Av Centrale CS40700

33 38058 Grenoble Cedex 9 - France

34 carole.peyrin@univ-grenoble-alpes.fr
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

The spatial resolution of the human visual field decreases considerably from the center to the periphery. However, several studies have highlighted the importance of peripheral vision for rapid scene categorization. Critically, low resolution of peripheral vision could trigger predictive mechanisms that would then guide the highly detailed central vision. In Experiment 1, we investigated if peripheral vision could influence the categorization of scenes in central vision, and vice-and-versa. We used large photographs of indoor and outdoor scenes from which we extracted a central disk and a peripheral ring. Stimuli were composed of a central disk and a peripheral ring that could be either semantically congruent or incongruent. Participants had to categorize the central disk while ignoring the peripheral ring or the peripheral ring while ignoring the central disk. Results revealed a semantic interference of peripheral vision on central vision, as strong as the interference of central vision on peripheral vision. In Experiment 2, we investigated the nature of the physical signal in peripheral vision that influences central vision processing. We used either an intact peripheral ring, or a peripheral ring whose amplitude spectrum was preserved (phase-scrambling), or a peripheral ring whose phase spectrum was preserved (amplitude-scrambling). Participants had to categorize the central disk while ignoring the peripheral ring. Results showed that only phase-preserved peripheral rings elicited a semantic interference effect as strong as the one observed with intact peripheral rings. Information contained in the phase spectrum, conveying the spatial configuration of the scene, may be critical to trigger peripheral predictions.

Keywords: Scene perception; peripheral vision; central vision; semantic interference; spatial configuration

view Only

Introduction

The human visual field is the visible portion of the environment for a given retinal image during eye fixation. Central vision is the small portion of the visual field where the eyes look directly. It is considered as the most important part of the visual field, certainly due to its very good spatial resolution, or visual acuity. Visual acuity, which is the ability to discern details, is the highest in a small area of central vision, the fovea. It is estimated to cover approximately 1° of eccentricity (Curcio & Allen, 1990; Curcio et al., 1990; Larson & Loschky, 2009; Yamada, 1969). The remaining part of the visual field projects onto the parafovea, which extends to about $4\text{--}5^\circ$ of retinal eccentricity, and onto the peripheral retina, which covers the entire remaining visual field (Larson & Loschky, 2009; Loschky, Nuthmann, Fortenbaugh, & Levi, 2017). Although our subjective visual experience seems rich and detailed (in part thanks to constant eye movements), the spatial resolution decreases drastically with eccentricity and only low spatial resolution information is available in most of the visual field. Indeed, the density of photoreceptors is the highest at the fovea and decreases with retinal eccentricity. This is further amplified at the cortical level, where central vision is overrepresented relative to peripheral vision (cortical magnification). Namely, more neurons of the primary visual cortex (V1) are allocated to the processing of information from the fovea than periphery (Daniel & Whitteridge, 1961). Due to its favored processing at both the retinal and cortical levels, central vision gives us the best visual acuity, allowing us to read, recognize faces and objects, and do a variety of other daily tasks.

However, several studies in visual recognition research have demonstrated the crucial role of low resolution peripheral information for a rapid categorization of scenes (Boucart et al., 2013; Larson & Loschky, 2009; Loschky et al., 2019; Trouilloud et al., 2020). In particular, peripheral vision alone is sufficient to perform rapid scene categorization. This has been first demonstrated by Larson and Loschky (2009), who used an experimental paradigm known as the window-scotoma to examine the relative contributions of central and peripheral vision during rapid scene categorization. Participants of that study were presented with either the central part of a scene (window condition), the peripheral part (scotoma condition), or the whole scene (control condition). Radii of window and scotoma stimuli varied according to four different values (1° , 5° , 10.8° , or 13.6°). According to the authors, as the central vision extends to about 5° of visual angle, the 5° radius condition was used to compare the relative contribution of central and peripheral vision. Results showed that categorization performances were better in the 5° scotoma condition (i.e., peripheral vision) than in the 5° window condition (i.e., central vision). Moreover, the performance in the scotoma condition when only peripheral information was available did not differ from the maximum performance obtained in the control condition where the whole scene was visible. Low resolution peripheral vision would therefore be more useful than high resolution central vision for quickly categorizing a scene. It should be noted that peripheral vision would have an advantage over central vision for scene categorization only because it covers a larger part of the visual field revealing more scene content. Indeed, Geuzebroek and van den Berg (2018) observed that when they presented the same amount of semantic content in central and peripheral vision (by manipulating the size of a scene), the advantage of peripheral vision disappeared. The process of scene categorization would

1
2
3 therefore be independent of the eccentricity of scene presentation if the amount of information
4 remains identical in central and peripheral vision, which is not the case in real-life conditions.
5

6 In a subsequent experiment using the same paradigm, Larson, Freeman, Ringer, and
7 Loschky (2014) manipulated the presentation times to examine how the relative contributions
8 of the central and peripheral visual field change over time (Larson, Freeman, Ringer, &
9 Loschky, 2014). The results revealed an advantage in favor of central processing during the
10 first 100 ms, after which the relative contribution of peripheral vision increases. To interpret
11 these results, the authors proposed the zoom-out hypothesis in which attention is first allocated
12 at the locus of fixation and expands outward over the span of a fixation. The advantage of
13 peripheral vision was also found by Trouilloud et al. (2020) who replicated the findings of
14 Larson and Loschky (2009) using stimuli that consider the cortical magnification factor. More
15 precisely, stimuli were built so that the image content presented in peripheral and central vision
16 activated approximately the same amount of neurons in the primary visual cortex. However, in
17 this study, the advantage of peripheral vision was observed even for very short presentation
18 time (33 ms), which contradicts the zoom-out hypothesis. To go further, Trouilloud et al. (2020)
19 conducted a second experiment to determine whether the coarse peripheral information could
20 be used to guide the processing of detailed information from central vision. They examined
21 whether scene analysis followed a peripheral to central analysis rather than a central to
22 peripheral analysis during rapid scene categorization. To this end, they created dynamic
23 sequences of the scene which participants had to categorize as indoor or outdoor scenes. The
24 sequences were composed of five versions of a scene revealing part of the visual information
25 through a set of rings of different eccentricities presented sequentially, from the peripheral to
26 the central part or from the central to the most peripheral part. The results revealed an advantage
27 for the peripheral to central processing sequences. Contrary to the zoom-out hypothesis, low
28 resolution of peripheral information could therefore allow a first categorization of the scene
29 which would then be validated by the analysis of the high resolution of central information.
30
31
32
33
34
35
36
37

38 According to predictive models of visual recognition (Bar, 2003; Bar, 2007; Trapp &
39 Bar, 2015; Kauffmann et al., 2014; Kveraga et al., 2007; Peyrin et al., 2010), low resolution
40 information is sufficient and useful to trigger cortical predictive mechanisms that would then
41 guide a more detailed visual analysis. Low spatial frequency content of the retinal image would
42 be rapidly projected to the cortex (especially to the orbitofrontal cortex), allowing to activate a
43 coarse representation of contextual conformation which could then be used to generate
44 predictions about the objects likely to be found. These predictions, or expectations, would thus
45 help to refine the possible interpretations of the object and facilitate the processing of the
46 detailed information contained in high spatial frequencies. Since low spatial frequencies are
47 mostly available in peripheral vision, Roux-Sibilon et al. (2019) studied the hypothesis that
48 information from peripheral vision may be used to generate predictive signals that guide the
49 processing of information from central vision. The authors tested the effect of a peripheral
50 scene background on the categorization of an object in central vision using a semantic
51 interference paradigm. The peripheral background was either congruent (e.g., a chair in an
52 indoor scene) or incongruent with the object (e.g., a chair in an outdoor scene). Also, the
53 peripheral background was presented either simultaneously to the object, 30 ms before, or 150
54 ms before. The authors observed that participants were faster to categorize the object when it
55 was in a congruent than incongruent environment, suggesting an automatic processing of
56
57
58
59
60

1
2
3 peripheral information. Importantly, this congruence effect only occurred when the peripheral
4 background was perceived slightly before the object (30 ms and 150 ms), and the effect of
5 congruence increased linearly with the stimulus onset asynchrony (SOA) between the
6 peripheral background and the central object, supporting the view that peripheral information
7 is used to guide visual processes in central vision in a predictive manner, rather than central
8 and peripheral information being processed concurrently and integrated in a feedforward
9 manner.

10
11
12 In a recent fMRI study, Peyrin et al. (2021) also observed the influence of peripheral
13 vision on the categorization of scenes in central vision. A scene to ignore was presented in
14 peripheral vision and a target scene to categorize appeared simultaneously in central vision.
15 The scene pairs were either semantically congruent or incongruent and could share the same
16 physical properties (similar amplitude spectrum and spatial configuration) or not. The results
17 showed that participants made more errors and were slower to categorize the central scene
18 when the peripheral scene was semantically incongruent than when it was semantically
19 congruent. At the cortical level, this semantic interference effect was associated with a stronger
20 activation of the orbitofrontal cortex, consistent with previous fMRI studies addressing the
21 predictive brain hypothesis in visual categorization (Bar et al., 2006; Kauffmann, Bourgin,
22 Guyader, & Peyrin, 2015; Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Peyrin et al., 2010).
23 Critically, this effect was in fact observed when the two scenes shared the same physical
24 properties, highlighting the importance of both semantic and physical information available in
25 peripheral vision in the generation of predictions. However, the nature of the physical features
26 that were used to generate predictions remains unknown in this study. In the Fourier domain,
27 an image is described by both its amplitude spectrum and its phase spectrum. The amplitude
28 spectrum provides the distribution of luminance contrast as a function of orientations and
29 spatial frequencies. The phase spectrum provides information about the relative position of the
30 objects in a scene, that is the spatial configuration of the scene. Given that the physically similar
31 scenes used in Peyrin et al. (2021) were matched both in terms of amplitude spectrum and
32 spatial configuration (i.e., pixel to pixel correlation), it could not be disambiguated whether it
33 is the information in the amplitude spectrum, in the phase spectrum, or both that is used to
34 generate predictions. A possible answer can be provided by a second experiment conducted by
35 Roux-Sibilon et al. (2019) in which the authors directly addressed how low-level physical
36 features of a peripheral scene background could influence the categorization of an object in
37 central vision. In this experiment, the peripheral background was either intact or phase
38 scrambled (i.e., spatial frequency and orientation distribution was preserved while the phase of
39 the scene was altered). An interference effect was only observed for intact scene background
40 suggesting that the information contained in the phase spectrum, rather than in the amplitude
41 spectrum alone, was automatically used in peripheral vision to generate predictions about the
42 category of the object in central vision.

43
44 Finally, some studies have shown that the categorization of the scene background can
45 be also influenced by the semantic congruence of an object in the scene (Davenport, 2007;
46 Davenport & Potter, 2004; Joubert et al., 2008). Therefore, Lukavský (2019) addressed the
47 reciprocal influence between central and peripheral vision during scene categorization. Based
48 again on a semantic interference paradigm, the author presented stimuli composed of a central
49 disk and a peripheral ring revealing the central or peripheral parts of a scene respectively. The
50
51
52
53
54
55
56
57
58
59
60

1
2
3 disk and the ring could belong to the same scene (congruent) or to two scenes from different
4 categories (incongruent). Participants had to categorize the central disk while ignoring the
5 peripheral ring or to categorize the peripheral ring while ignoring the central disk. Stimuli were
6 presented for 33 ms, followed by a blank screen, and then a mask (stimulus-to-mask onset
7 asynchrony of 200 ms). Results revealed that participants made more errors when the central
8 disk and the ring were incongruent than congruent, whatever the position in the visual field of
9 the stimuli to categorize (i.e., the central disk or the peripheral ring). This result indicates that
10 the information to be ignored was automatically processed and interfered with the processing
11 of the information to attend. Importantly, this interference effect was as strong for a central
12 categorization with a peripheral distractor than for a peripheral categorization with a central
13 distractor. This result indicates that scene categorization relies on both central and peripheral
14 visual information. To ensure that the absence of interaction was not due to a too long
15 presentation time between the stimulus and the mask (200 ms) that could favor the processing
16 of both central and peripheral information despite a precedence of one information on the other,
17 Lukavský (2019) conducted a second experiment with the same procedure but additionally
18 manipulated the stimulus-to-mask onset asynchrony. Even at the shortest asynchrony (33 ms),
19 the interference was as strong for the central and the peripheral distractors. According to the
20 author, this result contradicted the hypothesis of a prioritization of the central processing at
21 very short presentation time as stated in the zoom-out hypothesis of Larson et al. (2014). In
22 this context, we can assume that it also contradicts the hypothesis of a prioritization of the
23 peripheral processing as suggested by the advantage for the peripheral to central processing
24 sequences in Trouilloud et al. (2020). Lukavský (2019) interpreted the results in terms of
25 evidence accumulation (Mirolli et al., 2010) or random walk model (Gold & Schadlen, 2007)
26 postulating that all the perceived stimuli provide evidence supporting more or less several
27 possible hypotheses. Evidence is then gradually accumulated by the system until the decision
28 threshold of one of the hypotheses is reached. Since central and peripheral vision influence
29 each other, Lukavský (2019) suggested that evidence from different parts of the visual field is
30 gradually gathered during gist recognition. Peripheral and central information would not be
31 processed in isolation but concurrently, so that information extracted in peripheral vision would
32 influence the analysis of central content and vice versa. Importantly, this experimental
33 paradigm demonstrated the influence of peripheral vision on the central vision processing. It
34 should however be noted that the radius of the central disk was fixed to 5.54° of visual angle,
35 so that it contained foveal and parafoveal information, while the peripheral ring extended from
36 5.54° to 11.08° of eccentricity. Due to the cortical magnification, the central disk was over-
37 represented at the cortical level relative to the peripheral ring. Therefore, we believe that a
38 prioritization of the peripheral processing to trigger predictive mechanisms is still plausible
39 after considering this limit.

40
41
42
43
44
45
46
47
48
49
50
51
52 The aim of the present study was to test the hypothesis that peripheral vision is rapidly
53 used to trigger predictive mechanisms using a semantic interference paradigm inspired by
54 Lukavský (2019) but considering the cortical magnification factor. To this end, we used the
55 stimuli of Trouilloud et al. (2020) which allow us to stimulate different parts of the visual field
56 while controlling that approximately the same cortical surface is activated in V1. Participants
57 were presented simultaneously with a central disk (CD) and a peripheral ring (PR) revealing
58 part of scenes that could either belong to the same category (semantically congruent) or to
59
60

1
2
3 different categories (semantically incongruent). In Experiment 1, participants had to categorize
4 the central disk while ignoring the peripheral ring or categorize the peripheral ring while
5 ignoring the central disk in the same way as the Lukavský (2019) experiment. Based on
6 previous works using a semantic interference paradigm (Lukavský, 2019; Roux-Sibilon et al.,
7 2019; Peyrin et al., 2021), we expected to observe a semantic interference effect (i.e., impaired
8 performance in the semantically incongruent condition relative to the semantically congruent
9 condition) regardless of the position of the scene to be categorized (central vs. peripheral).
10 Moreover, if information in peripheral vision is rapidly processed to generate a predictive
11 signal that guides visual processing in central vision, we expected a greater interference for the
12 categorization of the central disk than the peripheral ring. In Experiment 2, we investigated the
13 nature of the physical signal in peripheral vision that influences central vision processing. We
14 used either an intact peripheral ring, or a peripheral ring whose amplitude spectrum was
15 preserved (phase-scrambling), or a peripheral ring whose phase spectrum was preserved
16 (amplitude-scrambling). Participants had to categorize the central disk while ignoring the
17 peripheral ring. As for Experiment 1, we expected to observe a semantic interference effect for
18 the intact peripheral ring. Furthermore, if information contained in the phase spectrum (i.e.,
19 spatial configuration of the scene) is used to generate predictions, we expected an interference
20 effect in this condition. Similarly, if the information contained in the amplitude spectrum (i.e.,
21 distribution of information across spatial frequencies and orientations) is used to generate
22 predictions, we expected an interference effect in this condition.
23
24
25
26
27
28
29
30
31

32 **Experiment 1**

33 **Method**

34 **Participants**

35
36
37
38
39 Twenty-four undergraduate students of Psychology from University Grenoble Alpes
40 (19 women, mean age \pm standard deviation: 20.83 \pm 2.47) participated in the experiment. All
41 were right handed, with normal or corrected-to-normal vision. The study was approved by the
42 ethics committee of the University Grenoble Alpes (CER-Grenoble Alpes, COMUE University
43 Grenoble Alpes, IRB00010290). All participants involved in the study gave their informed
44 written consent.
45
46
47

48 **Stimuli**

49
50 We used the stimuli from Trouilloud et al. (2020)'s experiment. The stimuli were
51 constructed from 50 color photographs from the Pixabay website (<https://pixabay.com/fr/>), a
52 photo sharing site under CC0 (Creative Commons Zero) license. Half of the photographs
53 represented indoor scenes (e.g., living room, kitchen, bathroom), and the other half represented
54 outdoor scenes (e.g., cityscape, mountain, beach). We choose an indoor vs outdoor
55 categorization task since the superordinate-level precedes basic-level distinctions for rapid
56 scene categorization (Rousselet, Joubert, & Fabre-Thorpe, 2005; Joubert, Rousselet, Fize, &
57 Fabre-Thorpe, 2007; Loschky & Larson, 2010; Kadar & Ben-Shahar, 2012).
58
59
60

Image manipulation was performed using MATLAB (Mathworks Inc., Sherborn, MA, USA). For each scene, we built a square image of 1500×1500 pixels centred on an object representative of its category. The object fitted into a 100×100 pixel square. For indoor scenes, the object could be a lamp, a book, a candle, a radio, a clock, a vase, a household appliance, a pair of glasses, dishes or kitchen utensils. For outdoor scenes, the object could be a house, a tree, a plant, a car, a building, a boat, a mountain or a bridge. All images were converted to 256 gray levels by averaging the luminance values from the three color channels at each pixel. All stimuli were normalized to obtain an average luminance of 0.5 and a RMS contrast (Root Mean Square) of 0.2 for pixel intensity values between 0 and 1 (i.e., an average luminance of 128 and an RMS contrast of 51 on a scale of 256 grey levels). The RMS contrast corresponded to the mean RMS of the whole set of images. Thus, we created 50 scenes of size 1500×1500 pixels (25 indoor and 25 outdoor scenes).

To cover both central and peripheral vision, the angular size of stimuli was fixed at $\sim 30^\circ$ of visual angle. Then, for each image, Trouilloud et al. (2020) built five stimuli consisting of a central disk and four peripheral rings (Fig. 1b) revealing different parts of the scene. The eccentricity and size of each stimuli was defined using an equation empirically derived from retinotopic measurement (Wu, Yan, Zhang, Jin, & Guo, 2012; for a similar procedure, see Geuzebroek & van den Berg, 2018):

$$y = \frac{\left(\int_{r_{inner}}^{r_{outer}} (553.99 - 123.98 \ln(x)) dx \right)}{7.5}$$

In this equation, y is the integrated V1 surface area in square millimeters, representing the activated cortical surface, and r_{inner} and r_{outer} are the inner and outer radii in degrees, respectively. This progression was chosen to better consider the properties of the visual system. Indeed, information from central and peripheral visual fields undergoes a deformation on the visual cortex: The central information is over-represented at cortical level, relative to peripheral information (i.e., cortical magnification; Daniel & Whitteridge, 1961). In order to consider these transformations at cortical level, the surface of the image revealed by the rings increased with their eccentricity. Based on this equation, Trouilloud et al. (2020) created five stimuli of different eccentricities (a central disk and four peripheral rings) but with the same surface of activation on V1 (20%, i.e. $\sim 136 \text{ mm}^2$). For the central stimuli, the scene was revealed through a small disk of only 1.66° radius of visual angle (i.e. 83 pixels). For the first peripheral ring, adjacent to the central disk, the inner edge was fixed at 1.66° visual angle (83 pixels) and the outer edge was fixed at 4.10° visual angle (205 pixels). For the second peripheral ring, adjacent to the first one, the inner edge was fixed at 4.10° visual angle (205 pixels) and the outer edge was fixed at 7.10° visual angle (355 pixels). For the third peripheral ring, adjacent to the second one, the inner edge was fixed at 7.10° of visual angle (355 pixels) and the outer edge was fixed at 10.70° of visual angle (535 pixels). Finally, for the fourth peripheral ring, adjacent to the third one, the inner edge was fixed at 10.70° visual angle (535 pixels) and the outer edge was fixed at 15° visual angle (750 pixels).

In the present experiment, we only used the central disk and the third peripheral ring from Trouilloud et al. (2020; Figure 1). We specifically used the R3 as the peripheral ring (PR)

for two reasons. First, it was the stimulus with the best categorization performance in peripheral vision (for average error rates and average correct response times). Second, Trouilloud et al (2020) showed that there is a large area of peripheral vision for which categorization is optimal (i.e., categorization of peripheral ring 1, 2, and 3 did not differ from the categorization of the whole scene). Beyond this area, recognition is less efficient. The central disk (CD) and the third peripheral ring (PR) stimuli were presented simultaneously. CD and PR could be either semantically congruent (e.g., a central indoor scene and the same peripheral indoor scene) or semantically incongruent (e.g., a central indoor scene and a peripheral outdoor scene). Stimuli were presented on a gray background of 0.5 average luminance.

Stimuli can be downloaded from <https://osf.io/q3rtw/>.

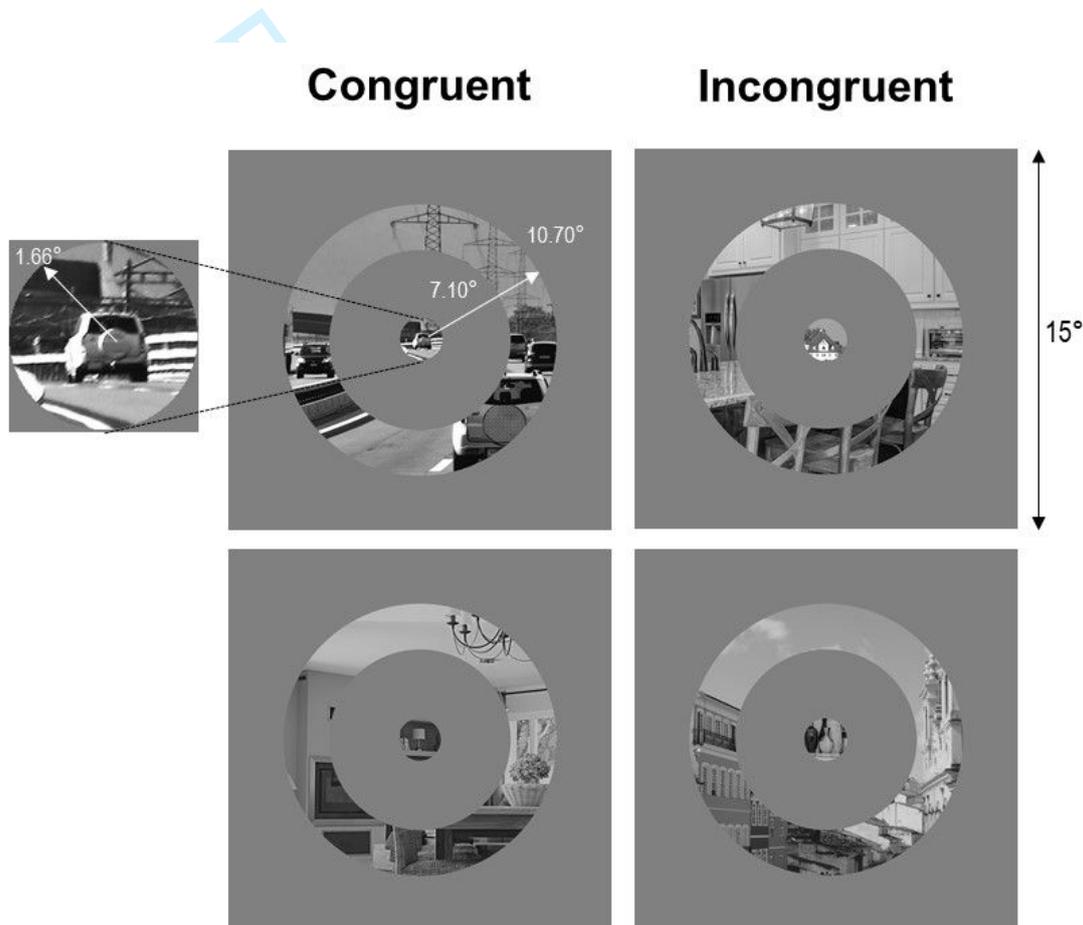


Figure 1. Example of stimuli used in Experiment 1. The central disk and the peripheral ring stimuli were presented simultaneously. They could be either semantically congruent (e.g., a central outdoor scene and the same peripheral outdoor scene) or semantically incongruent (e.g., a central outdoor scene and a peripheral indoor scene).

Procedure

Stimuli were displayed using the Psychtoolbox (Brainard, 1997; Pelli, 1997) implemented in MATLAB R2019 (MathWorks, Natick, MA, USA) on a 30' monitor (DELL ULTRASHARP), with a resolution of 2560×1600 pixels and a refreshing rate of 60 Hz. The

1
2
3 participants' heads were placed on a chinrest at 70 cm from the screen in order to respect the
4 stimuli angular size. All participants performed two sessions. In the first session, participants
5 had to categorize the central disk (the target) and ignore the peripheral ring (the distractor). In
6 the second session, participants had to categorize the peripheral ring (the target) and ignore the
7 central disk (the distractor). The two experimental sessions were counterbalanced between each
8 participant. For each session, a trial began with a central black fixation point presented for 500
9 ms on a 0.5 luminance background, followed by both the central disk and peripheral ring for
10 100 ms on a 0.5 luminance background and then, by a gray screen (0.5 luminance) of 1900 ms
11 during which participants could respond. Participants were asked to categorize the target (either
12 the central disk or the peripheral ring) as an indoor or an outdoor scene. Participants were
13 instructed to answer as correctly and as quickly as possible and to fixate on the centre of the
14 screen for the duration of the trial. They were instructed to press the corresponding response
15 key with the middle finger and the forefinger of the right hand. Response keys were
16 counterbalanced across participants. Each session included 160 trials (40 congruent stimuli
17 from the Indoor category, 40 congruent stimuli from the Outdoor category and 80 incongruent
18 stimuli) and lasted 45 min, with breaks every 80 trials. For each trial, response accuracy and
19 response time (in ms) were recorded. Before each experimental session, participants performed
20 a training session (25 trials for each session) using stimuli that were not included in the main
21 experiment.

22
23 In order to ensure that participants fixated on the centre of the screen for the duration
24 of a trial, we recorded eye movements throughout the experiment. Eye movements were
25 recorded using an Eyelink 1000 eye-tracker (SR Research) with a sampling rate of 1,000 Hz
26 and a nominal spatial resolution of 0.01° of visual angle. For each participant, only the right
27 eye was recorded using the "pupil-corneal reflection" mode. The Eyelink software
28 automatically detected saccades with the following thresholds: speed $>30^\circ/\text{s}$, acceleration
29 $>8,000^\circ/\text{s}^2$, and saccadic displacement $>0.15^\circ$. Fixations were detected when the pupil was
30 visible and no saccade was in progress. Blinks were detected during partial or total occlusion
31 of the pupil. Each session was preceded by a calibration procedure in which participants had
32 to orient their gaze toward nine separate dots appearing sequentially in a 3×3 grid that
33 occupied the entire display. A drift correction was performed every 10 trials. A new calibration
34 was done in the middle of the experiment.

35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

Data Analysis

Data were analysed using R (R Core Team, 2019) and *lme4* package (Bates, Mächler, et al., 2015) to perform the analyses on the error rates (ER) and correct response times (RT). To analyse the ER, we set up a mixed effect logistic (the dependent variable being a binary variable) regression model. The participant's response was coded 1 when incorrect and 0 when correct. To analyse RT, we set up a linear mixed effects model of the Congruence (congruent vs. incongruent) and the Target position (central vs. peripheral) in a 2×2 within-subjects design. Congruence, Target position and their interaction term were entered into each model as fixed effects (i.e., as the effects of the variables of interest). Intercepts for subjects, as well as subject-wise random slopes for the Congruence effect, Target position and their interaction were specified as random effects (i.e., as the effects of the variables to which we want to generalize our results). Based on previous works using semantic interference paradigm

(Lukavský, 2019; Roux-Sibilon et al., 2019; Peyrin et al., 2021), we expected to observe a semantic interference effect regardless of the position of the Target (central vs. peripheral). Moreover, if information in peripheral vision is used to generate a predictive signal that guides visual processing in central vision, the interference effect should be greater for the categorization of the central target than the peripheral target. To test these hypotheses, we calculated the main effects of Congruence, Target position as well as their interaction for ER and RT.

We used the method proposed by Bates, Kliegl et al. (2015) to construct parsimonious mixed models preventing convergence problems. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained with a Wald test for ER analyses and by Satterthwaite approximation for RT analyses with the *lmerTest* package (Kuznetsova et al., 2017). The significance threshold was set at .05 for ER and RT analyses. Effect sizes were estimated using Cohen's d_z (Lakens, 2013).

Eye movements were used to control that the gaze position on each trial was maintained within 1.7° of eccentricity from the centre of the screen (i.e. the radius of the CD stimulus). For each participant and each experimental session, we removed trials in which saccades were initiated with an amplitude beyond 1.7° around the fixation cross. This resulted in removing 0.78% of the trials. Data from one participant who shifted his gaze beyond 1.7° of eccentricity from the center of the screen in a majority of trials were also discarded from the analyses. The statistical analyses were therefore conducted on 23 participants (18 women, mean age: 20.77 ± 2.43). We also removed 0.19% aberrant RT lower than 0.

Data are available in the Open Science Framework repository: <https://osf.io/q3rtw/>.

Results

Results are shown in Figure 2. The analyses performed on ER revealed a significant main effect of the Congruence, $\beta = 0.67$, $z = 5.49$, $p < .001$. Participants made more errors when the two scenes were semantically incongruent ($10.10 \pm 6.28\%$) than congruent ($5.22 \pm 4.64\%$). This indicated that when the two scenes were semantically incongruent, the scene to ignore was processed at the semantic level and interfered with the categorization of the target (i.e., semantic interference effect). We also observed a significant main effect of the Target position, $\beta = 0.35$, $z = 2.41$, $p = 0.016$. Specifically, participants made fewer errors for categorizing the central disk ($6.09 \pm 4.99\%$) than the peripheral ring ($9.24 \pm 6.04\%$). However, the interaction between the Congruence and the Target position was not significant, $\beta = 0.10$, $z = 0.45$, $p = 0.656$, suggesting that the semantic interference effect was as important for categorizing a target in central vision with a distractor in peripheral vision as for categorizing a target in peripheral vision with a distractor in central vision.

The analyses performed on RT revealed a significant main effect of the Congruence, $\beta = 35.25$, $t(6670.97) = 8.08$, $p < 0.001$, $d_z = 1.685$. Participants were slower when the two scenes were semantically incongruent (683 ± 44 ms) than congruent (648 ± 45 ms), suggesting a semantic interference effect on response times. We also observed a significant main effect of the Target position, $\beta = 59.78$, $t(21.88) = 3.98$, $p < 0.001$, $d_z = 0.830$. Participants were faster for categorizing the central disk (637 ± 39 ms) than the peripheral ring (695 ± 49 ms). As for the ER analysis, the interaction between the Congruence and the Target position was not significant, $\beta = -5.25$, $t(22.06) = -0.368$, $p = 0.717$, $d_z = 0.077$, suggesting again that the

semantic interference effect was as important for categorizing a target in central vision with a distractor in peripheral vision as for categorizing a target in peripheral vision with a distractor in central vision.

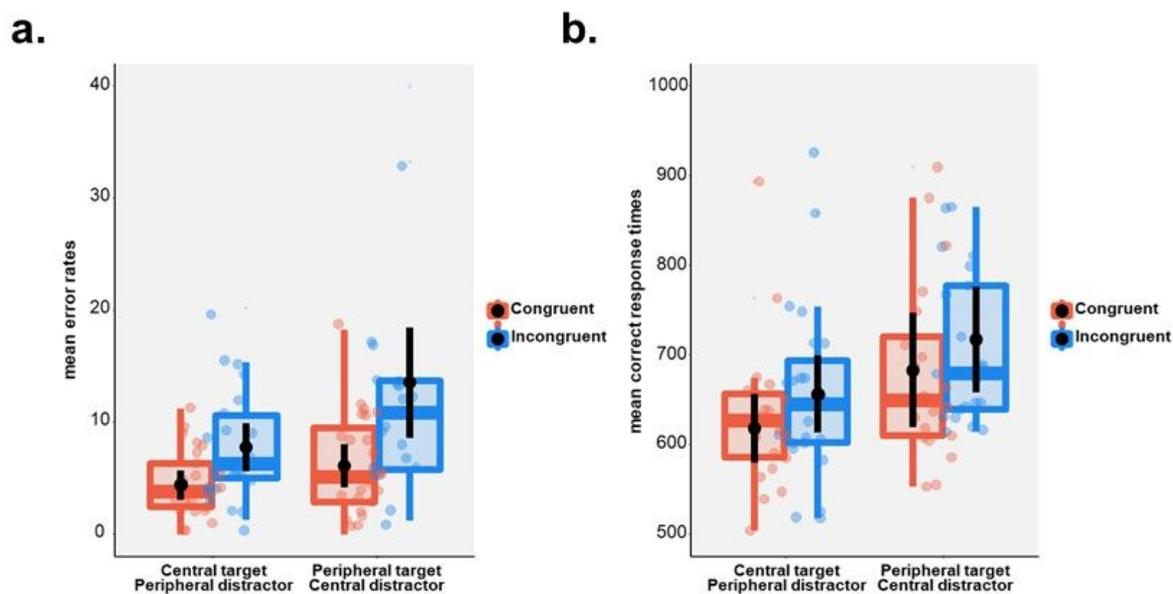


Figure 2. Box plots of (a) mean error rates in percentage and (b) mean correct response times in milliseconds during the categorization of the central target and the peripheral target according to their congruence (congruent, incongruent). Black dots and error bars indicate the mean and standard error, respectively. Color dots correspond to individual observations.

Discussion

The aim of this first experiment was to examine whether peripheral information could influence the processing of central information during rapid scene categorization. To this end, we used a semantic interference paradigm in which a central disk and a peripheral ring depicting two congruent or incongruent scenes were presented simultaneously. The task consisted in categorizing the scene contained in the central disk while ignoring the peripheral ring (i.e., peripheral distractor) or in categorizing the peripheral ring while ignoring the central disk (i.e., central distractor).

The results showed a semantic interference effect. Namely, participants were faster and made fewer errors to categorize congruent than incongruent scenes. When the distractor does not belong to the same category as the target, this conflicting information disturbs the categorization of the target scene, so that the visual system is more likely to make more categorization errors and needs more time to categorize the scene. Although information from the distractor is not relevant to the task, it is automatically processed and interferes with the categorization of the target scene. Based on our hypothesis that information in peripheral vision is used to generate a predictive signal that guides visual processing in central vision, we

1
2
3 expected the interference effect of the distractor to be larger for the categorization of the central
4 disk than the peripheral ring. However, this interference effect of peripheral information on
5 central categorization was as important as the interference effect of central information on
6 peripheral categorization, in line with the Lukavský (2019)'s findings. This result will be
7 addressed in the general discussion.
8
9

10 However, it should be noted that the task required the participants to keep focusing their
11 gaze at the center of the screen throughout the experimental trial (from the presentation of the
12 fixation point to their response), which may have added difficulty in ignoring this central
13 information when we asked participants to categorize peripheral stimuli while ignoring central
14 information. Besides, even if participants had no difficulties shifting their attention between
15 the two experimental sessions (central disk target vs. peripheral ring target), they had better
16 performances (both on error rates and response times) when categorizing the central disk with
17 a peripheral distractor (irrespective of its congruence) than when categorizing the peripheral
18 ring with a central distractor. This suggests that it is difficult to ignore the high-resolution
19 information from central vision. Nevertheless, keeping the gaze fixed in the centre of the screen
20 is inherent to conditions for studying vision.
21
22
23

24 Overall, the results of this first experiment are consistent with previous findings of
25 Lukavský (2019) showing mutual influence between central and peripheral vision during rapid
26 scene categorization. High-resolution information from central vision cannot not be ignored
27 and therefore influences peripheral information processing. However, the influence of
28 peripheral information on central vision processing appears to be as important as central
29 influence despite the low-resolution inherent to peripheral vision. Although irrelevant for the
30 task, peripheral information was also automatically processed and interfered with the
31 categorization of the central scene. To go further, we aimed at determining the nature of the
32 physical features used in peripheral information to influence central vision processing. To this
33 end, we conducted a second experiment (Experiment 2) using the same paradigm but we used
34 either an intact peripheral ring, or a peripheral ring whose amplitude spectrum was preserved
35 (phase-scrambling), or a peripheral ring whose phase spectrum was preserved (amplitude-
36 scrambling). In this way, we were able to independently assess the contribution of the phase
37 signal and that of the amplitude signal in the semantic interference of peripheral vision.
38
39
40
41
42
43
44
45

46 **Experiment 2**

47 **Methods**

48 **Participants**

49 Twenty-two undergraduate students of Psychology from University Grenoble Alpes
50 (21 women, mean age: 20.45 ± 1.71) participated in the experiment. The sample size was
51 chosen based on a power analysis with estimated effect size of $d_z = 1.296$ (effect size of the
52 interference effect for a central categorization in the RT analysis of Experiment 1) to achieve
53 power of 0.99 at an alpha level of 0.05. All were right handed, with normal or corrected-to-
54 normal vision. The study was performed within the same ethical framework as Experiment 1.
55
56
57
58
59
60

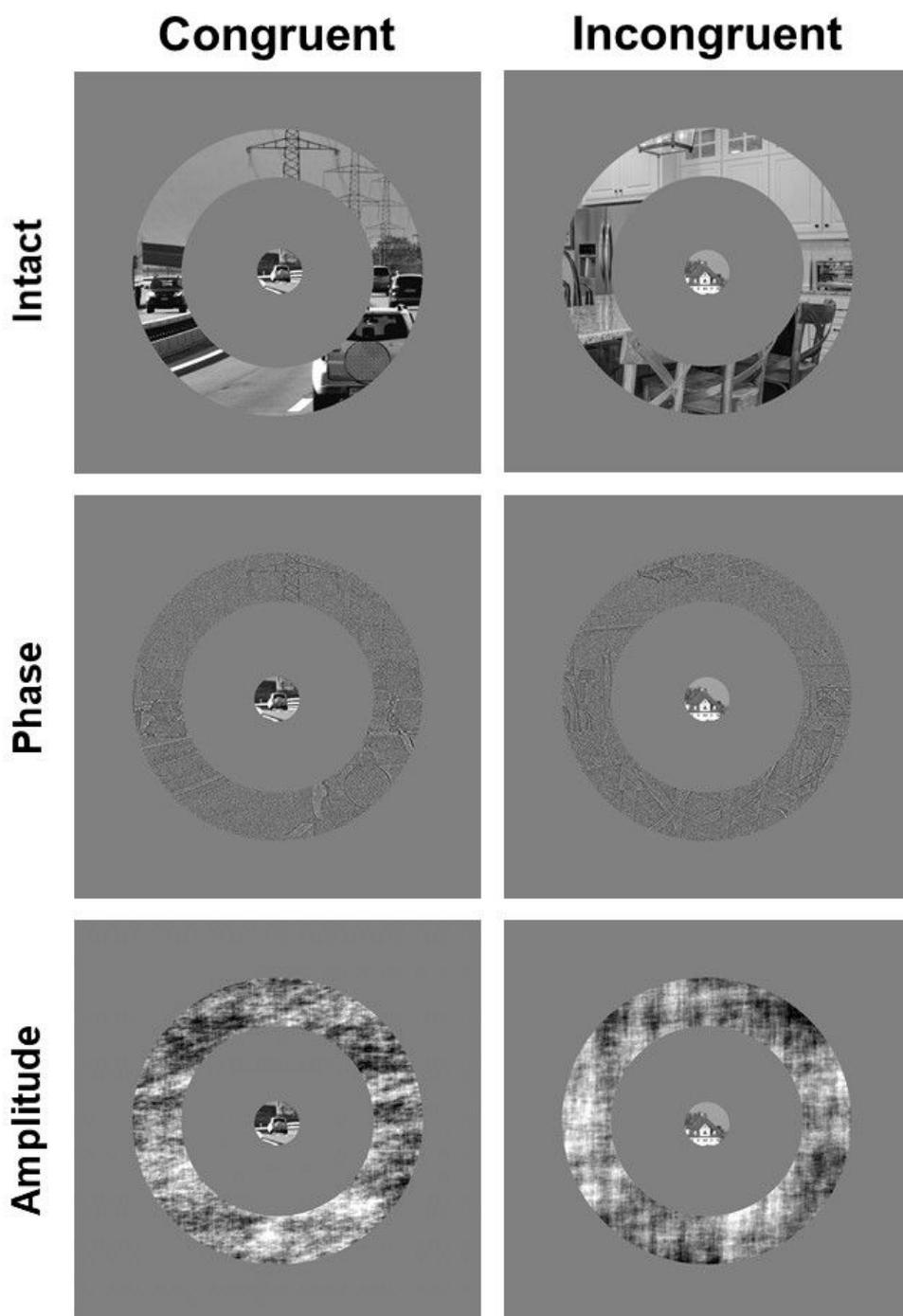
Stimuli and procedure

The central disk (CD) and peripheral ring (PR) stimuli were built from the same stimuli as those used in Experiment 1. In this experiment, we manipulated the features available in PR stimuli which could be either Intact, Phase-preserved (amplitude-scrambled) or Amplitude-preserved (phase-scrambled). Phase-preserved and Amplitude-preserved stimuli were created as follows. First, we calculated the Fourier transform of the scene to obtain its amplitude spectrum and its phase spectrum. We also created an image of the same size with white noise and computed its Fourier transform in order to obtain an amplitude spectrum and a phase spectrum of a white noise. To create the Phase-preserved stimuli, we then performed an inverse Fourier transform using the phase spectrum of the scene and the amplitude spectrum of the white noise. To create the Amplitude-preserved stimuli, the inverse Fourier transform was done using the amplitude spectrum of the scene and the phase spectrum of the white noise. All images brought back into the spatial domain were then equalized in terms of mean luminance (0.5) and RMS contrast (0.2) as in Experiment 1. We then built a third peripheral ring by following the procedure described in Experiment 1.

As in Experiment 1, CD and PR stimuli were presented simultaneously (Figure 3). They could be either semantically congruent (e.g., a central indoor scene and the same peripheral indoor scene) or semantically incongruent (e.g., a central indoor scene and a peripheral outdoor scene). The information contained in the PR stimulus could be either Intact, Phase-preserved or Amplitude-preserved.

Stimuli were displayed using the Psychtoolbox implemented in MATLAB R2019 (MathWorks, Natick, MA, USA) using the same 30' monitor as in Experiment 1 with a viewing distance of 70 cm in order to respect the stimuli angular size. All participants had to categorize the central scene and ignore the peripheral scene. A trial began with a central black fixation point presented for 500 ms on a 0.5 luminance background, followed by the stimulus for 100 ms on a 0.5 luminance background and then, by a gray screen (0.5 luminance) for 1900 ms during which participants could respond. Participants' task was similar to Experiment 1 (i.e., indoor vs. outdoor categorization). They were instructed to answer as correctly and as quickly as possible and to fixate on the centre of the screen for the duration of the experiment. Central fixation was monitored using the same Eyelink 1000 eye-tracker (SR Research) as in Experiment 1. They were instructed to press the corresponding response key with the middle finger and the forefinger of the right hand. Response keys were counterbalanced across participants. The experiment included 240 trials (80 intact PR including 40 congruent PR and 40 incongruent PR; 80 phase-preserved PR including 40 congruent PR and 40 incongruent PR; 80 amplitude-preserved PF including 40 congruent PR and 40 incongruent PR) and lasted 30 minutes, with a break in the middle of the experiment. Trials were fully randomized. For each trial, response accuracy and response time (in ms) were recorded. Before the experiment, participants performed a training session (25 trials) using stimuli that were not included in the main experiment.

Stimuli can be downloaded from <https://osf.io/q3rtw/>.



49
50
51
52
53
54
55
56
57
58
59
60

Figure 3. Example of stimuli used in Experiment 2. The central disk and the peripheral ring stimuli were presented simultaneously. They could be either semantically congruent (e.g., a central outdoor scene and the same peripheral outdoor scene) or semantically incongruent (e.g., a central outdoor scene and a peripheral indoor scene). The information contained in the peripheral ring stimulus could be either intact, phase-preserved (i.e., spatial configuration of the scene was preserved) or amplitude-preserved (i.e., distribution of information across spatial frequencies and orientations was preserved).

Data Analysis

Data were analysed using R (R Core Team, 2019) and *lme4* package (Bates, Mächler, et al., 2015) to perform the analyses on the error rates (ER) and correct response times (RT). To analyse the ER, we set up a mixed effects logistic regression model. The participant's response was coded 1 when incorrect and 0 when correct. To analyse RT, we set up a linear mixed effects model of the Congruence (congruent vs. incongruent) and the Physical features of peripheral information (intact vs. amplitude-preserved vs. phase-preserved) in a 2×3 within-subjects design. Congruence, Physical features and their interaction term were entered into each model as fixed effects. Intercepts for subjects, as well as subject-wise random slopes for the effect of Congruence, Physical features and their interaction were specified as random effects. As in Experiment 1, we expected to observe an interference effect for the intact scene. Specifically, we aimed to determine which type of peripheral information contributed to influence central vision categorization. Basically, the contribution of the physical features contained in the amplitude spectrum should result in an interference effect for peripheral rings whose amplitude spectrum was preserved. Similarly, the contribution of the physical features contained in the phase spectrum should result in an interference effect for peripheral rings whose phase spectrum was preserved. If both physical information are used in peripheral vision, we should observe interference effects in both conditions. To test these hypotheses, we calculated whether the interference effects were significant in each condition of Physical features for ER and RT.

We used the method proposed by Bates, Kliegl, et al. (2015) to construct parsimonious mixed models preventing convergence problems. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained with a Wald test for ER analyses and by Satterthwaite approximation for RT analyses with the *lmerTest* package (Kuznetsova et al., 2017). The significance threshold was set at .017 after Bonferroni correction for three non-orthogonal tests ($.05/3 = .017$) for ER and RT analyses. Effect sizes were estimated using Cohen's d_z (Lakens, 2013). Trials in which saccades were initiated from more than 1.7° around the fixation cross (i.e., radius of the central disk) were discarded from the analyses. This resulted in removing 0.08% of the trials.

Data are available in the Open Science Framework repository: <https://osf.io/q3rtw/>.

Results

Results are shown in Figure 4. The analyses performed on ER revealed neither a significant interference effect (i.e., Congruence effect) for the intact scene, $\beta = 0.70$, $z = 2.38$, $p = .0171$, nor a significant interference effect for the phase-preserved scene, $\beta = 0.09$, $z = 0.27$, $p = .784$, nor a significant interference effect for the amplitude-preserved scene, $\beta = -0.25$, $z = -1$, $p = .317$.

The analyses performed on RT revealed a significant interference effect for the intact scene, $\beta = 30.73$, $t(4981.13) = 4.14$, $p < 0.001$, $d_z = 0.883$, with participants being slower when the two scenes were semantically incongruent (665 ± 35 ms) than congruent (633 ± 34 ms). We also observed a significant interference effect for the phase-preserved scene, $\beta = 19.05$, $t(4981.03) = 2.60$, $p = .009$, $d_z = 0.554$, with participants being slower when the two scenes were semantically incongruent (643 ± 37 ms) than congruent (623 ± 33 ms). However, the

interference effect was not significant for the amplitude-preserved scene, $\beta = 14.18$, $t(4981.04) = 1.93$, $p = .054$, $d_z = 0.411$.

We performed post-hoc mixed models analyses on RT to examine whether the interference effect was actually larger in one of the conditions. We calculated as fixed effect the interaction between the Congruence (congruent vs. incongruent) and the Physical features (intact vs. amplitude-preserved or phase-preserved vs. amplitude-preserved or intact vs. phase-preserved) on the RT. However, these post-hoc analyses failed to demonstrate neither a stronger interference effect for intact than phase-preserved scenes, $\beta = 11.90$, $t(3311.09) = 1.13$, $p = .257$, $d_z = 0.241$, nor a stronger interference effect for intact than amplitude-preserved scenes, $\beta = 16.43$, $t(3293.21) = 1.59$, $p = .113$, $d_z = 0.339$, nor a stronger interference effect for phase-preserved than amplitude-preserved scenes, $\beta = 4.90$, $t(3337.08) = 0.47$, $p = .637$, $d_z = 0.100$.

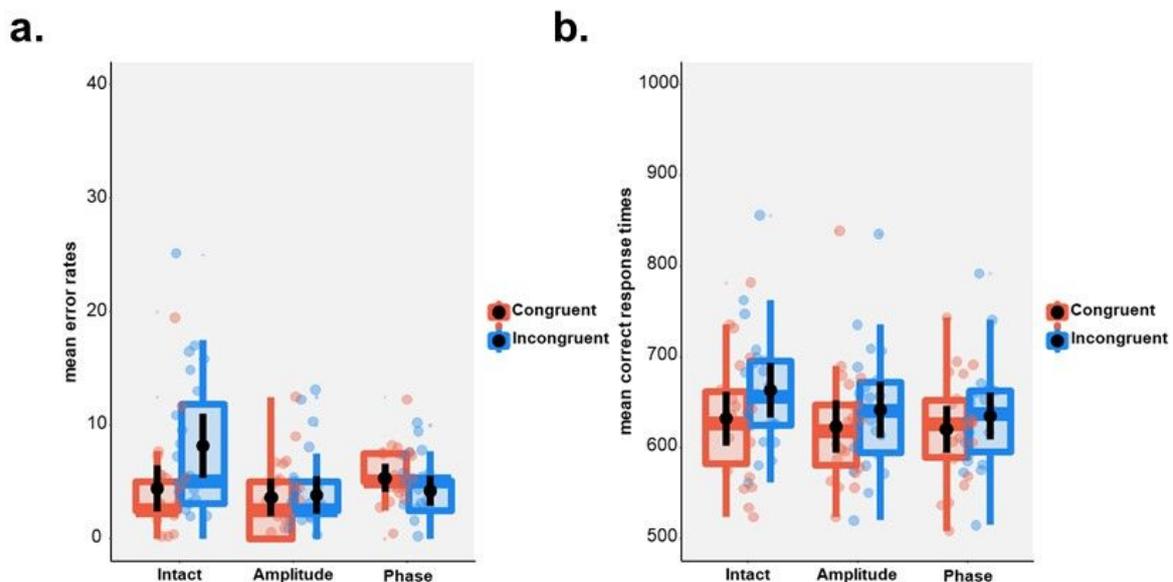


Figure 4. (a) Mean error rates in percentage and (b) mean correct response times in milliseconds during the categorization of the central disk with a peripheral ring containing an intact scene, an amplitude-preserved scene and a phase-preserved scene according to the congruence between the central disk and the peripheral ring (congruent, incongruent). Black dots and error bars indicate mean and standard error. Color dots are individual observations.

General discussion

Past studies suggest that low-resolution information from peripheral vision plays a critical role during scene categorization (Larson & Loschky, 2009; Loschky et al., 2019; Roux-Sibilon et al., 2018; Trouilloud et al., 2020) and would be useful to generate predictions which would then influence and guide the processing of detailed information from central vision (Bar, 2003; Bar, 2007; Peyrin et al., 2021; Roux-Sibilon et al., 2019). The aim of the present study was therefore to determine (1) whether peripheral information influences the categorization of

1
2
3 scenes in central vision and (2) which type of physical features from peripheral vision actually
4 guides central vision processing. To this end, we used a semantic interference paradigm in
5 which we presented simultaneously a central disk and a peripheral ring, both containing scenes
6 belonging to the same category (congruent condition) or not (incongruent condition).
7

8
9 In Experiment 1, participants had to categorize as an indoor or an outdoor scene either
10 the scene contained in the central disk while ignoring the peripheral ring or the scene contained
11 in the peripheral ring while ignoring the central disk. Participants had to fixate the centre of the
12 screen. Results showed a semantic interference effect indicating that the information to be
13 ignored was however processed and interfered with the processing of the information to attend.
14 As in the study of Lukavský (2019), this effect was present for both central and peripheral
15 categorization despite our control of the cortical magnification factor so that the central disk
16 and the peripheral ring activated the same amount of neurons in the primary visual cortex.
17 Importantly for our research issue, results demonstrated that despite its very low spatial
18 resolution, information available in peripheral vision influences and guides the processing of
19 the information that the participants fixated in central vision. This result could be thus
20 interpreted in the context of the predictive models of visual recognition (Bar, 2003, 2004, 2007;
21 Trapp & Bar, 2015; Kauffmann et al., 2014; Kveraga et al., 2007; Peyrin et al., 2010).
22 According to these models, a low resolution information is sufficient and useful to trigger
23 cortical predictive mechanisms that would then guide a more detailed visual analysis. Roux-
24 Sibilon et al. (2019) demonstrated that the low spatial resolution of peripheral information may
25 be used to generate predictions that guide the processing of central information. Similarly, in
26 our experiment, when the peripheral scene to ignore belongs to the same category as the central
27 scene, it may be used to generate valid predictions about the category of the central scene which
28 would then facilitate its categorization. On the other hand, when the peripheral scene belongs
29 to a different category than the central scene, its rapid processing would result in the generation
30 of invalid predictions, impairing the central categorization and resulting in lower categorization
31 performances.
32

33
34 In more ecological viewing conditions, the rapid processing of peripheral information
35 may also be used to generate predictions about the visual information present at the next
36 fixation. Herwig and Schneider (2014) showed that peripheral pre-saccadic input is used to
37 generate predictions about the appearance of the foveal post-saccadic input. In their study,
38 participants were presented with two peripheral objects (a triangular or circular sinusoidal
39 grating) and had to initiate a saccade toward one of them so that both objects were saccaded to
40 in an equal amount of trials. During the execution of the saccade, the spatial frequency of the
41 peripheral object could remain intact (normal condition) or change (swapped condition)
42 depending on the object that was targeted so that participants implicitly learnt associations
43 between the peripheral and foveal appearance of each object. In a subsequent “test” phase, one
44 of the two peripheral objects was presented and participants had to initiate a saccade toward it.
45 The object disappeared during the saccade and was replaced by a new object with a random
46 spatial frequency. Participants had to adjust the spatial frequency of the object (which was now
47 at fovea) in order to match the expected spatial frequency of the object they saw peripherally
48 before initiating their saccade. Results revealed participants' judgments were influenced by the
49 acquisition phase. Peripheral objects with the same spatial frequency were expected to have a
50 different spatial frequency when foveated depending on the association learnt during the
51
52
53
54
55
56
57
58
59
60

1
2
3 acquisition phase. This therefore indicated that associations between pre- and post-saccadic
4 appearance of visual stimuli are used to predict object features across saccades. Several studies
5 have also shown that peripheral information is used before a change in gaze location, allowing
6 to initiate a saccadic movement towards a relevant area of the scene (Cajar et al., 2020;
7 Nuthmann, 2014). Indeed, Nuthmann (2014) examined the importance of central and
8 peripheral vision during object search in a scene by restricting the visual field to its central part
9 using gaze-contingent spotlights of different eccentricities (i.e., masking either peripheral or
10 both peripheral and parafoveal vision) or its peripheral parts using blindspots of different
11 eccentricities (i.e., masking either foveal or both foveal and parafoveal vision). Results showed
12 that when participants searched the object with only high resolution information from foveal
13 vision, their accuracy and search times were significantly impaired relative to the control
14 condition where all scene information was available. However, when peripheral information
15 was available and foveal information was hidden, their search performances were as good as
16 when all the image content was displayed. Interestingly, the authors decomposed the scan
17 pattern of participants to identify which part of the visual field was useful to (1) localize and
18 (2) identify the object. In the blindspots condition (i.e., masking central parts), participants
19 were not impaired to localize the object but were slower to identify it. In the spotlights
20 condition (i.e., masking peripheral parts), the opposite results were found. Participants were
21 not impaired to identify the object but they needed more time to first localize it. These findings
22 therefore support the critical utility of peripheral vision relative to central vision to generate
23 saccades towards a relevant part of the scene and localize objects of interest. Overall, these
24 findings highlight the crucial role of peripheral vision in updating the gaze location to localize
25 an object and predict its features before its identification thanks to the high resolution of central
26 vision. Despite its low resolution, peripheral vision therefore appears essential to extract
27 enough information to quickly categorize objects or scenes and guide saccadic eye movements
28 to bring areas of interest into central vision, allowing a more detailed processing.

29
30
31 In this context, we formulated the hypothesis that the semantic interference of
32 peripheral vision would be stronger than the semantic interference of central vision. We failed
33 to observe a significant interaction between the congruence and the spatial position of the
34 information to be ignored. A posteriori, our hypothesis seems daring with regard to
35 methodological constraints of our experiment and natural vision habits. As we previously
36 pointed out, our task required participants to fixate the centre of the screen to avoid saccadic
37 movements throughout the experiment. This experimental instruction induced attentional
38 constraint to the task. While our attention is distributed over the entire visual field in normal
39 viewing conditions, participants here were submitted to focused attentional constraints. Brand
40 and Johnson (2018) investigated whether scene categorization required distributed attention
41 over the entire visual field or whether focused attention in central vision was sufficient. They
42 presented a large scene and they superimposed two rectangles, one at the centre of the screen
43 and the other surrounding the periphery of the scene. The task consisted in categorizing the
44 scene as natural or man-made and then in indicating the orientation of either the central
45 rectangle (i.e., focused attention) or the peripheral one (i.e., distributed attention). The results
46 revealed better scene categorization performances in the distributed than focused attention
47 condition, suggesting that distributed attention facilitated scene categorization at the
48 superordinate level. In other words, attention would be naturally distributed on the whole visual
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 field when perceiving a scene, and information from the whole visual field would be used to
4 categorize it at a superordinate level (as in our experiment). It is thus possible that our task
5 involves a central attentional bias which increases the weight of the interference from the centre
6 on the peripheral categorization and conversely, underestimates the weight of the interference
7 from the periphery on the central categorization. In addition, central information is intrinsically
8 the one with the highest spatial resolution and previous studies revealed that it is more difficult
9 to ignore HSF than LSF in central vision (Kauffmann et al., 2017). Besides, irrespective of the
10 interference effect, we observed that participants were faster and categorized the central disk
11 more accurately than the peripheral ring, potentially masking a greater interference effect of
12 peripheral vision. Future studies therefore need to ensure that attention is distributed over the
13 whole visual field.
14

15
16
17 To go further, we aimed at determining the nature of physical features from peripheral
18 vision which influence the processing of central vision information during scene
19 categorization. To this end, we performed a second experiment (Experiment 2) in which
20 participants were only requested to categorize the central disk while ignoring the peripheral
21 ring and we manipulated the physical features of the peripheral ring. The scene contained in
22 the peripheral ring was therefore either intact (i.e., containing semantic information), only its
23 amplitude spectrum was preserved (amplitude-preserved, i.e. containing only the distribution
24 of information across spatial frequencies and orientations), or only its phase spectrum was
25 preserved (phase-preserved, i.e. containing only information about the spatial structure of the
26 scene). If information from the phase spectrum influences central vision processing, we
27 expected an interference effect in the phase-preserved condition. If information from the
28 amplitude spectrum influences central vision processing, we expected an interference effect in
29 the amplitude-preserved condition. As in Experiment 1, we observed the semantic interference
30 effect on the response times when the peripheral scene was intact. Response times were longer
31 when the peripheral scene was incongruent than congruent with the central scene, suggesting
32 again that the peripheral input of low-resolution, even irrelevant, could not be ignored and was
33 integrated to the processing of central information. Unlike Experiment 1, we cannot however
34 conclude on an interference effect on error rates, the congruence effect being not significant
35 after the Bonferroni correction. It should however be noted that the very low error rates in
36 Experiment 2 (below 8%) suggest a ceiling effect. The categorization of a central disk may be
37 too easy to detect the expected effect on error rates. Importantly, the interference effect of
38 peripheral rings was also observed when the phase information of the peripheral ring was
39 preserved, but not when only the amplitude information was preserved. The incongruent
40 information contained in the phase spectrum, unlike amplitude spectrum information, delayed
41 the response as much as the information contained in the intact scene. In the same way, Roux-
42 Sibilon et al. (2019) observed that the interference effect of a scene background on the
43 categorization of an object in central vision disappeared when the amplitude spectrum of the
44 scene background was preserved, but the spatial configuration was altered (phase scrambling
45 of scene images). Altogether, these results suggest that semantic and spatially-arranged features
46 contained in the phase spectrum of a scene in peripheral vision are primordial to guide the
47 processing of information in central vision. It should be noted that the absence of interference
48 effect for amplitude-preserved peripheral scenes does not necessarily imply that amplitude
49 information is not used during scene categorization. Indeed, many studies have shown that
50
51
52
53
54
55
56
57
58
59
60

1
2
3 although this information alone may not be sufficient (Loschky & Larson, 2008; Loschky et
4 al., 2007) statistical regularities from the amplitude spectrum of scenes from different
5 categories can be exploited for rapid and coarse scene categorization (Field, 1987; Guyader et
6 al., 2004; Oliva & Torralba, 2001; Torralba & Oliva, 2003). However, our results suggest that
7 such information may be less relevant than phase information to guide the processing in central
8 vision.
9

10
11 To conclude, these two experiments show mutual influence between central and
12 peripheral information during rapid scene categorization. Despite the low-resolution from
13 peripheral content, the influence of peripheral information on central categorization is as
14 important as the influence of detailed information from central vision which could not be
15 ignored. Although irrelevant for the task, low resolution in peripheral vision would be
16 automatically processed and would interfere with central information processing, arguing in
17 favor of the predictive role of peripheral vision to guide the analysis of central vision content.
18 More importantly, these predictions would be based on the spatial configuration of the scene
19 contained in the phase spectrum. Unfortunately, the present study did not allow us to
20 disentangle the role of peripheral vision between an evidence accumulation model and a
21 proactive model. As previously performed by Roux-Sibilon et al. (2019), future studies could
22 manipulate the time during which peripheral information is processed before central
23 information onset in order to enhance predictive processes. An increasing interference effect
24 with SOA would support a predictive model rather than the simultaneous processing of central
25 and peripheral information.
26
27
28
29
30
31
32
33

34 **Disclosure statement**

35 No potential conflict of interest was reported by the authors.
36
37
38

39 **Funding**

40 This work was supported by NeuroCoG IDEX UGA in the framework of the “Investissements
41 d’avenir” program (ANR-15-IDEX-02).
42
43
44
45

46 **References**

- 47 Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object
48 Recognition. *Journal of Cognitive Neuroscience*, 15(4), 600-609.
49 <https://doi.org/10.1162/089892903321662976>
50
51 Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions.
52 *Trends in Cognitive Sciences*, 11(7), 280-289.
53 <https://doi.org/10.1016/j.tics.2007.05.005>
54
55 Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hamalainen,
56 M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down
57 facilitation of visual recognition. *Proceedings of the National Academy of Sciences*,
58 103(2), 449-454. <https://doi.org/10.1073/pnas.0507062103>
59
60

- 1
2
3 Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- 4
5
6 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- 7
8
9 Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research, 86*, 35-42. <https://doi.org/10.1016/j.visres.2013.04.006>
- 10
11
12 Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision, 10*(4), 433-436.
- 13
14
15 Brand, J., & Johnson, A. P. (2018). The effects of distributed and focused attention on rapid scene categorization. *Visual Cognition, 26*(6), 450-462. <https://doi.org/10.1080/13506285.2018.1485808>
- 16
17
18
19 Cajar, A., Engbert, R., & Laubrock, J. (2020). How spatial frequencies and color drive object search in real-world scenes: A new eye-movement corpus. *Journal of Vision, 20*(7), 8. <https://doi.org/10.1167/jov.20.7.8>
- 20
21
22
23 Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of comparative Neurology, 300*(1), 5-25.
- 24
25
26 Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of comparative neurology, 292*(4), 497-523.
- 27
28
29 Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology, 159*(2), 203-221.
- 30
31
32 Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition, 35*(3), 393-401. <https://doi.org/10.3758/BF03193280>
- 33
34
35 Davenport, J. L., & Potter, M. C. (2004). Scene Consistency in Object and Background Perception. *Psychological Science, 15*(8), 559-564. <https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- 36
37
38 Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a, 4*(12), 2379-2394.
- 39
40
41 Geuzebroek, A. C., & van den Berg, A. V. (2018). Eccentricity scale independence for scene perception in the first tens of milliseconds. *Journal of Vision, 18*(9), 9. <https://doi.org/10.1167/18.9.9>
- 42
43
44 Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience, 30*(1), 535-574. <https://doi.org/10.1146/annurev.neuro.29.051605.113038>
- 45
46
47
48 Guyader, N., Chauvin, A., Peyrin, C., Héroult, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies, 327*(4), 313-318.
- 49
50
51
52 Herwig, A., & Schneider, W. X. (2014). Predicting object features across saccades: Evidence from object recognition and visual search. *Journal of Experimental Psychology: General, 143*(5), 1903-1922. <https://doi.org/10.1037/a0036781>
- 53
54
55
56 Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision, 8*(13), 11-11. <https://doi.org/10.1167/8.13.11>
- 57
58
59
60

- 1
2
3 Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene
4 context: Fast categorization and object interference. *Vision Research*, *47*(26),
5 3286-3297. <https://doi.org/10.1016/j.visres.2007.09.013>
6
7 Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The Neural Bases of the
8 Semantic Interference of Spatial Frequency-based Information in Scenes. *Journal of*
9 *Cognitive Neuroscience*, *27*(12), 2394-2405. https://doi.org/10.1162/jocn_a_00861
10
11 Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the
12 neural network underlying coarse-to-fine categorization of visual scenes. A dynamic
13 causal modeling study. *Brain and Cognition*, *99*, 46-56.
14 <https://doi.org/10.1016/j.bandc.2015.07.004>
15
16 Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency
17 processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*.
18 <https://doi.org/10.3389/fnint.2014.00037>
19
20 Kauffmann, L., Roux-Sibilon, A., Beffara, B., Mermillod, M., Guyader, N., & Peyrin, C.
21 (2017). How does information from low and high spatial frequencies interact during
22 scene categorization? *Visual Cognition*, *25*(9-10), 853-867.
23 <https://doi.org/10.1080/13506285.2017.1347590>
24
25 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in
26 linear mixed effects models. *Journal of statistical software*, *82*(1), 1-26.
27
28 Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-
29 down facilitation in recognition. *Journal of Neuroscience*, *27*(48), 13232-13240.
30
31 Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A
32 practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*.
33 <https://doi.org/10.3389/fpsyg.2013.00863>
34
35 Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal
36 dynamics of scene gist recognition. *Journal of Experimental Psychology: Human*
37 *Perception and Performance*, *40*(2), 471-487. <https://doi.org/10.1037/a0034986>
38
39 Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision
40 to scene gist recognition. *Journal of Vision*, *9*(10), 6-6. <https://doi.org/10.1167/9.10.6>
41
42 Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene
43 categorization, including the natural/man-made distinction. *Journal of Vision*, *8*(1), 4-4.
44
45 Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before
46 basic-level distinctions in scene gist processing. *Visual Cognition*, *18*(4), 513-536.
47
48 Loschky, L. C., Nuthmann, A., Fortenbaugh, F. C., & Levi, D. M. (2017). Scene perception
49 from central to peripheral vision. *Journal of Vision*, *17*(1), 6.
50 <https://doi.org/10.1167/17.1.6>
51
52 Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007).
53 The importance of information localization in scene gist recognition. *Journal of*
54 *Experimental Psychology: Human Perception and Performance*, *33*(6), 1431.
55
56 Loschky, L. C., Szaffarczyk, S., Beugnet, C., Young, M. E., & Boucart, M. (2019). The
57 contributions of central and peripheral vision to scene- gist recognition with a 1808
58 visual field. *Journal of Vision*, *21*.
59
60 Lukavský, J. (2019). Scene categorization in the presence of a distractor. *Journal of Vision*,
19(2), 6. <https://doi.org/10.1167/19.2.6>

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Mirolli, M., Ferrauto, T., & Nolfi, S. (2010). Categorisation through evidence accumulation in an active vision system. *Connection Science*, 22(4), 331-354. <https://doi.org/10.1080/09540091.2010.505976>
- Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 342.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3), 145-175.
- Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10, 437-442.
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., Marendaz, C., & Vuilleumier, P. (2010). The Neural Substrates and Timing of Top-Down Processes during Coarse-to-Fine Categorization of Visual Scenes: A Combined fMRI and ERP Study. *Journal of Cognitive Neuroscience*, 22(12), 2768-2780. <https://doi.org/10.1162/jocn.2010.21424>
- Peyrin, C., Roux-Sibilon, A., Trouilloud, A., Khazaz, S., Joly, M., Pichat, C., Boucart, M., Krainik, A., & Kauffmann, L. (2021). Semantic and Physical Properties of Peripheral Vision Are Used for Scene Categorization in Central Vision. *Journal of Cognitive Neuroscience*, 33(5), 799-813. https://doi.org/10.1162/jocn_a_01689
- Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, 12(6), 852-877.
- Roux-Sibilon, A., Trouilloud, A., Kauffmann, L., Guyader, N., Mermillod, M., & Peyrin, C. (2019). Influence of peripheral vision on object categorization in central vision. *Journal of Vision*, 19(14), 7. <https://doi.org/10.1167/19.14.7>
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, 14(3), 391-412. <https://doi.org/10.1088/0954-898X/14/3/302>
- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition: Predictions and competition. *Annals of the New York Academy of Sciences*, 1339(1), 190-198. <https://doi.org/10.1111/nyas.12680>
- Trouilloud, A., Kauffmann, L., Roux-Sibilon, A., Rossel, P., Boucart, M., Mermillod, M., & Peyrin, C. (2020). Rapid scene categorization: From coarse peripheral vision to fine central vision. *Vision Research*, 170, 60-72. <https://doi.org/10.1016/j.visres.2020.02.008>
- Wu, J., Yan, T., Zhang, Z., Jin, F., & Guo, Q. (2012). Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging. *Human brain mapping*, 33(7), 1727-1740.
- Yamada, E. (1969). Some structural features of the fovea centralis in the human retina. *Archives of ophthalmology*, 82(2), 151-159.

Bibliographie

- Alpern, M. (1952). Metacontrast; historical introduction. *American Journal of Optometry*.
- Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic Organization of Human Ventral Visual Cortex. *Journal of Neuroscience*, 29(34), 10638-10652.
<https://doi.org/10.1523/JNEUROSCI.2807-09.2009>
- 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(5), 617-629.
<https://doi.org/10.1068/p190617>
- Baizer, J. S., Ungerleider, L. G., & Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *Journal of Neuroscience*, 11(1), 168-190.
<https://doi.org/10.1523/JNEUROSCI.11-01-00168.1991>
- Baldassano, C., Fei-Fei, L., & Beck, D. M. (2016). Pinpointing the peripheral bias in neural scene-processing networks during natural viewing. *Journal of Vision*, 16(2), 9.
<https://doi.org/10.1167/16.2.9>
- Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, 15(4), 600-609.
<https://doi.org/10.1162/089892903321662976>
- Bar, M. (2007). The proactive brain : Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280-289. <https://doi.org/10.1016/j.tics.2007.05.005>
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hamalainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2), 449-454.
<https://doi.org/10.1073/pnas.0507062103>
- Bar, M., & Ullman, S. (1996). Spatial Context in Recognition. *Perception*, 25(3), 343-352.
<https://doi.org/10.1068/p250343>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Bechetolle, A., Arnould, B., Bron, A., Baudouin, C., Renard, J.-P., Sellem, E., Brouquet, Y., Denis, P., Nordmann, J.-P., & Rigeade, M.-C. (2008). Measurement of health-related quality of life with glaucoma : Validation of the Glau-QoL\copyright 36-item questionnaire. *Acta ophthalmologica*, 86(1), 71-80.
- Benson, D. F., & Greenberg, J. P. (1969). Visual Form Agnosia : A Specific Defect in Visual Discrimination. *Archives of Neurology*, 20(1), 82-89.
<https://doi.org/10.1001/archneur.1969.00480070092010>
- 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. *The Journal of Physiology*, 203(1), 237-260.1.
- Bonnar, L., Gosselin, F., & Schyns, P. G. (2002). Understanding Dali's Slave Market with the Disappearing Bust of Voltaire : A Case Study in the Scale Information Driving Perception. *Perception*, 31(6), 683-691. <https://doi.org/10.1068/p3276>
- Bonnet, C. (1986). La théorie de la détection du signal. *Manuel Pratique de Psychophysique*, 59-86.
- Boucard, C. C., Hernowo, A. T., Maguire, R. P., Jansonius, N. M., Roerdink, J. B. T. M., Hooymans, J. M. M., & Cornelissen, F. W. (2009). Changes in cortical grey matter density associated with long-standing retinal visual field defects. *Brain*, 132(7), 1898-1906.
<https://doi.org/10.1093/brain/awp119>

- Boucart, M., Dinon, J.-F., Desprez, P., Desmettre, T., Hladiuk, K., & Oliva, A. (2008). Recognition of facial emotion in low vision : A flexible usage of facial features. *Visual neuroscience*, *25*(4), 603-609.
- Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, *86*, 35-42.
<https://doi.org/10.1016/j.visres.2013.04.006>
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*(5241), 177-178.
- Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.
- Brand, J., & Johnson, A. P. (2018). The effects of distributed and focused attention on rapid scene categorization. *Visual Cognition*, *26*(6), 450-462.
<https://doi.org/10.1080/13506285.2018.1485808>
- Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research*, *15*(12), 1411-1412.
[https://doi.org/10.1016/0042-6989\(75\)90200-X](https://doi.org/10.1016/0042-6989(75)90200-X)
- Breitmeyer, B. G. (2014). Contributions of magno- and parvocellular channels to conscious and non-conscious vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641), 20130213. <https://doi.org/10.1098/rstb.2013.0213>
- Breitmeyer, B. G., Kafaligönül, H., Ögmen, H., Mardon, L., Todd, S., & Ziegler, R. (2006). Meta-and paracontrast reveal differences between contour-and brightness-processing mechanisms. *Vision research*, *46*(17), 2645-2658.
- Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking : A comparison, review, and update. *Perception & psychophysics*, *62*(8), 1572-1595.
- Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, *7*(8), 880-886. <https://doi.org/10.1038/nn1278>
- Broca, P. (1865). *Instructions générale pour les recherches anthropologiques:(Anatomie et physiologie.)*. V. Masson et fils.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2-3), 96-107.
[https://doi.org/10.1016/S0165-0173\(01\)00085-6](https://doi.org/10.1016/S0165-0173(01)00085-6)
- Bullier, J., Schall, J. D., & Morel, A. (1996). Functional streams in occipito-frontal connections in the monkey. *Behavioural Brain Research*, *76*(1), 89-97. [https://doi.org/10.1016/0166-4328\(95\)00182-4](https://doi.org/10.1016/0166-4328(95)00182-4)
- Cajar, A., Engbert, R., & Laubrock, J. (2020). How spatial frequencies and color drive object search in real-world scenes : A new eye-movement corpus. *Journal of Vision*, *20*(7), 8.
<https://doi.org/10.1167/jov.20.7.8>
- Campagne, A., Fradcourt, B., Pichat, C., Baciú, M., Kauffmann, L., & Peyrin, C. (2016). Cerebral correlates of emotional and action appraisals during visual processing of emotional scenes depending on spatial frequency : A pilot study. *Plos one*, *11*(1), e0144393.
- Campbell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, *197*(3), 551-566. <https://doi.org/10.1113/jphysiol.1968.sp008574>
- Caplette, L., West, G., Gomot, M., Gosselin, F., & Wicker, B. (2014). Affective and contextual values modulate spatial frequency use in object recognition. *Frontiers in psychology*, *5*, 512.
- Cesarei, A. D., Mastroia, S., & Codispoti, M. (2013). Early Spatial Frequency Processing of Natural Images : An ERP Study. *PLOS ONE*, *8*(5), e65103.
<https://doi.org/10.1371/journal.pone.0065103>
- Chen, W. W., Wang, N., Cai, S., Fang, Z., Yu, M., Wu, Q., Tang, L., Guo, B., Feng, Y., Jonas, J. B., Chen, X., Liu, X., & Gong, Q. (2013). Structural Brain Abnormalities in Patients with Primary Open-Angle Glaucoma : A Study with 3T MR Imaging. *Investigative Ophthalmology & Visual Science*, *54*(1), 545. <https://doi.org/10.1167/iovs.12-9893>
- Collin, C. A. (2006). Spatial-frequency thresholds for object categorisation at basic and subordinate levels. *Perception*, *35*(1), 41-52.

- Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Perception & psychophysics*, 67(2), 354-364.
- Corveleyn, X., Lenoble, Q., Rouland, J.-F., & Boucart, M. (2020). What is the nature of the reach-and-grasp deficit in glaucoma? *Journal of Glaucoma*, 29(9), 799-806.
- Crabb, D. P., Smith, N. D., Glen, F. C., Burton, R., & Garway-Heath, D. F. (2013). How does glaucoma look? : Patient perception of visual field loss. *Ophthalmology*, 120(6), 1120-1126.
- Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of comparative Neurology*, 300(1), 5-25.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of comparative neurology*, 292(4), 497-523.
- Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology*, 159(2), 203-221.
- Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, 35(3), 393-401. <https://doi.org/10.3758/BF03193280>
- Davenport, J. L., & Potter, M. C. (2004). Scene Consistency in Object and Background Perception. *Psychological Science*, 15(8), 559-564. <https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- David, E., Beitner, J., & Vö, M. L.-H. (2020). *Effects of Transient Loss of Vision on Head and Eye Movements during Visual Search in a Virtual Environment*. 26.
- David, S. V., Hayden, B. Y., & Gallant, J. L. (2006). Spectral Receptive Field Properties Explain Shape Selectivity in Area V4. *Journal of Neurophysiology*, 96(6), 3492-3505. <https://doi.org/10.1152/jn.00575.2006>
- De Cesarei, A., & Codispoti, M. (2011). Scene Identification and Emotional Response : Which Spatial Frequencies Are Critical? *Journal of Neuroscience*, 31(47), 17052-17057. <https://doi.org/10.1523/JNEUROSCI.3745-11.2011>
- De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, 22(5), 545-559.
- De Valois, R. L., & De Valois, K. K. (1980). Spatial vision. *Annual review of psychology*, 31(1), 309-341.
- De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision research*, 22(5), 531-544.
- de Haan, E. H., & Cowey, A. (2011). On the usefulness of 'what' and 'where' pathways in vision. *Trends in cognitive sciences*, 15(10), 460-466.
- Dive, S., Rouland, J. F., Lenoble, Q., Szaffarczyk, S., McKendrick, A. M., & Boucart, M. (2016). Impact of peripheral field loss on the execution of natural actions : A study with glaucomatous patients and normally sighted people. *Journal of glaucoma*, 25(10), e889-e896.
- Dubey, S., Bedi, H., Bedi, M., Matah, P., Sahu, J., Mukherjee, S., & Chauhan, L. (2019). Impact of visual impairment on the wellbeing and functional disability of patients with glaucoma in India. *Journal of Current Ophthalmology*.
- Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38(4), 659-671.
- Duncan, R. O., Sample, P. A., Weinreb, R. N., Bowd, C., & Zangwill, L. M. (2007). Retinotopic organization of primary visual cortex in glaucoma : Comparing fMRI measurements of cortical function with visual field loss. *Progress in Retinal and Eye Research*, 26(1), 38-56. <https://doi.org/10.1016/j.preteyeres.2006.10.001>
- Elliott, D. B., Yang, K. C., & Whitaker, D. (1995). Visual acuity changes throughout adulthood in normal, healthy eyes : Seeing beyond 6/6. *Optometry and vision science: official publication of the American Academy of Optometry*, 72(3), 186-191.
- 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 (New York, NY: 1991)*, 7(2), 181-192.

- Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E.-J., & Shadlen, M. N. (1994). FMRI of human visual cortex. *Nature*.
- Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in cognitive sciences*, 4(9), 345-352.
- Enroth-Cugell, C., & Robson, J. G. (s. d.). *Functional Characteristics and Diversity of Cat Retinal Ganglion Cells*. 18.
- Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, 187(3), 517-552. <https://doi.org/10.1113/jphysiol.1966.sp008107>
- Entzman, L., Guyader, N., Kauffmann, L., Lenouvel, J., Charles, C., Peyrin, C., Vuillaume, R., & Mermillod, M. (in press). The role of emotional content and perceptual saliency during the programming of saccades toward faces. *Cognitive Science*.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598-601. <https://doi.org/10.1038/33402>
- Everson, R. M., Prashanth, A. K., Gabbay, M., Knight, B. W., Sirovich, L., & Kaplan, E. (1998). Representation of spatial frequency and orientation in the visual cortex. *Proceedings of the National Academy of Sciences*, 95(14), 8334-8338.
- Fechtner, R. D., & Weinreb, R. N. (1994). Mechanisms of optic nerve damage in primary open angle glaucoma. *Survey of ophthalmology*, 39(1), 23-42.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y., 1(1))*, 1-47. <https://doi.org/10.1093/cercor/1.1.1-a>
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, 4(12), 2379-2394.
- Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, 16(3), 553-567.
- Foster, K. H., Gaska, J. P., Nagler, M., & Pollen, D. A. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *The Journal of physiology*, 365(1), 331-363.
- Frezzotti, P., Giorgio, A., Motolese, I., De Leucio, A., Iester, M., Motolese, E., Federico, A., & De Stefano, N. (2014). Structural and Functional Brain Changes beyond Visual System in Patients with Advanced Glaucoma. *PLoS ONE*, 9(8), e105931. <https://doi.org/10.1371/journal.pone.0105931>
- Frezzotti, P., Giorgio, A., Toto, F., De Leucio, A., & De Stefano, N. (2016). Early changes of brain connectivity in primary open angle glaucoma : Early Connectivity Changes in Glaucoma. *Human Brain Mapping*, 37(12), 4581-4596. <https://doi.org/10.1002/hbm.23330>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815-836. <https://doi.org/10.1098/rstb.2005.1622>
- Geschwind, N., & Galaburda, A. M. (1984). *Cerebral dominance : The biological foundations*. Harvard University Press.
- Geschwind, N., & Levitsky, W. (1968). Human brain : Left-right asymmetries in temporal speech region. *Science*, 161(3837), 186-187.
- Geuzebroek, A. C., & van den Berg, A. V. (2018). Eccentricity scale independence for scene perception in the first tens of milliseconds. *Journal of Vision*, 18(9), 9. <https://doi.org/10.1167/18.9.9>
- Gittings, N. S., & Fozard, J. L. (1986). Age related changes in visual acuity. *Experimental Gerontology*, 21(4-5), 423-433. [https://doi.org/10.1016/0531-5565\(86\)90047-1](https://doi.org/10.1016/0531-5565(86)90047-1)
- Glen, F. C., & Crabb, D. P. (2015). Living with glaucoma : A qualitative study of functional implications and patients' coping behaviours. *BMC ophthalmology*, 15(1), 1-15.
- Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, 30(1), 535-574. <https://doi.org/10.1146/annurev.neuro.29.051605.113038>
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1), 20-25.

- Grainger, J., Tydgat, I., & Issel , J. (2010). Crowding affects letters and symbols differently. *Journal of Experimental Psychology: Human Perception and Performance*, 36(3), 673.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10-11), 1409-1422. [https://doi.org/10.1016/S0042-6989\(01\)00073-6](https://doi.org/10.1016/S0042-6989(01)00073-6)
- Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual Receptive Fields of Neurons in Inferotemporal Cortex of the Monkey. *Science*, 166(3910), 1303-1306. <https://doi.org/10.1126/science.166.3910.1303>
- Gu rin-Dugu , A., & Oliva, A. (2000). Classification of scene photographs from local orientations features. *Pattern Recognition Letters*, 21(13-14), 1135-1140.
- Gupta, N. (s. d.). *Glaucoma as a neurodegenerative disease*. 5.
- Gutierrez, P., Wilson, M. R., Johnson, C., Gordon, M., Cioffi, G. A., Ritch, R., Sherwood, M., Meng, K., & Mangione, C. M. (1997). Influence of glaucomatous visual field loss on health-related quality of life. *Archives of ophthalmology*, 115(6), 777-784.
- Guyader, N., Chauvin, A., Boucart, M., & Peyrin, C. (2017). Do low spatial frequencies explain the extremely fast saccades towards human faces? *Vision research*, 133, 100-111.
- Guyader, N., Chauvin, A., Peyrin, C., H rault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, 327(4), 313-318.
- Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization : Implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1264.
- Harwerth, R. S., & Levi, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision research*, 18(11), 1579-1586.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity Bias as an Organizing Principle for Human High-Order Object Areas. *Neuron*, 34(3), 479-490. [https://doi.org/10.1016/S0896-6273\(02\)00662-1](https://doi.org/10.1016/S0896-6273(02)00662-1)
- Hausmann, M., Slabbekoorn, D., Van Goozen, S. H., Cohen-Kettenis, P. T., & G nt rk n, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral neuroscience*, 114(6), 1245.
- Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch, P., Schapiro, M. B., & Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, 88(5), 1621-1625. <https://doi.org/10.1073/pnas.88.5.1621>
- Henriksson, L., Nurminen, L., Hyv rinen, A., & Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *Journal of Vision*, 8(10), 5-5.
- H rault, J., Oliva, A., & Guerin-Dugue, A. (1997). Scene categorisation by curvilinear component analysis of low frequency spectra. *ESANN'97: European symposium on artificial neural networks (Bruges, 16-18 April 1997)*, 91-96.
- Herwig, A., & Schneider, W. X. (2014). Predicting object features across saccades : Evidence from object recognition and visual search. *Journal of Experimental Psychology: General*, 143(5), 1903-1922. <https://doi.org/10.1037/a0036781>
- Hochstein, S., & Ahissar, M. (2002). View from the Top : Hierarchies and Reverse Hierarchies in the Visual System. *Neuron*, 36(5), 791-804. [https://doi.org/10.1016/S0896-6273\(02\)01091-7](https://doi.org/10.1016/S0896-6273(02)01091-7)
- Hodapp, E., Parrish, R. K., & Anderson, D. R. (1993). *Clinical decisions in glaucoma*. Mosby Incorporated.
- Ho-Phuoc, T., Guyader, N., & Gu rin-Dugu , A. (2010). A functional and statistical bottom-up saliency model to reveal the relative contributions of low-level visual guiding factors. *Cognitive Computation*, 2(4), 344-359.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3), 574-591.

- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106-154.2.
- 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*(2), 196-204.
- Issa, N. P., Trepel, C., & Stryker, M. P. (2000). Spatial frequency maps in cat visual cortex. *Journal of Neuroscience*, *20*(22), 8504-8514.
- Jian-Yu, E., Mihailovic, A., Schrack, J. A., Garzon, C., Li, T., Friedman, D. S., West, S. K., Gitlin, L. N., & Ramulu, P. Y. (2021). Longitudinal changes in daily patterns of objectively measured physical activity after falls in older adults with varying degrees of glaucoma. *EClinicalMedicine*, *40*, 101097.
- Johnson, J. (s. d.). *Our Peripheral Vision is Poor*. 13.
- Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of vision*, *3*(7), 4-4.
- Joubert, O. R., Fize, D., Rousset, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13), 11-11. <https://doi.org/10.1167/8.13.11>
- Joubert, O. R., Rousset, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context : Fast categorization and object interference. *Vision research*, *47*(26), 3286-3297.
- Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of vision*, *12*(13), 16-16.
- Kaplan, E. (2004). The M, P, and K pathways of the primate visual system. *The visual neurosciences*, *1*, 481-493.
- Kaplan, E., & Shapley, R. M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences*, *83*(8), 2755-2757.
- Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The Neural Bases of the Semantic Interference of Spatial Frequency-based Information in Scenes. *Journal of Cognitive Neuroscience*, *27*(12), 2394-2405. https://doi.org/10.1162/jocn_a_00861
- 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. <https://doi.org/10.1016/j.visres.2014.11.013>
- Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A dynamic causal modeling study. *Brain and Cognition*, *99*, 46-56. <https://doi.org/10.1016/j.bandc.2015.07.004>
- Kauffmann, L., Khazaz, S., Peyrin, C., & Guyader, N. (2021). Isolated face features are sufficient to elicit ultra-rapid and involuntary orienting responses toward faces. *Journal of Vision*, *21*(2), 4-4.
- Kauffmann, L., Peyrin, C., Chauvin, A., Entzmann, L., Breuil, C., & Guyader, N. (2019). Face perception influences the programming of eye movements. *Scientific reports*, *9*(1), 1-14.
- Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*. <https://doi.org/10.3389/fnint.2014.00037>
- Kauffmann, L., Roux-Sibilon, A., Beffara, B., Mermillod, M., Guyader, N., & Peyrin, C. (2017). How does information from low and high spatial frequencies interact during scene categorization? *Visual Cognition*, *25*(9-10), 853-867. <https://doi.org/10.1080/13506285.2017.1347590>
- Kimchi, R., & Pirkner, Y. (2015). Multiple level crowding : Crowding at the object parts level and at the object configural level. *Perception*, *44*(11), 1275-1292.
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*(2), 145-168. <https://doi.org/10.1016/j.bandc.2007.06.007>

- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science : A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lam, A. K., To, E., Weinreb, R. N., Yu, M., Mak, H., Lai, G., Chiu, V., Wu, K., Zhang, X., & Cheng, T. P. (2020). Use of virtual reality simulation to identify vision-related disability in patients with glaucoma. *JAMA ophthalmology*, 138(5), 490-498.
- Lamb, M. R., & Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception & Psychophysics*, 54(6), 773-784. <https://doi.org/10.3758/BF03211802>
- Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 471-487. <https://doi.org/10.1037/a0034986>
- Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, 9(10), 6-6. <https://doi.org/10.1167/9.10.6>
- Laycock, R., Crewther, D. P., Fitzgerald, P. B., & Crewther, S. G. (2007). Evidence for fast signals and later processing in human V1/V2 and V5/MT+ : A TMS study of motion perception. *Journal of Neurophysiology*, 98(3), 1253-1262.
- Lee, B. L., Gutierrez, P., Gordon, M., Wilson, M. R., Cioffi, G. A., Ritch, R., Sherwood, M., & Mangione, C. M. (1998). The Glaucoma Symptom Scale : A brief index of glaucoma-specific symptoms. *Archives of ophthalmology*, 116(7), 861-866.
- Lenoble, Q., Corveleyn, X., Tran, T. H. C., Rouland, J.-F., & Boucart, M. (2019). Can I reach it? A study in age-related macular degeneration and glaucoma patients. *Visual Cognition*, 27(9-10), 732-739.
- Lenoble, Q., Lek, J. J., & McKendrick, A. M. (2016). Visual object categorisation in people with glaucoma. *British Journal of Ophthalmology*, 100(11), 1585-1590. <https://doi.org/10.1136/bjophthalmol-2015-308289>
- Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision : Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of vision*, 2(2), 3-3.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, 4(5), 533-539. <https://doi.org/10.1038/87490>
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth : Anatomy, physiology, and perception. *Science*, 240(4853), 740-749. <https://doi.org/10.1126/science.3283936>
- Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the natural/man-made distinction. *Journal of Vision*, 8(1), 4-4.
- Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, 18(4), 513-536.
- Loschky, L. C., Nuthmann, A., Fortenbaugh, F. C., & Levi, D. M. (2017). Scene perception from central to peripheral vision. *Journal of Vision*, 17(1), 6. <https://doi.org/10.1167/17.1.6>
- Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbelle, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1431.
- Loschky, L. C., Szaffarczyk, S., Beugnet, C., Young, M. E., & Boucart, M. (2019). The contributions of central and peripheral vision to scene- gist recognition with a 1808 visual field. *Journal of Vision*, 21.
- Loschky, L., McConkie, G., Yang, J., & Miller, M. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*, 12(6), 1057-1092. <https://doi.org/10.1080/13506280444000652>
- Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding : Selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, 7(2), 24-24.
- Lukavský, J. (2019). Scene categorization in the presence of a distractor. *Journal of Vision*, 19(2), 6. <https://doi.org/10.1167/19.2.6>

- Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4), 176-184. [https://doi.org/10.1016/S1364-6613\(02\)01870-3](https://doi.org/10.1016/S1364-6613(02)01870-3)
- Maniglia, M., Soler, V., Cottureau, B., & Trotter, Y. (2018). Spontaneous and training-induced cortical plasticity in MD patients : Hints from lateral masking. *Scientific Reports*, 8(1), 90. <https://doi.org/10.1038/s41598-017-18261-6>
- Maunsell, J. H., & Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49(5), 1127-1147. <https://doi.org/10.1152/jn.1983.49.5.1127>
- Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49(5), 1127-1147. <https://doi.org/10.1152/jn.1983.49.5.1127>
- McCotter, M., Gosselin, F., Sowden, P., & Schyns, P. (2005). The use of visual information in natural scenes. *Visual Cognition*, 12(6), 938-953.
- McKendrick, A. M., Badcock, D. R., & Morgan, W. H. (2004). Psychophysical Measurement of Neural Adaptation Abnormalities in Magnocellular and Parvocellular Pathways in Glaucoma. *Investigative Ophthalmology & Visual Science*, 45(6), 1846. <https://doi.org/10.1167/iovs.03-1225>
- Merigan, W. H., Byrne, C. E., & Maunsell, J. H. (1991). Does primate motion perception depend on the magnocellular pathway? *Journal of Neuroscience*, 11(11), 3422-3429. <https://doi.org/10.1523/JNEUROSCI.11-11-03422.1991>
- Merigan, W. H., & Maunsell, J. H. R. (1993). How Parallel are the Primate Visual Pathways? *Annual Review of Neuroscience*, 16(1), 369-402. <https://doi.org/10.1146/annurev.ne.16.030193.002101>
- Mirolli, M., Ferrauto, T., & Nolfi, S. (2010). Categorisation through evidence accumulation in an active vision system. *Connection Science*, 22(4), 331-354. <https://doi.org/10.1080/09540091.2010.505976>
- Morel, A., & Bullier, J. (1990). Anatomical segregation of two cortical visual pathways in the macaque monkey. *Visual Neuroscience*, 4(6), 555-578. <https://doi.org/10.1017/S0952523800005769>
- Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic bulletin & review*, 8(3), 454-469.
- Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental brain research*, 227(3), 367-377.
- Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Le Bas, J.-F., & Peyrin, C. (2013). Retinotopic and Lateralized Processing of Spatial Frequencies in Human Visual Cortex during Scene Categorization. *Journal of Cognitive Neuroscience*, 25(8), 1315-1331. https://doi.org/10.1162/jocn_a_00397
- 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. <https://doi.org/10.1371/journal.pone.0038493>
- Musel, B., Hera, R., Chokron, S., Alleysson, D., Chiquet, C., Romanet, J.-P., Guyader, N., & Peyrin, C. (2011). Residual abilities in age-related macular degeneration to process spatial frequencies during natural scene categorization. *Visual Neuroscience*, 28(6), 529-541. <https://doi.org/10.1017/S0952523811000435>
- 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(10), 2287-2297. https://doi.org/10.1162/jocn_a_00643
- Navon, D. (1977). Forest before trees : The precedence of global features in visual perception. *Cognitive psychology*, 9(3), 353-383.
- Nelson, P., Aspinall, P., & O'Brien, C. (1999). Patients' perception of visual impairment in glaucoma : A pilot study. *British Journal of Ophthalmology*, 83(5), 546-552.

- Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 342.
- Odden, J. L., Mihailovic, A., Boland, M. V., Friedman, D. S., West, S. K., & Ramulu, P. Y. (2020). Assessing Functional Disability in Glaucoma : The Relative Importance of Central Versus Far Peripheral Visual Fields. *Investigative Ophthalmology & Visual Science*, *61*(13), 23-23.
- 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. <https://doi.org/10.1006/cogp.1997.0667>
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene : A holistic representation of the spatial envelope. *International journal of computer vision*, *42*(3), 145-175.
- Orban, G. A. (2008). Higher order visual processing in macaque extrastriate cortex. *Physiological Reviews*, *88*(1), 59-89. <https://doi.org/10.1152/physrev.00008.2007>
- Özgen, E., Sowden, P., Schyns, P., & Daoutis, C. (2005). Top-down attentional modulation of spatial frequency processing in scene perception. *Visual Cognition*, *12*(6), 925-937. <https://doi.org/10.1080/13506280444000580>
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, *82*(5), 2490-2502. <https://doi.org/10.1152/jn.1999.82.5.2490>
- Pelli, D. G., Robson, J. G., & J, A. J. W. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences*, 187-199.
- Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics : Transforming numbers into movies. *Spatial vision*, *10*, 437-442.
- Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*(3), 329-342. <https://doi.org/10.1007/BF00239352>
- Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103-112. <https://doi.org/10.1016/j.neuroimage.2018.10.086>
- Peyrin, C., Baci, 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. <https://doi.org/10.1016/j.neuroimage.2004.06.020>
- 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*(2), 278-282. [https://doi.org/10.1016/S0278-2626\(03\)00126-X](https://doi.org/10.1016/S0278-2626(03)00126-X)
- 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. <https://doi.org/10.1016/j.brainres.2005.12.051>
- Peyrin, C., Mermillod, M., Chokron, S., & Marendaz, C. (2006). Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. *Brain and Cognition*, *62*(3), 214-220. <https://doi.org/10.1016/j.bandc.2006.05.005>
- Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., Marendaz, C., & Vuilleumier, P. (2010). The Neural Substrates and Timing of Top-Down Processes during Coarse-to-Fine Categorization of Visual Scenes : A Combined fMRI and ERP Study. *Journal of Cognitive Neuroscience*, *22*(12), 2768-2780. <https://doi.org/10.1162/jocn.2010.21424>
- Peyrin, C., Ramanoël, S., Roux-Sibilon, A., Chokron, S., & Hera, R. (2017). Scene perception in age-related macular degeneration : Effect of spatial frequencies and contrast in residual vision. *Vision Research*, *130*, 36-47. <https://doi.org/10.1016/j.visres.2016.11.004>
- Peyrin, C., Roux-Sibilon, A., Trouilloud, A., Khazaz, S., Joly, M., Pichat, C., Boucart, M., Krainik, A., & Kauffmann, L. (2021). Semantic and Physical Properties of Peripheral Vision Are Used for Scene Categorization in Central Vision. *Journal of Cognitive Neuroscience*, *33*(5), 799-813. https://doi.org/10.1162/jocn_a_01689

- 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*(2), 464-473.
<https://doi.org/10.1016/j.neuroimage.2005.06.006>
- Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative Ophthalmology*, *11*(5), 368-377.
- Pokorny, J., & Smith, V. C. (1997). Psychophysical signatures associated with magnocellular and parvocellular pathway contrast gain. *JOSA A*, *14*(9), 2477-2486.
- Polyak, S. L. (1941). *The retina*.
- Quattoni, A., & Torralba, A. (2009). Recognizing indoor scenes. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 413-420.
- Quigley, H. A., & Broman, A. T. (2006). The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*, *90*(3), 262-267.
- Quigley, H. A., Dunkelberger, G. R., & Green, W. R. (1988). Chronic human glaucoma causing selectively greater loss of large optic nerve fibers. *Ophthalmology*, *95*(3), 357-363.
- Raiguel, S., Van Hulle, M. M., Xiao, D.-K., Marcar, V. L., Lagae, L., & Orban, G. A. (1997). Size and shape of receptive fields in the medial superior temporal area (MST) of the macaque. *NeuroReport*, *8*(12), 2803-2808.
- Ramanoël, S., Kauffmann, L., Cousin, E., Dojat, M., & Peyrin, C. (2015). Age-Related Differences in Spatial Frequency Processing during Scene Categorization. *PLOS ONE*, *10*(8), e0134554.
<https://doi.org/10.1371/journal.pone.0134554>
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex : A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79-87.
<https://doi.org/10.1038/4580>
- Rotshtein, P., Schofield, A., Funes, M. J., & Humphreys, G. W. (2010). Effects of spatial frequency bands on perceptual decision : It is not the stimuli but the comparison. *Journal of Vision*, *10*(10), 25-25. <https://doi.org/10.1167/10.10.25>
- Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual cognition*, *12*(6), 852-877.
- Roux-Sibilon, A. (2020). Reconnaissance visuelle : Étude de l'influence de la vision périphérique dans le cadre d'un modèle fréquentiel prédictif [Thèse de doctorat, Université Grenoble Alpes]. Thèses. <http://www.theses.fr/2020GRALS029>
- Roux-Sibilon, A., Rutgé, F., Aptel, F., Attye, A., Guyader, N., Boucart, M., Chiquet, C., & Peyrin, C. (2018). Scene and human face recognition in the central vision of patients with glaucoma. *PLOS ONE*, *13*(2), e0193465. <https://doi.org/10.1371/journal.pone.0193465>
- Roux-Sibilon, A., Trouilloud, A., Kauffmann, L., Guyader, N., Mermillod, M., & Peyrin, C. (2019). Influence of peripheral vision on object categorization in central vision. *Journal of Vision*, *19*(14), 7. <https://doi.org/10.1167/19.14.7>
- Rovamo, J., Virsu, V., & Näsänen, R. (1978). Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature*, *271*(5640), 54-56.
<https://doi.org/10.1038/271054a0>
- Rubens, A. B., & Benson, D. F. (1971). Associative Visual Agnosia. *Archives of Neurology*, *24*(4), 305-316. <https://doi.org/10.1001/archneur.1971.00480340037003>
- Sabbah, N., Sanda, N., Authié, C. N., Mohand-Saïd, S., Sahel, J.-A., Habas, C., Amedi, A., & Safran, A. B. (2017). Reorganization of early visual cortex functional connectivity following selective peripheral and central visual loss. *Scientific Reports*, *7*(1), 43223.
<https://doi.org/10.1038/srep43223>
- Sabel, B. A., & Gudlin, J. (2014). Vision restoration training for glaucoma : A randomized clinical trial. *JAMA ophthalmology*, *132*(4), 381-389.
- Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system : Structure and function. *Physiological Reviews*, *75*(1), 107-154.
<https://doi.org/10.1152/physrev.1995.75.1.107>

- Sasaki, Y., Hadjikhani, N., Fischl, B., Liu, A. K., Marret, S., Dale, A. M., & Tootell, R. B. H. (2001). Local and global attention are mapped retinotopically in human occipital cortex. *Proceedings of the National Academy of Sciences*, *98*(4), 2077-2082. <https://doi.org/10.1073/pnas.98.4.2077>
- Schafer, A., Rouland, J. F., Peyrin, C., Szaffarczyk, S., & Boucart, M. (2018). Glaucoma affects viewing distance for recognition of sex and facial expression. *Investigative ophthalmology & visual science*, *59*(12), 4921-4928.
- Schiller, P. H., Logothetis, N. K., & Charles, E. R. (1990). Functions of the colour-opponent and broad-band channels of the visual system. *Nature*, *343*(6253), 68-70. <https://doi.org/10.1038/343068a0>
- Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show Me the Features! Understanding Recognition From the Use of Visual Information. *Psychological Science*, *13*(5), 402-409. <https://doi.org/10.1111/1467-9280.00472>
- Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges : Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, *5*(4), 195-200. <https://doi.org/10.1111/j.1467-9280.1994.tb00500.x>
- Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile : When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3), 243-265. [https://doi.org/10.1016/S0010-0277\(98\)00069-9](https://doi.org/10.1016/S0010-0277(98)00069-9)
- Sereno, M. I., McDonald, C. T., & Allman, J. M. (1994). Analysis of retinotopic maps in extrastriate cortex. *Cerebral Cortex*, *4*(6), 601-620.
- 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*(3), 259-273. <https://doi.org/10.1068/p150259>
- Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to spatial location. *Perception*, *16*(1), 103-111.
- Sloan, L. L. (1959). New test Charts for the Measurement of Visual Acuity at far and Near Distances*. *American Journal of Ophthalmology*, *48*(6), 807-813. [https://doi.org/10.1016/0002-9394\(59\)90626-9](https://doi.org/10.1016/0002-9394(59)90626-9)
- Smith, N. D., Crabb, D. P., & Garway-Heath, D. F. (2011). An exploratory study of visual search performance in glaucoma. *Ophthalmic and Physiological Optics*, *31*(3), 225-232.
- Smith, N. D., Glen, F. C., & Crabb, D. P. (2012). Eye movements during visual search in patients with glaucoma. *BMC ophthalmology*, *12*(1), 1-11.
- Stievenard, A., Rouland, J. F., Peyrin, C., Warniez, A., & Boucart, M. (2021). Sensitivity to central crowding for faces in patients with glaucoma. *Journal of Glaucoma*, *30*(2), 140-147.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, *262*(5134), 685-688. <https://doi.org/10.1126/science.8235589>
- Tanaka, K. (1996). Inferotemporal Cortex and Object Vision. *Annual Review of Neuroscience*, *19*(1), 109-139. <https://doi.org/10.1146/annurev.ne.19.030196.000545>
- Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, *66*(1), 170-189. <https://doi.org/10.1152/jn.1991.66.1.170>
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520-522. <https://doi.org/10.1038/381520a0>
- Tootell, R. B., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, *218*(4575), 902-904.
- Tootell, R. B., Switkes, E., Silverman, M. S., & Hamilton, S. L. (1988). Functional anatomy of macaque striate cortex. II. Retinotopic organization. *Journal of neuroscience*, *8*(5), 1531-1568.
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391-412. <https://doi.org/10.1088/0954-898X/14/3/302>
- Tran, T. H. C., Rambaud, C., Desprez, P., & Boucart, M. (2010). Scene perception in age-related macular degeneration. *Investigative ophthalmology & visual science*, *51*(12), 6868-6874.

- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition : Predictions and competition. *Annals of the New York Academy of Sciences*, 1339(1), 190-198. <https://doi.org/10.1111/nyas.12680>
- Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, 4(2), 157-165. [https://doi.org/10.1016/0959-4388\(94\)90066-3](https://doi.org/10.1016/0959-4388(94)90066-3)
- Van Essen, D. C., & DeYoe, E. A. (1995). *Concurrent processing in the primate visual cortex*.
- Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, 6, 370-375. [https://doi.org/10.1016/0166-2236\(83\)90167-4](https://doi.org/10.1016/0166-2236(83)90167-4)
- Vanmarcke, S., & Wagemans, J. (2016). Individual differences in spatial frequency processing in scene perception : The influence of autism-related traits. *Visual Cognition*, 24(2), 115-131. <https://doi.org/10.1080/13506285.2016.1199625>
- VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing : From early perception to decision-making. *Journal of cognitive neuroscience*, 13(4), 454-461.
- Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental brain research*, 37(3), 475-494.
- Viswanathan, A. C., McNaught, A. I., Poinoosawmy, D., Fontana, L., Crabb, D. P., Fitzke, F. W., & Hitchings, R. A. (1999). Severity and stability of glaucoma : Patient perception compared with objective measurement. *Archives of Ophthalmology*, 117(4), 450-454.
- Vlad-Debuschere, R., Guyader, N., & Guérin-Dugué, A. (2015). A bio-inspired model of central and peripheral vision for scene categorization. *XXVème colloque GRETSI (GRETSI 2015)*.
- Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2 : Single-cell study. *The European Journal of Neuroscience*, 11(4), 1239-1255. <https://doi.org/10.1046/j.1460-9568.1999.00531.x>
- Voyer, D. (1996). On the magnitude of laterality effects and sex differences in functional lateralities. *Laterality*, 1(1), 51-84.
- Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, 56(2), 366-383.
- Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision research*, 51(7), 718-737.
- Wang, A. Y., Lee, P. Y., Bui, B. V., Jobling, A. I., Greferath, U., Brandli, A., Dixon, M. A., Findlay, Q., Fletcher, E. L., & Vessey, K. A. (2020). Potential mechanisms of retinal ganglion cell type-specific vulnerability in glaucoma. *Clinical and Experimental Optometry*, 103(5), 562-571.
- Wang, J., Li, T., Sabel, B. A., Chen, Z., Wen, H., Li, J., Xie, X., Yang, D., Chen, W., Wang, N., Xian, J., & He, H. (2016). Structural brain alterations in primary open angle glaucoma : A 3T MRI study. *Scientific Reports*, 6(1), 18969. <https://doi.org/10.1038/srep18969>
- Weber, A. J., Chen, H., Hubbard, W. C., & Kaufman, P. L. (2000). *Experimental Glaucoma and Cell Size, Density, and Number in the Primate Lateral Geniculate Nucleus*. 41(6), 11.
- West, S., Munoz, B., Brandeen-Roche, K., Zeger, S., Schein, O., Fried, L., The See Project Team, & Rubin, G. (1997). A comprehensive assessment of visual impairment in a population of older Americans : The SEE study. *American Journal of Ophthalmology*, 124(2), 272-273. [https://doi.org/10.1016/S0002-9394\(14\)70808-6](https://doi.org/10.1016/S0002-9394(14)70808-6)
- Wiesmann, S. L., Caplette, L., Willenbockel, V., Gosselin, F., & Vö, M. L.-H. (2021). Flexible time course of spatial frequency use during scene categorization. *Scientific Reports*, 11(1), 14079. <https://doi.org/10.1038/s41598-021-93252-2>
- Witelson, S. F. (1983). Bumps on the brain : Right-left anatomic asymmetry as a key to functional Lateralization. In *Language functions and brain organization* (p. 117-144). Elsevier.
- Witelson, S. F., & Kigar, D. L. (1988). *Asymmetry in brain function follows asymmetry in anatomical form : Gross, microscopic, postmortem and imaging studies*.
- Wu, J., Yan, T., Zhang, Z., Jin, F., & Guo, Q. (2012). Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging. *Human brain mapping*, 33(7), 1727-1740.

- Xu, X., Anderson, T. J., & Casagrande, V. A. (2007). How do functional maps in primary visual cortex vary with eccentricity? *Journal of Comparative Neurology*, 501(5), 741-755.
- Yamada, E. (1969). Some structural features of the fovea centralis in the human retina. *Archives of ophthalmology*, 82(2), 151-159.
- Yucel, Y. H., Zhang, Q., Weinreb, R. N., Kaufman, P. L., & Gupta, N. (2001). Atrophy of Relay Neurons in Magno- and Parvocellular Layers in the Lateral Geniculate Nucleus in Experimental Glaucoma. 42(13), 7.