Développement et mise en oeuvre de modèles d'attention visuelle

Tien Ho Phuoc

UNIVERSITE DE GRENOBLE
INSTITUT POLYTECHNIQUE DE GRENOBLE

N° attribué par la bibliothèque

THESE

pour obtenir le grade de

DOCTEUR DE L'Université de Grenoble
délivré par l'Institut polytechnique de Grenoble

Spécialité : "Signal, Image, Parole, Télécoms"

préparée au laboratoire **GIPSA-Lab**

dans le cadre de **l'École Doctorale *"Electronique, Electrotechnique, Automatique et Traitement du Signal"***

présentée et soutenue publiquement

par

HO-PHUOC Tien

le 6 mai 2010

DEVELOPPEMENT ET MISE EN OEUVRE DE MODELES D'ATTENTION VISUELLE

Directrice de thèse: **Anne Guérin-Dugué**
Co-directrice de thèse: **Nathalie Guyader**

JURY

M. Jean-Luc Schwartz,	Président du jury
M. Patrick Lambert,	Rapporteur
M. Abdelhakim Saadane,	Rapporteur
M. Frédéric Landragin,	Examinateur
Mme. Anne Guérin-Dugué,	Directrice de thèse
Mme. Nathalie Guyader,	Co-directrice de thèse

Résumé

Pour explorer le monde qui nous entoure nous bougeons sans cesse les yeux alternant entre des mouvements rapides “les saccades” et des moments d’immobilisation “les fixations”. Quels sont les facteurs guidant ces mouvements oculaires ? Comment les interpréter et les évaluer quantitativement ? Cette thèse aborde ces questions lors de l’exploration libre de scènes naturelles, sous deux aspects : la modélisation et le recueil de données comportementales avec des enregistrements oculométriques.

Le modèle proposé s’inspire fortement de la biologie et propose de prédire les régions dites saillantes (qui attirent les yeux) en utilisant un certain nombre de caractéristiques visuelles de bas niveaux, selon une démarche de traitement ascendante (“bottom-up”) compatible avec le contexte choisi d’exploration libre des scènes naturelles. Bien qu’il s’agisse de l’exploration de scènes statiques, un modèle dynamique spatio-temporel est également proposé considérant les séquences temporelles alternant les phases de stabilisation durant les fixations et les phases de déplacement durant les saccades. Les données comportementales et les données physiologiques ont permis l’établissement du modèle, ses évolutions et améliorations successives, puis sa validation.

Ainsi, nous montrons que bien que la couleur soit présente partout et apparaisse dans plusieurs modèles de la littérature, celle-ci influence peu les mouvements oculaires des sujets. De même nous montrons que programmer plusieurs saccades en parallèle à partir d’un point de fixation comme cela a été montré dans des expériences de recherche de cible n’est pas compatible avec les données comportementales.

Cette thèse propose aussi de nombreux outils méthodologiques pour comparer des données comportementales à des données issues d’un modèle et propose également une manière de tester l’importance relative de plusieurs caractéristiques visuelles de bas niveau sur la prédiction des mouvements oculaires.

Abstract

To explore the world around us, we move without cease our eyes altering between rapid movements, "saccades", and immobile moments, "fixations". What factors guide these eye movements ? How to interpret and evaluate quantitatively them ? This thesis addressed these problems in the context of free viewing of natural scenes according two aspects : modeling and behavioural data recording with eye movements experiments.

The proposed model was inspired mainly by the biology of the human visual system and predicted the salient regions (which attract the eyes) by using some low-level visual features according to the bottom-up architecture, which is compatible with the context of free viewing of natural scenes. Even though considering static scenes, we also proposed a spatio-temporal model to take into account temporal sequences of stabilization phases during fixations and movement phases during saccades. The behavioural and physiological data helped to build, improve successively and validate the model.

We presented that although colour was often used in most models in the literature, it had little influence on subjects' eye movements. We also showed that programming several saccades in parallel from a fixation point, which had been observed in the context of visual search, was not compatible with the experimental data.

This thesis also described several methodological tools to compare experimental data with the data predicted by the model and proposed a method to evaluate the relative contributions of low-level visual features to the prediction of eye movements.

Remerciements

Tout d'abord, je tiens à remercier mes directrices de thèse, Anne Guérin-Dugué et Nathalie Guyader, pour leur disponibilité et leurs encadrements enthousiastes tout au long de ma thèse. Ce travail n'aurait pas été abouti sans leurs aides précieuses. Elles m'ont beaucoup appris à la fois par leur qualité scientifique et par leur méthode pédagogique.

Je tiens ensuite à remercier Jean-Luc Schwart d'avoir accepté de présider mon jury. Merci également à Patrick Lambert et Abdelhakim Saadane pour leurs rapports détaillés sur mon manuscrit et des commentaires très instructifs. Je tiens aussi à remercier Frédéric Langdragin pour sa participation au jury en tant qu'examinateur et pour des questions intéressantes.

Je remercie le department DIS (LIS), en particulier l'équipe GPIG (Géométrie, Perception, Images, Gestes), pour une ambiance de travail agréable tout au long de ma thèse. Merci à Jeanny Hérault, Denis Pellerin, Barthélémy Durette, Sophie Marat, Granjon Lionel, Ionescu Gelu pour leurs aides précieuses et des discussions fructeuses. Merci aussi à David Alleyson et Sophie Achard pour leurs conseils instructifs. Je remercie également toutes les personnes qui ont participé à mes expériences oculométriques.

Je tiens à remercier Marie-No, Lucia et Jean-Marc Sache ainsi qu'à toute l'équipe administrative et informatique du Gipsa-Lab pour leurs soutiens pendant ma thèse et lors de la preparation de la soutenance.

Enfin, je remercie mes parents et mes proches, qui m'ont toujours encouragé. Cette thèse est dédiée à mes parents pour tout ce qu'ils m'ont donné et inspiré.

Table des matières

Introduction générale

Lorsque nous regardons le monde qui nous entoure nous bougeons sans cesse les yeux. Ces mouvements oculaires correspondent en fait à une succession de mouvements très rapides des deux yeux simultanément (les saccades) avec entre deux mouvements des phases plus ou moins longues durant lesquelles les yeux se stabilisent sur une région particulière du monde environnant (les fixations). Ces mouvements oculaires sont indispensables pour la perception visuelle en permettant de capter la très grande majorité de l'information sensorielle venant du monde extérieur. Le reste correspondant bien entendu aux sons, mais également en fonction des situations aux odeurs, au goût et au toucher.

La compréhension de comment et surtout de vers quelles régions bougeons nous les yeux pour un environnement visuel particulier est d'un très grand intérêt pour de nombreuses disciplines, et avant tout, pour une meilleure compréhension de notre système visuel. Cette meilleure compréhension de notre système visuel et plus particulièrement des mouvements oculaires sont depuis quelques dizaines d'années très utiles en clinique pour, par exemple, aider au diagnostique de certaines pathologies (maladies basse-vision, schizophrénie, ou maladie d'Alzheimer). Ces travaux sont également indispensables en sciences de l'ingénieur ou encore en marketing ou en "design". Ainsi comprendre où les gens regardent peut permettre de développer des algorithmes de compression et de transmission d'images liés aux zones regardées, ou encore de créer des outils de visualisation d'images adaptés. Il n'est également pas rare de voir que ce type d'études permet à des publicitaires de construire différemment leurs bandes annonces, à des magasins d'agencer leurs rayons d'une certaine manière, à des sites Internet de présenter leurs pages web sous des angles différents pour l'objectif final d'attirer l'attention des clients ou des lecteurs. A la lecture de ces quelques lignes, les lecteurs comprendront donc aisément pourquoi comprendre comment et où les gens regardent est un challenge d'un intérêt scientifique, sociétal et économique grandissant.

Il est également à noter que l'engouement décrit ci-dessus a considérablement augmenté depuis quelques années grâce aux nouvelles technologies qui permettent d'enregistrer les mouvements oculaires de plus en plus vite et dans des situations de plus en plus écologiques. Ainsi, les nouveaux appareils d'enregistrement sont portables, ne nécessitent plus forcément d'être envahissants (simple caméra posée en face d'un utilisateur), et permettent donc à l'utilisateur d'agir le plus naturel possible. De plus, les appareils permettent maintenant de connaître avec des précisions remarquables la position des yeux toutes les millisecondes.

La thèse, présentée dans ce manuscrit, se place dans ce contexte de mieux comprendre comment et vers quels endroits vont se porter nos yeux devant tel ou tel stimulus visuel.

Comme nous l'avons dit, les gens explorent les scènes visuelles en déplaçant leurs yeux vers des zones particulières appelées "zones d'intérêt". Nous verrons par la suite que ce sont ces zones d'intérêt qui sont analysées par notre système visuel en détails : elles sont tout d'abord captées par l'œil et prétraitées par les cellules de la rétine puis transmises au cortex visuel. C'est grâce à la perception successive de ces zones d'intérêt que nous formons le percept continu et complet de notre environnement visuel.

A quoi correspondent ces zones d'intérêt ? Ou plutôt qu'est-ce qui définit ces zones d'intérêt ? Deux situations particulières se distinguent facilement. Dans la première, nous regardons une scène visuelle sans but précis ; nos yeux vont bouger et être attirés par des particularités du stimulus visuel seul. On dit alors que les mouvements oculaires sont guidés par des facteurs dits de bas niveau ou exogènes. Dans la deuxième situation, nous regardons une scène visuelle avec un but précis, par exemple à la recherche d'un objet. Nos yeux seront alors principalement guidés par ce que nous cherchons (but) et toute zone susceptible de ressembler à cet objet précis sera regardée. On dit alors que les mouvements oculaires sont guidés par des facteurs dits de haut niveau ou endogènes. Des études ont montré qu'en fait les deux mécanismes, exogènes et endogènes, pouvaient être liés et que bien souvent ces deux mécanismes intervenaient en même temps avec des décours temporels légèrement différents. Ainsi, le stimulus visuel et donc les processus exogènes guideraient les mouvements oculaires immédiatement à l'apparition de la scène et les processus endogènes viendraient ensuite prendre le relais. Il est à noter qu'indépendamment de la nature de ce qui a guidé le mouvement des yeux, celui-ci est précédé par un déplacement de l'attention visuelle, une capacité du cerveau à sélectionner la zone vers laquelle se fera le mouvement des yeux.

Dans cette thèse, nous tenterons d'apporter des réponses pour aider à mieux comprendre comment et vers quels endroits les gens regardent. Pour ce faire, nous nous sommes fixés un cadre de travail. Tout d'abord, nous adopterons une double approche : une approche par la modélisation et une approche comportementale. Grâce à la première, nous proposons un modèle "computationnel" de perception visuelle ; ce modèle s'inspire des premiers étages du système visuel et permet de faire ressortir dans une scène les régions dites saillantes. Ces régions sont celles qui sont les plus susceptibles d'attirer l'attention visuelle et donc les yeux. Un tel modèle est appelé dans la littérature, *modèle d'attention visuelle* ou encore *modèle de saillance*. La seconde approche nous permettra grâce à une expérience utilisant l'oculométrie de connaître les endroits regardés par un certain nombre de personnes sur des images de scènes naturelles. Nous avons choisi de nous placer dans des conditions de laboratoire. Ainsi, nous étudierons l'exploration d'images statiques présentées sur l'écran d'un ordinateur. Les observateurs auront la tête fixe et visionneront les images librement sans but précis. Les deux approches se nourrissent l'une l'autre. L'approche comportementale fournira une base de données réelle correspondant à la "vérité ter-

rain" ; elle permet de confronter et d'améliorer le modèle proposé. Le modèle proposé permettra à son tour de tester ou de prédire les données comportementales.

Les modèles d'attention visuelle extraient les caractéristiques visuelles des stimuli et/ou le but d'une tâche, la sémantique d'une image afin de créer une carte représentant les zones saillantes. Le modèle d'attention visuelle que nous proposons est uniquement basé sur des facteurs de bas niveau comme la luminance, la couleur, l'orientation pour prédire les mouvements oculaires durant les premières secondes de visualisation. Notre modèle s'inspire de la biologie du système visuel et modélise le fonctionnement des principales cellules de la rétine et du cortex visuel primaire. Au niveau de la rétine, une image est décomposée en trois images différentes : une image de luminance et deux images de chrominance (les oppositions rouge - vert et bleu - jaune). Ces trois images subissent ensuite des traitements qui ont pour but principal de renforcer leurs contrastes. Au niveau du cortex visuel primaire, les images sont ensuites décomposées en cartes d'énergie correspondant à différentes orientations et différentes fréquences spatiales du stimulus visuel. Ces cartes sont ensuite normalisées et des mécanismes d'interactions entre cartes permettent de faire ressortir les zones saillantes. Enfin, une carte de saillance unique est créée par fusion des différentes cartes d'énergie.

Alors que les caractéristiques visuelles sont nombreuses, leurs rôles dans les mouvements oculaires diffèrent et restent encore une question ouverte. Nous testerons dans cette thèse, l'influence de la couleur vis-à-vis de celle de la luminance. En effet, de nombreux modèles utilisent la couleur comme une caractéristique visuelle jouant un rôle dans l'attention visuelle, mais peu d'études ont réellement tenté de mesurer l'importance relative de cet attribut par rapport à la luminance. Pour cela, nous proposons d'étudier les fixations de sujets sur des images en couleur et sur les mêmes images en niveau de gris. Nous avons de plus quantifié par une approche méthodologique originale les contributions d'un grand nombre de caractéristiques visuelles de bas niveau aux mouvements oculaires.

En nous basant sur notre modèle d'attention visuelle, nous avons également étudié un problème courant dans les recherches sur les mouvements oculaires qui est celui de "la programmation de saccade". Ainsi, il a été montré dans des expériences de recherche de cible que le système visuel programme à partir d'un même point de vue les deux fixations (ou les deux saccades) suivantes. Nous souhaitons tester si cette programmation de plusieurs saccades en parallèle à partir d'un même point de vue existe également lors de l'exploration libre de scènes naturelles. Nous avons ainsi voulu répondre à la question "à partir d'une fixation, combien de saccades programmons-nous ?" Pour y répondre nous avons simulé en utilisant trois modèles de programmation de saccade, chacun basé sur notre modèle d'attention visuelle, des points de fixations sur une base d'images. Et nous avons confronté ces trois modèles à des données réelles.

Enfin, nous avons approfondi le modèle de la rétine en examinant les traitements rétiniens dans un processus spatio-temporel. Nous modélisons le fonctionnement de la rétine lors d'une succession de fixations et de saccades durant l'exploration de

scènes. En outre, nous étudions un échantillonnage spatialement variant d'une image avec un taux de plus en plus important du centre à la périphérie en tenant compte de la densité non uniforme des cellules ganglionnaires. Cet échantillonnage causera la déformation de l'image à sa périphérie ainsi que la déformation de son spectre. En examinant cette déformation, nous pouvons affiner notre modèle d'attention visuelle.

Plan de la thèse

Le plan de la thèse est le suivant.

Le chapitre 1 est consacré à l'état de l'art en introduisant les propriétés des mouvements oculaires, la programmation de saccade et les modèles d'attention visuelle les plus courants de la littérature.

Le chapitre 2 présente notre modèle d'attention visuelle ; celui-ci est inspiré des caractéristiques biologiques du système visuel depuis le fonctionnement des cellules rétiniennes jusqu'au fonctionnement des cellules corticales. Ce modèle permet de prédire sur des images statiques les zones les plus susceptibles d'être regardées durant les premières secondes de visualisation.

Le chapitre 3 passe en revue les métriques de la littérature qui permettent de comparer la sortie d'un modèle d'attention visuelle avec des données oculométriques. Dans ce chapitre, nous proposons également des approches méthodologiques originales pour trouver quelles sont les caractéristiques visuelles de bas niveau qui expliquent au mieux un jeu de données oculométriques.

Le chapitre 4 permet de tester par une expérience comportementale l'influence de la couleur sur les mouvements oculaires obtenus lors de l'exploration libre de scènes naturelles. Il évalue également quelles sont les contributions relatives de différentes caractéristiques visuelles de bas niveau aux mouvements oculaires obtenus.

Le chapitre 5 évalue trois différents modèles de programmation de saccade : un modèle pour lequel à partir d'un même point de vue toutes les saccades suivantes sont programmées, un modèle pour lequel les deux saccades suivantes sont programmées et enfin un modèle pour lequel à partir d'un point de vue seule la saccade suivante est programmée.

Enfin, le chapitre 6 représente le traitement spatio-temporel de la rétine et l'échantillonnage spatialement variant d'une image. Ce chapitre ouvre des pistes de travail pour continuer à appréhender les modèles d'attention visuelle spatio-temporelle et toujours mieux comprendre comment l'homme explore son environnement visuel et comment il intègre au cours de cette exploration l'information visuelle.

Chapitre 1

Mouvements oculaires et modèles d'attention visuelle

1 Mouvements oculaires lors de l'exploration de scènes

L'exploration de notre environnement est une tâche essentielle pour acquérir des informations sur le monde qui nous entoure. Cette exploration est d'autant plus importante que la grande majorité des informations acquises sont des informations visuelles. Comprendre les mécanismes impliqués dans cette exploration est un enjeu de taille dans de nombreux domaines de recherche (vision artificielle, neuroscience, psychologie, etc).

L'exploration d'une scène visuelle correspond à des mouvements plus ou moins rapides des yeux et s'effectue sans même que nous en ayant forcément conscience. Selon des études de l'Université de Pennsylvanie, l'œil humain est capable de capter 8,75 Mégabits d'information par seconde [Koch et al., 2006]. Pour comparaison, une image en couleur de taille 768 $\times$ 1024 pixels occupe 18 Mégabits. Ainsi, la quantité d'information captée par l'œil semble relativement faible comparée à la quantité d'information qui arrive constamment sur les yeux. Comment le système visuel "comble"-t-il cette ressource limitée ? La réponse se cache dans les mécanismes sous-jacents au système visuel humain. L'une des manières d'aborder la compréhension du système visuel humain est l'étude des mouvements oculaires, ces mouvements qui nous permettent d'explorer une scène visuelle.

1.1 Mouvements oculaires

Il y a deux principaux types de mouvements oculaires : le mouvement saccadique et le mouvement de poursuite. Le premier est celui que l'on utilise pour explorer une scène et durant lequel nos yeux se fixent successivement sur différentes positions dans la scène. Le deuxième est celui que l'on utilise lorsque les yeux suivent un objet en mouvement. Alors que nous nous intéressons, dans cette thèse, à l'exploration d'une scène visuelle statique, nous nous concentrons uniquement sur le mouvement saccadique.

Les mouvements oculaires saccadiques comportent des *saccades* et des *fixations*. D'après [Cassin and Solomon, 1990], une saccade est un mouvement rapide, simultané de deux yeux dans la même direction. La saccade est un phénomène binoculaire qui a été observé depuis l'antiquité. La binocularité est remarquée par Aristotle (384-322 BC)[1], et puis, plus de deux mille ans plus tard par Hermann Helmholtz (1864)[2]. Une saccade est décrite par son point de départ, son amplitude (en degrés angulaires), sa vitesse (en degrés par seconde) et sa direction. L'amplitude d'une saccade est la distance parcourue par l'œil (en degrés angulaires). Une saccade est réalisée dans une direction pour mettre une région de la scène visuelle au centre de la rétine (fovéa). La plupart des saccades s'effectuent en quelques dizaines de millisecondes. Une autre grandeur caractéristique d'une saccade est sa latence qui est le temps de déclenchement de la saccade entre l'apparition d'un stimulus et le début de la saccade. La latence d'une saccade à l'apparition d'une cible soudaine en périphérie est de l'ordre de 200 à 250 ms. Cette latence dépend de plusieurs facteurs tels que la condition expérimentale, l'âge, etc. A titre d'exemple, les enfants ont une latence plus longue que les adultes [Munoz et al., 1998; Yang et al., 2002].

Une saccade est liée à une fixation, une autre notion essentielle pour les mouvements oculaires. Une fixation a lieu lorsque les yeux se stabilisent sur une position du champ visuel pendant un certain temps. Une fixation est caractérisée par sa position spatiale et sa durée. La durée moyenne d'une fixation est de 200 à 250 ms. Les mouvements oculaires saccadiques se composent d'une succession de saccades et de fixations (Fig. 1.1). A chaque fixation, une petite région de la scène, où les yeux se stabilisent, est analysée en détails. Ce mécanisme est compatible avec la structure anatomique de la rétine. La densité des cônes est très grande dans la fovéa et diminue rapidement avec l'excentricité par rapport à la fovéa [Curcio et al., 1990]. Ainsi, des saccades permettent de placer des zones d'intérêt sur la fovéa pour une analyse détaillée durant des fixations. Le tableau 1.1 représente les valeurs moyennes des amplitudes de saccades, des durées de saccades, des latences de saccades et des durées de fixations obtenues à partir des expériences oculométriques.

Il est à noter que les mouvements oculaires sont différents d'un sujet à l'autre. Autrement dit, deux sujets qui regardent une même scène visuelle n'auront pas les mêmes mouvements (ou trajets) oculaires. Toutefois, lorsque l'on examine les mouvements oculaires pour un certain nombre de sujets, on peut révéler des tendances systématiques dans ces mouvements. Ces tendances seront étudiées à la section §1.4.

Les mouvements oculaires ont une relation étroite avec l'attention visuelle. Selon [Henderson, 1993; Hoffman, 1998], chaque mouvement oculaire vers une position est précédé par un déplacement de l'attention vers cette position. Dans ce cas-là, on parle d'*"overt attention"*. Il est important de noter qu'il est possible de déplacer son attention sans bouger les yeux ; l'attention peut être dirigée vers plusieurs positions

[1] Aristotle : "*So when one eye moves, the common source of sight is also set in motion ; and when this moves, the other eye moves also*" [Wade and Tatler, 2005]

[2] Hermann Helmholtz : "*We can not turn one eye up, the other down ; we can not move both eyes at the same time to the outer angle*" [Wade and Tatler, 2005]

FIG. 1.1 – Exemple des saccades (traits) et des fixations (points) d'un sujet lors de l'exploration de la scène présentée pendant 3 s avec un départ de l'exploration visuelle au centre de la scène.

TAB. 1.1 – Valeurs moyennes de certains paramètres des saccades et des fixations obtenues lors de l'exploration de stimuli visuels.

Paramètres	Valeur
Durée de saccade	Quelques dizaines de ms
Latence de saccade	200-250 ms
Amplitude de saccade	7°
Durée de fixation	200-250 ms

dans le champ visuel sans impliquer de mouvements oculaires. Cette attention est appelée *"covert attention"*. Dans les études concernant les mouvements oculaires, on parle uniquement d' *"overt attention"*.

1.2 Expérience oculométrique

Une expérience oculométrique permet d'enregistrer les mouvements oculaires. Cet enregistrement est effectué à l'aide des systèmes oculométriques. En 1935, Buswell [Buswell, 1935] est le premier à avoir utilisé l'oculométrie pour enregis-

trer les mouvements oculaires de personnes regardant un tableau. Cette technique a également été utilisée dans des tâches de lecture [Buswell, 1922, 1937]. Depuis, de très nombreuses études utilisent l'oculométrie avec différents types d'images, et différentes tâches pour comprendre les mécanismes sous jacents de la vision. Suivant la tâche, les expériences oculométriques se séparent en deux catégories : "exploration libre" (sans aucune tâche) et "exploration sous contrainte" (avec une tâche particulière). Pour la première, les sujets regardent librement un stimulus visuel sans consigne particulière. Pour la deuxième, les sujets explorent une scène avec une consigne bien précise (par exemple, chercher le nombre de personnages, chercher un objet, les personnes présentées semblent-elles heureuses ?, etc).

Malgré de nombreuses expériences depuis celle de Buswell en 1935, il n'est pas facile d'étudier les mouvements oculaires dans des conditions réelles, c'est-à-dire dans un contexte interactif entre les sujets et le monde extérieur. Ainsi, on préfère une expérience simplifiée en laboratoire. Les scènes visuelles sont présentées sur l'écran d'un ordinateur. Les sujets sont installés face à l'écran à une distance pré-définie et un système de caméras et/ou lumière infra rouge permet d'enregistrer les mouvements oculaires lorsque les sujets visionnent les scènes visuelles. Avec l'évolution des systèmes oculométriques (précision, rapidité d'enregistrement, etc), l'enregistrement des mouvements oculaires est de plus en plus précis. De plus la simplicité des nouveaux systèmes ont amené une augmentation considérable du nombre d'études utilisant l'oculométrie.

1.3 Facteurs influençant les mouvements oculaires

Comme nous l'avons dit si l'on examine les trajets oculaires de plusieurs sujets sur une même scène, on peut révéler des tendances systématiques dans ces trajets. Les mouvements oculaires ne sont donc pas aléatoires mais dépendent de plusieurs facteurs qui peuvent être regroupés en deux groupes.

Facteurs dits de "bas niveau" ou exogènes : ce sont les caractéristiques de bas niveau associées aux stimuli visuels comme la luminance, l'orientation ou la couleur. A titre d'exemple, une barre verticale parmi des barres horizontales attire l'attention et les yeux de même qu'un objet rouge sur un fond vert. Quand on étudie les caractéristiques de bas niveau sur les régions qui correspondent à des fixations, on observe qu'elles sont différentes de celles sur des régions choisies aléatoirement dans l'image. Cela supporte l'idée que les facteurs de bas niveau, associés aux stimuli visuels, peuvent contribuer à l'explication des mouvements oculaires. Il existe un grand nombre de recherches dans la littérature soutenant cette idée [Mannan et al., 1997; Krieger et al., 2000; Reinagel and Zador, 1999; Parkhurst et al., 2002; Parkhurst and Niebur, 2003, 2004; Tatler et al., 2005, 2006; Baddeley and Tatler, 2006]. Le contraste de luminance est le plus souvent évoqué comme un facteur de bas niveau attirant l'attention et les yeux [Parkhurst et al., 2002; Parkhurst and Niebur, 2004]. De plus, Baddeley et Tatler [Baddeley and Tatler, 2006] ont observé que les contours, vus comme des contrastes locaux importants, sont un facteur essentiel influençant les mouvements oculaires.

Facteurs dits de "haut niveau" ou endogènes : ce sont des facteurs dépendants des processus cognitifs, de la mémoire, de l'état émotionnel ou encore de la tâche, etc. Ces facteurs sont donc sujet-dépendants. Alors que plusieurs études ont montré le rôle important des facteurs de haut niveau dans les mouvements oculaires, ces facteurs sont difficiles à quantifier et à contrôler. Comme les facteurs de bas niveau, on peut trouver un grand nombre d'études soutenant la contribution des facteurs de haut niveau aux mouvements oculaires. Même dans les premières études concernant l'oculométrie, on a observé l'influence des facteurs de haut niveau lors de l'exploration de tableaux [Buswell, 1935]. Dans [Yarbus, 1967], il remarque une forte relation entre les fixations et la tâche. Ces travaux sont souvent cités dans la littérature pour illustrer l'influence des facteurs de haut niveau sur les mouvements oculaires. Il a montré qu'en fonction de la tâche, plus particulièrement de la question posée au sujet, les mouvements oculaires réalisés sont différents. Il a aussi montré que les facteurs de haut niveau allaient influencer de manière prépondérante par rapport aux facteurs de bas niveau les mouvements oculaires. Récemment, plusieurs études ont également mis en évidence le rôle des autres facteurs de haut niveau comme la sémantique et le contexte dans les mouvements oculaires [Henderson and Hollingworth, 1999; Henderson et al., 1999; Land and Hayhoe, 2001; Turano et al., 2003; Oliva et al., 2003; Castelhano et al., 2009; Henderson, 2003].

Ainsi, les études sur les mouvements oculaires ont confirmé le rôle des facteurs de bas niveau et des facteurs de haut niveau dans ces mouvements. Cependant l'influence de ces deux facteurs sur les mouvements oculaires diffère dans le temps avec une prédominance des facteurs de bas niveau au début de l'exploration et une augmentation de l'importance des facteurs de haut niveau avec le temps. Dans cette thèse, nous nous intéresserons uniquement à l'étude des facteurs de bas niveau, et donc à l'étude des premiers mouvements oculaires à l'apparition d'une scène.

1.4 Caractéristiques générales des mouvements oculaires

Des tendances générales ont été observées en examinant les mouvements oculaires pour un grand nombre de sujets. Nous présentons les caractéristiques souvent rencontrées lors de l'exploration de scènes.

1.4.1 Le biais de centralité

Lorsque l'on explore une scène, le centre de la scène a une forte probabilité d'attirer l'attention et donc les yeux. Ce phénomène appelé, *biais de centralité*, est noté dans plusieurs recherches [Buswell, 1935; Parkhurst et al., 2002; Parkhurst and Niebur, 2003; Tatler, 2007]. Alors que ce biais de centralité est souvent observé, les explications pour ce phénomène ne font pas l'unanimité.

La première explication avancée vient du fait que le sujet commence l'exploration de la scène au centre. En effet, dans la majorité des expériences un point de fixation central précède l'apparition de la scène. Or, la plupart des saccades ont de petites amplitudes de l'ordre de 7° [Bahill et al., 1975]. Cela pourrait expliquer pourquoi les sujets fixent plus la zone autour du centre que la périphérie s'ils commencent leur

exploration au centre de la scène.

Une deuxième explication concerne la nature de la scène. Plusieurs études ont montré que les yeux sont attirés par certaines caractéristiques de bas niveau dans une scène [Reinagel and Zador, 1999; Parkhurst et al., 2002; Parkhurst and Niebur, 2003, 2004; Tatler et al., 2005, 2006; Baddeley and Tatler, 2006]. Or, les stimuli utilisés souvent dans les expériences sont des scènes naturelles photographiées. Un photographe a tendance à mettre les objets d'intérêt au centre.

Pourtant, dans ses travaux [Tatler, 2007], Tatler a démontré que les deux causes ci-dessus ne peuvent être complètement responsables du biais de centralité des mouvements oculaires. D'abord, pour l'explication concernant le point de départ de l'exploration, une séquence artificielle de saccades est engendrée à l'aide du modèle de marche aléatoire ("*random walk*"). Ces marches ont des propriétés semblables à celles obtenues à partir de saccades réelles. Concrètement, les longueurs et les orientations des marches suivent des distributions expérimentales des amplitudes et des directions de saccades et ainsi, les petites marches sont engendrées plus souvent que les grandes marches. Une séquence artificielle de saccades est ainsi générée par le modèle à partir d'un même point de départ que dans une expérience menée en parallèle. Les résultats ont montré que la distribution spatiale des fixations générées par le modèle de marche aléatoire ne ressemble pas à celle expérimentale et ne présente pas de biais de centralité. Pour tester l'hypothèse de forte distribution des caractéristiques saillantes au centre de l'image, Tatler a choisit deux bases d'images : l'une avec les caractéristiques saillantes au centre et l'autre avec les caractéristiques saillantes à la périphérie. Les deux distributions spatiales des fixations obtenues sur ces deux bases d'images présentent un biais de centralité. Cela implique que le biais de centralité ne provient pas simplement de la concentration des objets saillants au centre de la scène.

Enfin, Tatler [Tatler, 2007] explique le biais de centralité par le fait que le centre de la scène apporte des avantages stratégiques pour les sujets :

- Le centre de la scène pourrait être la position optimale pour extraire de l'information.
- Même si le centre n'est pas riche en information, il est considéré comme la position optimale à partir de laquelle on extrait un maximum d'information de la scène.
- Le biais de centralité est la conséquence de la tendance à mettre l'œil au centre de son orbite ; cette position de l'œil coïncide avec le centre de la scène.

1.4.2 Distributions concernant les saccades et les fixations

Lorsque l'on étudie les mouvements oculaires de sujets visionnant une image, on peut observer des tendances globales sur les saccades et les fixations. La majorité des saccades ont de petites amplitudes. Ainsi, la distribution des amplitudes de saccades n'est pas uniforme mais s'apparente à une loi de Poisson [Bahill et al., 1975; Gajewski et al., 2005; Tatler et al., 2006]. Il est intéressant de noter que l'étalement de la distribution varie avec la tâche ; il y a plus de longues saccades dans une expérience de recherche visuelle que dans une expérience d'exploration libre [Tatler

et al., 2006]. Le mode de la distribution de Poisson se situe entre 2° et 4° pour l'exploration libre et entre 4° et 6° pour la recherche visuelle.

La distribution des directions de saccades n'est pas uniforme non plus. Les saccades sont effectuées majoritairement dans les directions horizontale et verticale [Tatler and Vincent, 2008].

D'autres études ont regardé les relations entre des saccades ou des fixations consécutives. Motter et Belky [Motter and Belky, 1998] ont remarqué que de longues saccades sont suivies par de courtes saccades et *vice versa* dans une expérience de recherche de stimuli artificiels. Dans [Tatler and Vincent, 2008], Tatler a observé pour des scènes naturelles, une saccade longue ($> 7°$) est précédée par une saccade plus courte. De plus, une saccade très courte ($1-3°$) est également précédée par une autre saccade courte ($< 5°$), ce qui forme une séquence de saccades courtes. Cela est contradictoire avec la conclusion de Motter et Belky [Motter and Belky, 1998]. Néanmoins, il faut noter que les stimuli dans les deux expériences sont différents. Pour les stimuli artificiels, il y a souvent peu de cibles, et ces cibles sont nettement espacées. En revanche, pour les scènes naturelles, les zones d'intérêt sont nombreuses et certaines peuvent être proches.

Concernant les durées de fixations consécutives, des études dans la littérature ont montré des tendances différentes. Les résultats dans [Tatler and Vincent, 2008] ont révélé qu'une fixation de 100 à 200 ms a tendance à être précédée par une fixation courte (< 220 ms) tandis qu'une fixation de moins de 100 ms ou de plus de 200 ms est précédée par une longue fixation (> 220 ms). En outre, Tatler a observé une augmentation des durées de fixations au cours du temps. Ce résultat a également été observé dans [Antes, 1974; Pannasch et al., 2008] sur des scènes naturelles.

Des études ont également analysé la relation entre les amplitudes de saccades et les durées de fixations. Dans [Pelz and Canosa, 2001], ils ont montré une faible corrélation entre la durée de fixation et l'amplitude de la saccade suivante. Selon [Velichkovsky et al., 2005], en traçant l'amplitude de la saccade suivante en fonction de la durée de fixation courante, on observe une forte diminution de l'amplitude de la saccade autour de la durée de fixation de 180 ms. Ainsi, il a séparé les fixations en deux catégories : les fixations courtes et les fixations longues. La première catégorie représente des durées de fixations inférieures à 180 ms ; la deuxième des durées de fixations supérieures à 180 ms. Un résultat similaire est aussi obtenu dans [Tatler and Vincent, 2008]. De plus, dans cette étude, les fixations dont la durée est inférieure à 180 ms sont encore divisées en deux autres catégories à l'aide du seuil de 80 ms. Ils montrent ainsi que les fixations de durée entre 80 ms et 180 ms sont suivies par une longue saccade, et les fixations de durée très courte (< 80 ms) ou longue (> 180 ms) sont suivies par une saccade plus courte.

Pour être plus exhaustif, l'étude de Tatler [Tatler and Vincent, 2008] examine aussi la relation entre l'amplitude de saccade et la durée de la fixation suivante. Les résultats ont montré que la connaissance de l'amplitude d'une saccade ne permet pas de prédire la durée de la fixation suivante. Par contre, on peut utiliser la durée

d'une fixation pour caractériser la saccade précédente : une fixation d'une durée comprise entre 200 et 300 ms a tendance à être précédée par une saccade longue tandis qu'une fixation plus courte ou plus longue est précédée par une saccade plus courte.

1.4.3 Modes d'exploration : "ambiant" et "focal"

Les distributions des durées de fixations et des amplitudes de saccades au cours du temps peuvent être liées à deux modes d'exploration de scènes : "ambiant" et "focal" [Trevarthen, 1968; Velichkovsky et al., 2005]. Le mode "ambiant" représente une exploration dans un large espace du champ visuel. Par contre, le mode "focal" représente une exploration sur une petite zone. Au niveau temporel, l'exploration "focale" est plus longue. Dans [Pannasch et al., 2008], ils parlent d'autres dichotomies semblables aux modes "ambiant" et "focal" comme *"noticing"* et *"examining"* ou *"spacial"* et *"figural"*.

Les deux modes d'exploration "ambiant" et "focal" peuvent correspondre respectivement à deux périodes durant l'exploration d'une scène : fixations plus courtes, saccades plus longues au début de l'exploration et fixations plus longues, saccades plus courtes à la fin de l'exploration [Antes, 1974; Pannasch et al., 2008]. Dans [Tatler and Vincent, 2008], les deux modes d'exploration sont également confirmés, mais ils ne trouvent pas deux périodes temporelles séparées pour ces deux modes. Tandis qu'une période de courtes saccades correspondant au mode "focal" est confirmée, une période de longues saccades ne l'est pas. De plus, le mode "ambiant" n'est pas limité au début de l'exploration mais existe durant toute l'exploration.

2 Programmation de saccade

Comme nous l'avons vu, l'exploration d'une scène s'effectue en échantillonnant le champ visuel à l'aide des séquences saccade-fixation pour placer des zones d'intérêt sur la fovéa ; là où une analyse détaillée a lieu. Nous allons dans cette section nous intéresser à la question : "Combien de saccades sont programmées à partir d'une fixation ?"

Nous appelons la position à partir de laquelle la ou les saccades suivantes sont programmées et effectuées : le "point de vue". Ce point de vue correspond aussi à la fixation courante.

2.1 Programmation de saccades en parallèle

La programmation de saccades en parallèle concerne le fait qu'à partir d'un point de vue, on peut programmer les deux saccades suivantes ; autrement dit, la programmation des deux saccades se recouvre temporellement. Ce mécanisme permet de réduire le temps entre les deux saccades. Des expériences de la littérature

soutiennent ce mécanisme. Dans des tâches de recherche de cible [Hooge and Erkelens, 1998], les sujets semblent avoir tendance à préparer à partir d'une fixation plus qu'une saccade. Bien que les sujets aient été explicitement demandés d'effectuer une saccade dans une direction, ils n'arrivent pas toujours à le faire. Seulement 65 - 80% des mouvements oculaires sont dans bonnes directions. Les travaux de Hooge et Erkelens [Hooge and Erkelens, 1998] ainsi que ceux de Zelinsky [Zelinsky, 1996] ont indiqué que les gens effectuent plus de saccades que nécessaire. Cela semble anormal quand on sait que le système visuel fonctionne de manière à traiter le plus efficacement possible les informations de l'extérieur en disposant des ressources matérielles limitées. Cette irrégularité n'est expliquée que par la programmation de saccades en parallèle qui suppose que lorsque l'exécution d'une saccade est en cours, la préparation de la saccade suivante a déjà commencé [McPeek et al., 2000].

Dans [Becker and Jürgens, 1979], Becker et Jurgens ont également observé une programmation de saccades en parallèle lors d'une expérience particulière en deux étapes. Dans cette expérience (Fig. 1.2), il est demandé à un sujet d'effectuer deux saccades. Il est supposé que la première saccade est provoquée par le premier stimulus, la seconde saccade par le second stimulus. Dans ce cas, la latence B de la deuxième saccade dans la figure 1.2 est constante et indépendante de la préparation de la première saccade. Ce résultat soutient l'idée que la programmation de la deuxième saccade commence avant la fin de la première saccade. La deuxième saccade a donc été programmée parallèlement à la programmation de la première saccade.

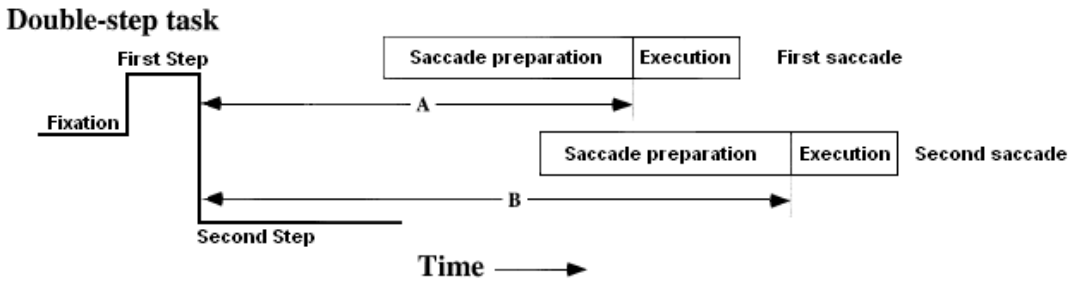


FIG. 1.2 – Illustration d'une séquence de l'expérience en deux étapes de Becker et Jurgens. La latence (B) de la deuxième saccade est constante (voir texte)(extrait de [McPeek et al., 2000]).

La programmation de saccades en parallèle est également confirmée dans une expérience de recherche visuelle [McPeek et al., 2000]. Dans cette expérience (Fig. 1.3), à chaque essai, il y a une cible et deux distracteurs ; ces distracteurs ont la même couleur mais différente de la couleur de la cible. Il est demandé au sujet d'effectuer une saccade vers la cible. A chaque essai, la couleur de la cible est aléatoire (rouge ou vert) ; les distracteurs étant de l'autre couleur. Quand la couleur de la cible dans un essai est différente de celle de l'essai précédent, des sujets se trompent et font une saccade vers un distracteur, ce qui avait été observé dans [McPeek et al., 1999]. Ensuite, une deuxième saccade est effectuée vers la cible après un intervalle très court. Ce phénomène implique que la deuxième saccade a été programmée en parallèle de

la première saccade. La programmation de saccades en parallèle caractérisée par un court intervalle inter-saccade est aussi observée dans d'autres expériences de recherche visuelle [Viviani and Swensson, 1982; Theeuwes et al., 1998] ou lors de la lecture [Morrison, 1984].

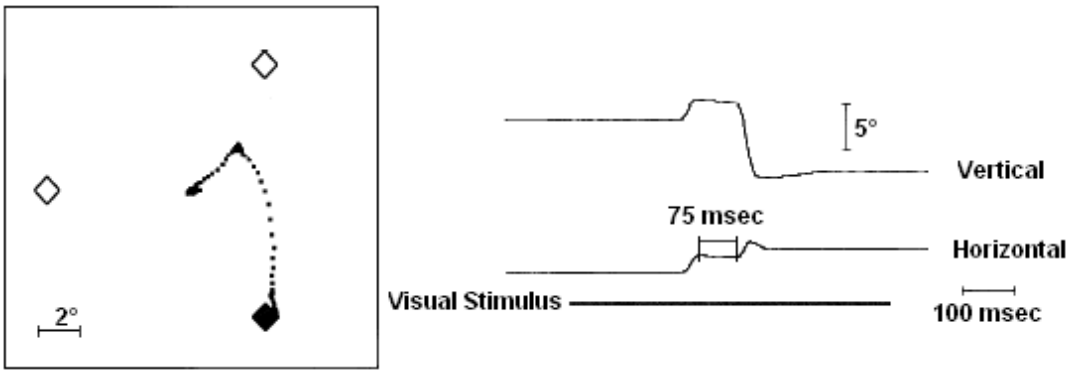


FIG. 1.3 – Illustration de la programmation de saccades en parallèle. A gauche : la première saccade se dirige vers le distracteur, la deuxième vers la cible. Les courbes en pointillé représentent des positions de l'œil au cours du temps. A droite : les positions horizontale et verticale de l'œil au cours du temps. L'intervalle très court entre les deux saccades démontre la programmation en parallèle [McPeek et al., 2000].

De plus, pour vérifier que la deuxième saccade est programmée en même temps que la première, McPeek [McPeek et al., 2000] veut s'assurer qu'un changement de la position de la cible pendant l'exécution de la première saccade ne modifie pas la deuxième. Il reprend l'expérience ci-dessus en modifiant la position de la cible de manière que lorsqu'un sujet fait une saccade vers un distracteur, les positions de la cible et de l'autre distracteur sont échangées. Les résultats expérimentaux ont montré que dans la plupart des cas (90%), les sujets dirigent la deuxième saccade vers la position ancienne de la cible.

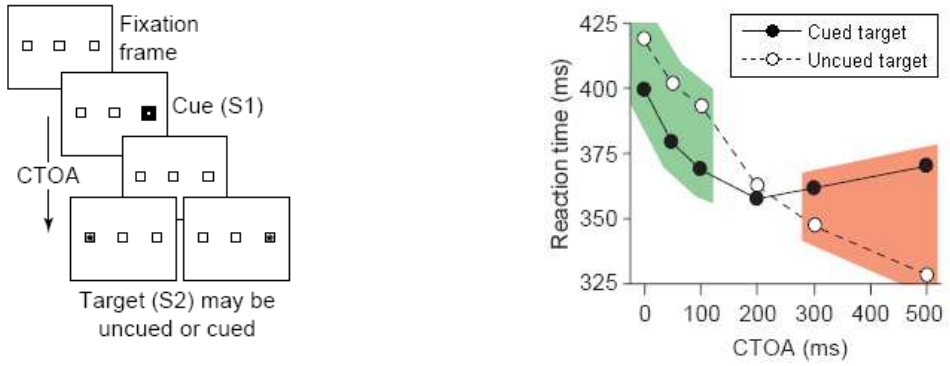
Les conclusions sur la programmation de saccades en parallèle sont élaborées à partir des expériences utilisant des stimuli artificiels et simples [Becker and Jürgens, 1979; Hooge and Erkelens, 1998; McPeek et al., 2000]. Le fait que ce mécanisme existe aussi dans des cas de scènes naturelles n'est pas immédiat et n'a pour le moment pas été testé. Contrairement au cas des stimuli simples, il y a beaucoup plus d'information dans une scène naturelle.

2.2 Inhibition de retour

L'inhibition de retour (IOR - *"Inhibition Of Return"*) correspond au mécanisme dirigeant les yeux vers une nouvelle position plutôt que celle que l'on vient de fixer. Dans cette situation, on privilégie les nouvelles zones tandis que les zones déjà regardées sont évitées. Le phénomène IOR est observé pour la première fois en 1984 par

Posner and Cohen [Posner and Cohen, 1984] puis ensuite par d'autres auteurs [Tassinari et al., 1987]. Le terme "inhibition de retour" est proposé par Posner et ses collègues [Posner et al., 1985] et puis largement utilisé dans la neuroscience cognitive et la psychologie expérimentale [Lupianez et al., 2006].

Dans l'expérience de Posner [Posner and Cohen, 1984] dans un premier temps, un indice est flashé à la périphérie de l'écran. Ensuite, une cible apparaît soit à la même position soit à l'opposé (Fig. 1.4a) et le sujet doit aller fixer cette cible. Selon la durée entre l'apparition de l'indice et celle de la cible (CTOA - *"cue-target onset asynchrony"*), Posner observe différents comportements des sujets. Si cette durée est courte (< 100 ms), le temps de réaction est plus court si la cible apparaît à la même position que l'indice. Par contre, à partir de 300 ms, le temps de réaction est plus long si la cible apparaît à la position de l'indice (Fig. 1.4b). De plus, dans [Samuel and Kat, 2003], Samuel et Kat ont étudié l'inhibition de retour pour la durée CTOA allant jusqu'à plus de 3 s. Le résultat de cette étude ont montré que l'inhibition de retour est stable pour un CTOA de 300 à 1600 ms.



(a) (b)

FIG. 1.4 – Expérience représentant l'inhibition de retour : (a) Déroulement de l'expérience : un carré présenté à la périphérie suivi par une cible qui peut être à la même position que le carré ou à la position opposée ; (b) Temps de réaction (à la cible) en fonction de la durée entre l'apparition du carrée et celle de la cible (CTOA - *"cue-target onset asynchrony"*) (extrait de [Klein, 2000]).

Depuis l'inhibition de retour est également utilisée dans les modèles d'attention visuelle comme celui d'Itti [Itti et al., 1998]. Ainsi le mécanisme IOR permet de prédire à partir d'une carte de saillance une séquence de fixations. Une fixation est choisie comme la position où la saillance est maximale après voir masqué les fixations choisies précédemment.

3 Modèle d'attention visuelle

Nous savons que la perception d'une scène visuelle est effectuée à l'aide des mouvements oculaires en utilisant des saccades et des fixations. Grâce aux saccades, les

yeux se déplacent vers des zones d'intérêt, l'une après l'autre, dans une scène. Ce sont à ces zones que la plupart de l'information visuelle est extraite pour la perception. Comment prédire ces zones ? C'est la question à laquelle plusieurs études ont tenté de répondre en proposant des modèles d'attention visuelle.

En s'appuyant sur les facteurs influençant les mouvements oculaires déjà révélés dans la littérature, un grand nombre d'études propose des modèles qui permettent de prédire des fixations lors de l'exploration de scènes. Ainsi, les modèles d'attention visuelle sont divisés en deux catégories principales : modèle ascendant (*"Bottom-Up"*) piloté par des facteurs de bas niveau et modèle descendant (*"Top-Down"*) piloté par des facteurs de haut niveau. En réalité, dans les modèles descendants, les facteurs de haut niveau sont souvent utilisés pour moduler la contribution des facteurs de bas niveau à l'attention visuelle. Nous utiliserons le terme "saillance" pour les processus attentionnels associés aux facteurs de bas niveau et "pertinence" pour les processus attentionnels associés aux facteurs de haut niveau[3]. Ainsi, un modèle descendant comporte à la fois la saillance et la pertinence tandis qu'un modèle ascendant ne concerne que la saillance et est aussi appelé "modèle de saillance".

3.1 Exemples de modèles ascendants

Ce type de modèle d'attention visuelle s'appuie sur l'hypothèse que l'attention est attirée par les caractéristiques de bas niveau des stimuli [Reinagel and Zador, 1999; Parkhurst et al., 2002; Parkhurst and Niebur, 2004; Tatler et al., 2005; Baddeley and Tatler, 2006]. Avec cette hypothèse, le modèle ascendant permet de prédire les fixations dans des conditions très contrôlées. Ce sont les premières fixations effectuées pendant une durée très courte (1-2 s) après l'apparition des stimuli et pour une exploration libre de scènes visuelles (aucune tâche n'est demandée au sujet). Ces conditions ont pour objectif de limiter le plus possible l'influence des facteurs de haut niveau.

Nous commençons par le modèle de Koch et Ullman [Koch and Ullman, 1985] qui est considéré comme le premier modèle d'attention visuelle. Pour ce modèle conceptuel, nous nous concentrons sur son architecture qui a influencé beaucoup d'autres modèles. Pour les modèles héritant de ce modèle, nous présentons plus en détails leurs implémentations.

3.1.1 Le modèle de Koch et Ullman

Ce modèle, proposé par Koch et Ullman [Koch and Ullman, 1985], est inspiré par les études de Treisman et Gelade en 1980 sur la théorie de l'intégration des caractéristiques pour l'attention visuelle [Treisman and Gelade, 1980]. Selon cette théorie, l'attention visuelle est guidée par la combinaison des caractéristiques de bas niveau comme l'intensité de luminosité, la couleur et l'orientation. Ainsi, dans ce premier modèle de Koch and Ullman (Fig. 1.5), une image d'entrée est décomposée

[3]La saillance et la pertinence sont appelées respectivement la saillance physique et la saillance cognitive selon Landragin [Landragin, 2004].

en plusieurs cartes, une carte par caractéristique de bas niveau. Ensuite, dans ces cartes, les positions saillantes émergent en supposant que la saillance d'une position dépend de sa différence par rapport aux positions voisines. Finalement, les cartes de caractéristique sont sommées pour créer la carte unique qui s'appelle la carte de saillance (*"saliency map"*). Cette carte de saillance, combinée avec le mécanisme WTA (*"Winner-Take-All"*), permet de prédire les positions que les sujets fixent. Le maximum de la carte correspondra à la prédiction de la première fixation, puis le maximum suivant correspondra à la prédiction de la deuxième fixation, etc. Désormais, la notion de carte de saillance est largement utilisée dans des études concernant le modèle d'attention visuelle.

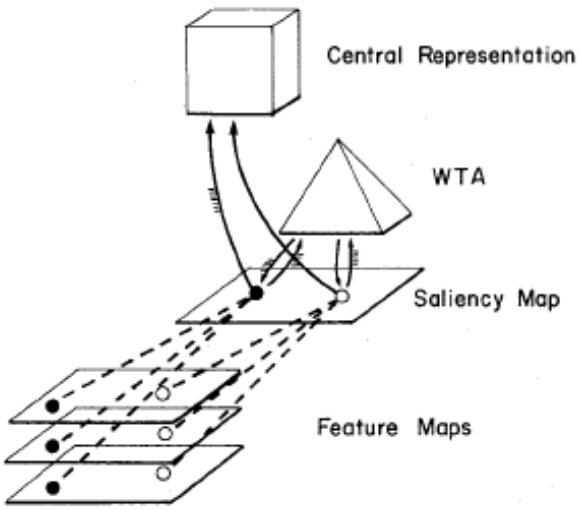


FIG. 1.5 – Le modèle d'attention visuelle de Koch et Ullman [Koch and Ullman, 1985]. La carte de saillance (*"saliency map"*) est créée par la fusion des cartes de caractéristique (*"feature maps"*) de bas niveau.

Le modèle de Koch et Ullman joue un rôle important en présentant le modèle de base sur lequel s'appuient de nombreux modèles d'attention visuelle. Les modèles développés à partir du modèle de Koch et Ullman apportent des améliorations au niveau de l'implémentation mais conservent l'architecture générale de celui-ci.

3.1.2 Le modèle d'Itti

Itti et collaborateurs [Itti et al., 1998] ont développé le modèle ascendant d'attention visuelle le plus répandu aujourd'hui (Fig. 1.6).

Comme le modèle original de Koch et Ullman, le modèle d'Itti décompose un stimulus visuel en caractéristiques visuelles de bas niveau comme l'orientation, l'intensité et la couleur. L'intensité correspond à la valeur moyenne des trois canaux r, g, b représentés dans l'espace RGB :

$$I = \frac{r + g + b}{3}$$

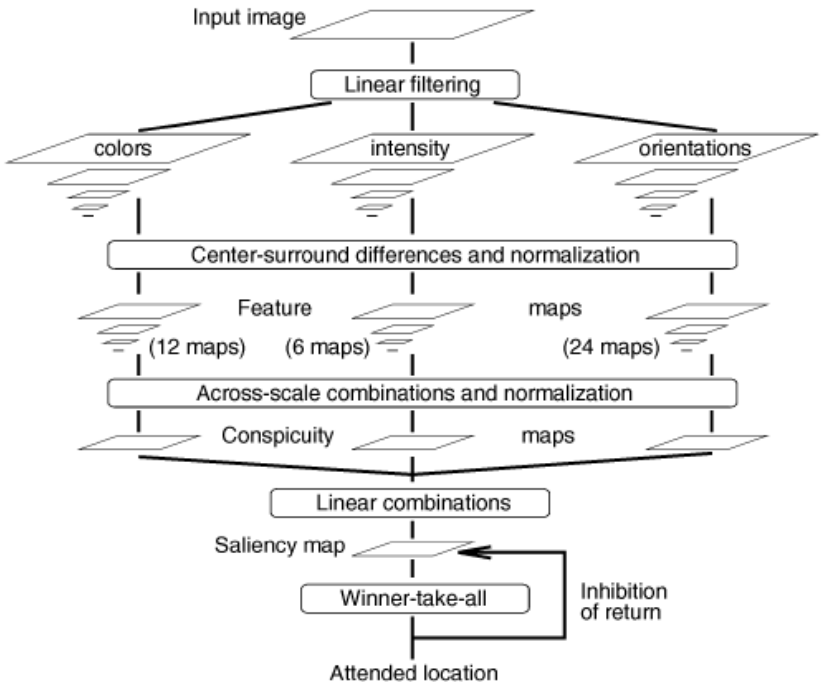


FIG. 1.6 – Le modèle d'attention visuelle proposé par Itti [Itti et al., 1998]

L'intensité I est ensuite décomposée par une pyramide passe-bas multirésolution à 8 niveaux (niveau 0 représente la carte d'intensité initiale). Ainsi, on obtient une pyramide $I(\sigma)$ où σ représente la résolution, $\sigma \in [0..8]$.

Les quatre couleurs R (rouge), G (vert), B (bleu) et Y (jaune) sont extraites selon les équations suivantes :

$$\begin{aligned} R &= r - \frac{g+b}{2} \\ G &= g - \frac{r+b}{2} \\ B &= b - \frac{r+g}{2} \\ Y &= \frac{r+g}{2} - \frac{|r-g|}{2} - b \end{aligned}$$

Comme l'intensité, chaque couleur est décomposée par une pyramide passe-bas. Ainsi, il y a 4 pyramides $R(\sigma)$, $G(\sigma)$, $B(\sigma)$, $Y(\sigma)$ pour les 4 couleurs.

Pour la caractéristique "orientation", elle est extraite de l'intensité I par des pyramides de Gabor $O(\sigma, \theta)$ où $\sigma \in [0..8]$ représente la résolution de la pyramide et $\theta \in \{0°, 45°, 90°, 135°\}$ l'orientation.

Ensuite, le contraste est extrait en effectuant la différence entre les valeurs à différents niveaux d'une pyramide. Pour l'intensité, les valeurs des niveaux de résolution plus fine c sont soustraites aux valeurs des niveaux de résolution plus grossière s :

$$\mathcal{I}(c, s) = |I(c) \ominus I(s)| \tag{1.1}$$

avec $c = \{2, 3, 4\}$, $s = c + \delta$ et $\delta = \{3, 4\}$. L'opérateur $\ominus$ représente la soustraction des valeurs à deux niveaux différents d'une pyramide ; cette soustraction nécessite une interpolation de la carte $I(s)$ pour qu'elle puisse avoir la même taille que $I(c)$. Ainsi, pour l'intensité, on obtient 6 cartes de trait (*"feature maps"*) $\mathcal{I}(c, s)$.

Normalisation Les cartes de trait seront sommées en vue de la création de la carte de saillance. Alors que la dynamique de ces cartes peut être différente car elles proviennent de différentes caractéristiques, il est nécessaire d'avoir une normalisation. De plus, cette normalisation renforce les cartes de trait qui ont un petit nombre de pics et diminue celles qui ont beaucoup de pics équivalents. Ainsi, la normalisation de chaque carte de trait est effectuée de la manière suivante :

- Normaliser chaque pixel entre $[0, M]$. Ainsi, la valeur maximale globale de chaque carte est M.
- Calculer la valeur moyenne des maxima locaux m.
- Multiplier la carte par $(M - m)^2$.

Fusion Pour chaque caractéristique (intensité, orientation ou couleur) toutes les cartes de trait sont fusionnées pour créer une carte de caractéristique (*"conspicuity map"*). Cette carte est aussi normalisée par la normalisation décrite ci-dessus. Enfin, la carte de saillance finale est construite en fusionnant les trois cartes de caractéristique.

En résumé, le modèle d'Itti n'est pas complexe et il est efficace au niveau du temps de calcul et de la qualité de la carte de saillance. De plus, la carte de saillance peut être combinée avec les mécanismes de WTA (*"Winner-Take-All"*) et d'IOR (*"Inhibition Of Return"*) pour choisir les fixations au cours du temps. La première fixation est choisie comme le maximum de la carte de saillance. Cette position est ensuite masquée avant de chercher le maximum suivant pour la deuxième fixation. Grâce à ses avantages, le modèle d'Itti est souvent repris dans d'autres études concernant l'attention visuelle ou la reconnaissance d'objet [Miau and Itti, 2001; Walther et al., 2002; Dhavale and Itti, 2003; Peters and Itti, 2008].

Bien que le modèle d'Itti soit conçu en imitant le fonctionnement du système visuel (cf. chapitre 2), l'aspect biologique modélisé reste limité. La normalisation utilisée dans le modèle est loin d'être justifiée par des propriétés biologiques. Dans [Itti and Koch, 2001], ils ont amélioré l'étape de normalisation en utilisant le filtre DoG (*"Difference Of Gaussians"*) plus biologiquement plausible. Ce filtrage est effectué avec un certain nombre d'itérations pour une carte afin de renforcer la saillance des positions différentes des positions voisines. Le principe de ce filtrage peut être expliqué par les champs récepteurs *"center-surround"* de certaines cellules dans le système visuel (cf. chapitre 2).

3.1.3 Le modèle de Le Meur

Dans le contexte de modèle ascendant, Le Meur [Le Meur et al., 2006] a proposé un autre modèle de saillance qui est également inspiré de l'architecture du modèle de Koch et Ullman (Fig. 1.7).

Dans le modèle de Le Meur, les caractéristiques de bas niveau sont représentées dans un espace psycho-visuel. Ainsi, une image en couleur est décomposée en trois composantes : une composante achromatique A et deux composantes chromatiques Cr_1, Cr_2. Ces deux composantes chromatiques représentent également l'opposition de couleurs comme dans le système visuel humain. Chacune des trois composantes est normalisée en appliquant la fonction CSF (*"Contrast Sensitivity Function"*) correspondante, qui représente la sensibilité au contraste en fonction de la fréquence spatiale et de l'orientation. Ainsi, chaque composante, A, Cr_1 ou Cr_2, est pondérée par une CSF spécifique.

Ensuite, les composantes sont décomposées en différentes fréquences spatiales et orientations (Fig. 1.7). Concrètement, il y a 17 canaux distribués dans 4 bandes de fréquences pour la composante achromatique et 5 canaux dans 2 bandes de fréquences pour chacune des composantes chromatiques. Puis, ces différents canaux

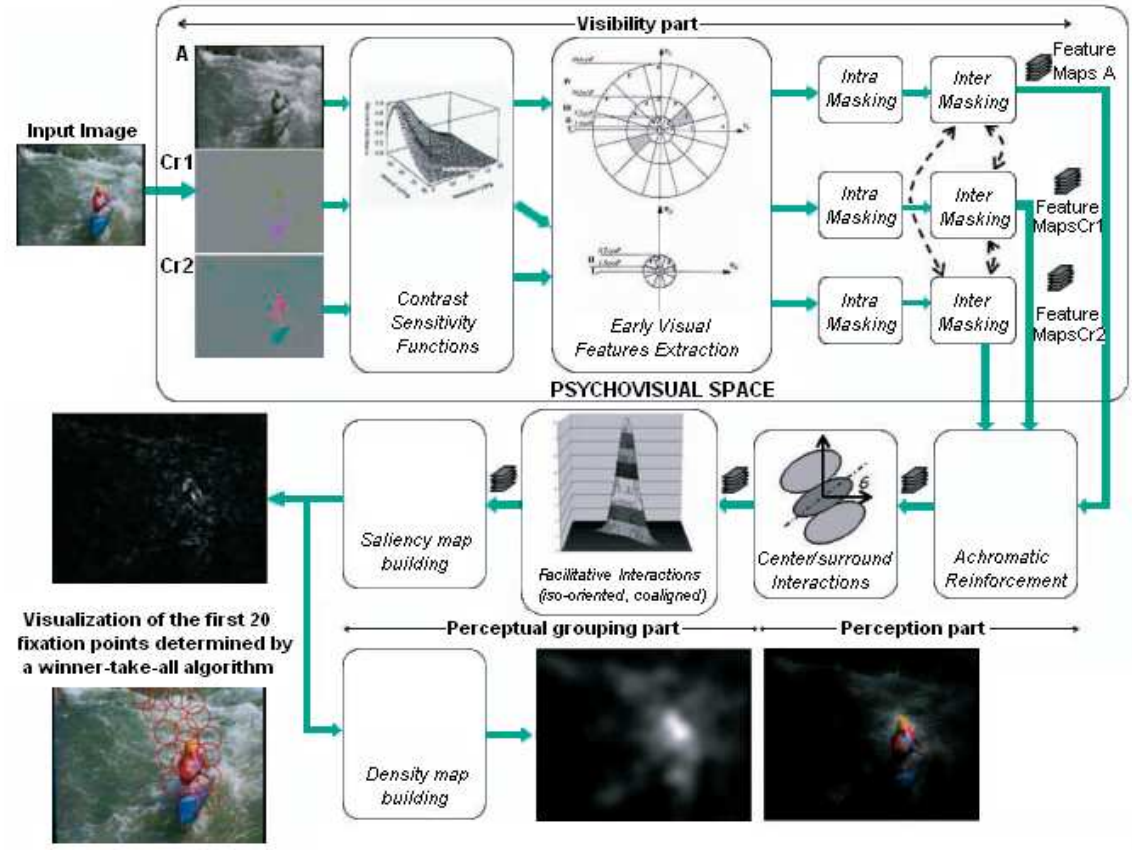


FIG. 1.7 – Le modèle d'attention visuelle proposé par Le Meur [Le Meur et al., 2006]

vont passer dans une étape de masquage qui modélise le fait que la réponse d'une cellule corticale dépend des réponses d'autres cellules. Il y a principalement deux types de masquage : intra et inter-composante.

Après avoir été représentées dans un espace psycho-visuel, les composantes sont soumises à des traitements au niveau perceptif. On note que ces traitements sont

appliqués principalement pour chacun des canaux de la composante achromatique. D'abord, la composante achromatique est renforcée par les composantes chromatiques aux positions dont le contraste de chrominance est fort. Ensuite, l'effet *"center-surround"* effectué par le filtre DoG est utilisé pour réduire la redondance d'information. Enfin des interactions permettant de renforcer des objets correspondant à une même orientation et alignés sont réalisées.

La carte de saillance est créée par la somme des canaux de la composante achromatique. De plus, cette carte de saillance est multipliée par un masque gaussien pour modéliser le fait que l'acuité visuelle est la plus forte au centre (où les yeux fixent) et diminue avec l'excentricité par rapport au centre.

L'avantage du modèle de Le Meur est de représenter en détails les composantes de bas niveau dans un espace psycho-visuel en utilisant des résultats d'expériences psychophysiques. Néanmoins, alors que ce modèle a abordé les traitements dans le cortex visuel, le traitement effectué par la rétine n'est pas exploité. De plus, les masquages intra-composante et inter-composante semblent complexes et coûteux car il y a beaucoup de paramètres libres difficiles à justifier, et enfin la correction de la carte de saillance par l'acuité est réalisée *a posteriori*.

3.2 Exemples de modèles descendants

L'influence des facteurs de haut niveau sur les mouvements oculaires a été observée dans plusieurs études. Dans les études de Buswell et Yarbus [Buswell, 1935; Yarbus, 1967], les positions fixées par les sujets sont dépendantes de la tâche. Cependant, contrairement aux facteurs de bas niveau, les facteurs de haut niveau sont beaucoup plus difficiles à modéliser. Cela explique qu'il existe peu de modèles descendants d'attention visuelle.

L'idée principale des modèles descendants est d'utiliser la pertinence associée aux facteurs de haut niveau pour moduler la saillance provenant des facteurs de bas niveau. L'intégration des facteurs de haut niveau est effectuée dans un contexte contrôlé, par exemple comme dans une tâche de recherche visuelle.

3.2.1 Le modèle de Navalpakkam & Itti

Le modèle descendant de Navalpakkam & Itti [Navalpakkam and Itti, 2005] est conçu pour une tâche de recherche visuelle dans des scènes naturelles. C'est un modèle complexe modélisant plusieurs aspects importants de la vision comme l'estimation de la pertinence d'une entité (homme, tête, main, etc) par rapport à la tâche, la pondération des caractéristiques de bas niveau en se basant sur la tâche, la détection et la reconnaissance d'objet, etc. Pour créer une carte d'attention visuelle finale, le modèle de Navalpakkam & Itti combine deux cartes différentes : une carte de saillance guidée par la tâche et une carte de pertinence représentant la pertinence des positions de la scène par rapport à la cible.

La carte de saillance guidée par la tâche a été tout d'abord proposée par Wolfe [Wolfe, 1994]. L'idée principale est d'utiliser les informations *a priori* d'une cible pour moduler les caractéristiques de bas niveau d'une scène afin de renforcer la probabilité de trouver cette cible. Par exemple, si l'on veut chercher une barre rouge, la pondération de la carte de la couleur rouge sera renforcée dans la fusion pour construire la carte de saillance. Le modèle de Navalpakkam & Itti suit également cette idée. Ainsi, la carte de saillance est créée selon le modèle proposé par Koch et Ullman [Koch and Ullman, 1985] en utilisant trois caractéristiques de bas niveau : l'intensité, la couleur et l'orientation. Ensuite, les caractéristiques similaires à celle de la cible sont renforcées. De plus, on renforce aussi les entités dans la scène qui sont similaires à la cible en utilisant la détection d'objet par apprentissage. Enfin, on obtient une carte de saillance guidée par la tâche de recherche d'une cible particulière.

L'estimation de la carte de pertinence fait appel à : la mémoire de travail et la mémoire à long terme (Fig. 1.8). La mémoire à long terme contient des connaissances de base comme la description et la relation des entités. La tâche est décrite sous forme *"qui est en train de faire quoi à qui"* et puis représentée par des mots-clés "objet", "sujet" et "action". Dans cette première version, de simples relations sont utilisées comme *"être un"*, *"partie de"*, *"contient"*, etc. Une liste de priorité est également créée pour estimer la pertinence d'une entité en se basant sur sa relation avec la cible. Par exemple, dans une tâche de recherche de main, on trouve un doigt et un homme. Alors, la pertinence du doigt (main *"contient"* doigt) est plus grande que celle de l'homme (main *"fait partie de"* l'homme). C'est la mémoire de travail qui estime la pertinence d'une entité en disposant d'un graphe représentant des entités de pertinence et leurs relations. Pour une entité donnée, la mémoire de travail cherche d'abord sa pertinence dans le graphe si cette entité y existe déjà, si non elle fait appel à la mémoire à long terme. Et puis, le graphe de la mémoire de travail est mis à jour à travers des entités reconnues. Ainsi, la mémoire de travail, combinée avec la mémoire à long terme, permet de construire une carte de pertinence qui estime la pertinence des positions spatiales de la scène.

Finalement, la carte d'attention visuelle finale est le produit pixel par pixel entre la carte de saillance guidée par la tâche et la carte de pertinence. Dans un premier test, le modèle n'effectue qu'une tâche de recherche d'une cible et espère la détecter le plus vite possible.

En résumé, le modèle de Navalpakkam & Itti permet de modéliser l'influence de la tâche lors d'une recherche visuelle pour des scènes naturelles. Bien que ce modèle ne soit pas encore complet (il y a quelques parties à implémenter), il représente une architecture intéressante du modèle intégrant des facteurs de haut niveau dans l'attention visuelle.

3.2.2 Le modèle de Peters & Itti

Dans la même motivation d'intégrer l'influence de la tâche au modèle d'attention visuelle, Peters et Itti [Peters and Itti, 2007] proposent un autre modèle simple en tenant compte des facteurs de bas niveau et de haut niveau. Ce modèle est conçu

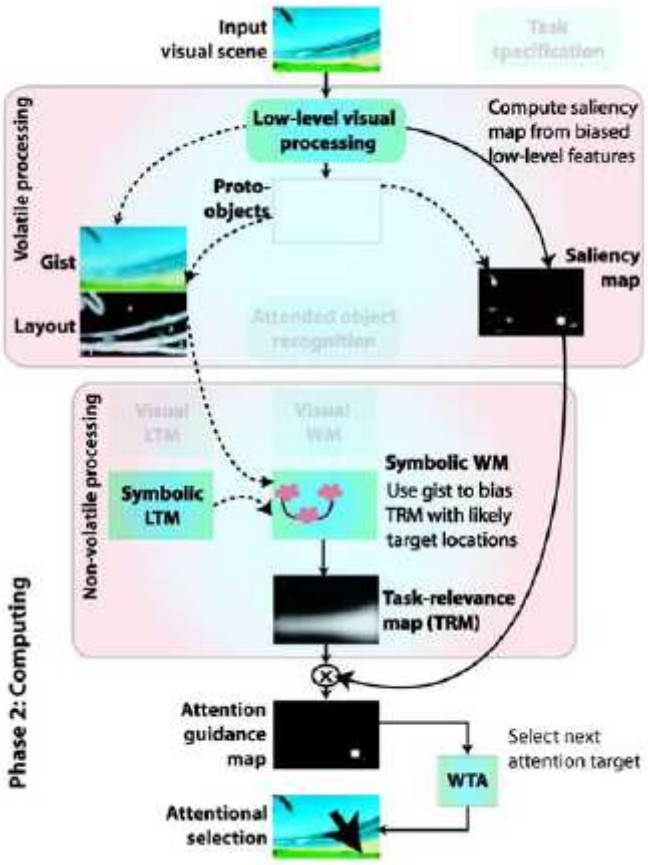


FIG. 1.8 – Un exemple de la combinaison de la carte de saillance guidée par la tâche et de la carte de pertinence. Les courbes en pointillé représentent les parties n'ayant pas été implémentées dans la première version du modèle de Navalpakkam & Itti [Navalpakkam and Itti, 2005].

pour prédire les mouvements oculaires des sujets dans un contexte interactif. Les fixations sont enregistrées lors de la visualisation des jeux vidéo.

Comme dans le modèle de Navalpakkam & Itti, le modèle de Peters & Itti combine également la carte de saillance et la carte de pertinence à la tâche. La carte de saillance se calcule selon le modèle de [Itti et al., 1998].

La carte de pertinence est obtenue par l'apprentissage sur une base de données des vidéos et des fixations enregistrées. L'objectif est de trouver la relation entre les caractéristiques d'une image (frame) et la carte de densité de fixations (la distribution des positions de fixations sur cette image). Après apprentissage, pour une image (ou une frame) de test, on peut calculer les caractéristiques de cette image et en déduire la carte de densité de fixations ou la carte de pertinence à la tâche.

La carte d'attention visuelle finale est la combinaison de la carte de saillance et de la carte pertinence en utilisant la multiplication pixel par pixel. Ce modèle est moins complexe que celui de Navalpakkam et Itti et plutôt computationnel. Il

permet également de montrer le rôle de l'intégration des facteurs de haut niveau dans le modèle d'attention visuelle.

3.2.3 Le modèle de Torralba

Récemment, une autre approche a été développée pour calculer la carte d'attention visuelle basée sur une modélisation statistique. L'idée principale de cette méthode est que la saillance d'un objet est inversement proportionnelle à sa probabilité d'apparition. Ainsi, un objet a plus de possibilité d'attirer les yeux lorsque la probabilité d'apparition de cet objet est faible. L'outil souvent utilisé dans cette approche est l'inférence Bayésienne [Oliva et al., 2003; Torralba, 2003a,b; Torralba et al., 2006; Itti and Baldi, 2006].

Ici, nous allons présenter le modèle descendant proposé par Torralba [Torralba et al., 2006] ; ce modèle est souvent cité dans la littérature pour l'approche statistique (Fig. 1.9). Le modèle de Torralba exploite le contexte de l'image pour moduler la saillance des caractéristiques locales de bas niveau dans une tâche de recherche visuelle (recherche d'un objet). En utilisant la formule de Bayes, Torralba a montré que la carte d'attention visuelle peut être calculée en recherchant le niveau d'attention de chaque position $\mathbf{x}$ suivant l'équation suivante :

$$S(\mathbf{x}) = \frac{1}{p(L|G)} p(\mathbf{x}|O = 1, G) \tag{1.2}$$

où $S(\mathbf{x})$ est le niveau d'attention de la position $\mathbf{x}$
L est le vecteur de caractéristique locale à la position $\mathbf{x}$
G est le vecteur de contexte de l'image
O représente l'objet recherché (la cible) ($O = 0$ représente l'absence de la cible dans l'image et $O = 1$ représente la présence de la cible dans l'image)

Ici, le facteur $\frac{1}{p(L|G)}$ représente la contribution à la saillance qui est inversement proportionnelle à la probabilité d'apparition d'un vecteur de caractéristique. Le vecteur de caractéristique locale L est calculé à partir des caractéristiques de bas niveau comme l'orientation, la couleur, l'intensité et le contraste et cela pour chaque position sur l'image. Le vecteur de contexte G est aussi calculé à partir des caractéristiques de bas niveau mais représente l'information globale de l'image concernant plusieurs fréquences spatiales et orientations. Le facteur $p(L|G)$ est estimé à l'aide d'un apprentissage.

Le deuxième facteur $p(\mathbf{x}|O = 1, G)$ de l'équation 1.2 représente la probabilité de trouver l'objet sous le contexte G pour chaque position $\mathbf{x}$. Ce facteur reflète la relation entre le contexte ou l'information globale de l'image et la position de la cible. Le facteur $p(\mathbf{x}|O = 1, G)$ est également estimé par apprentissage.

Ainsi, le niveau d'attention d'une position est la saillance associée aux caractéristiques locales et modulée par la pertinence associée au contexte. Pour une image de test donnée, le vecteur de contexte G de cette image est d'abord calculé. Pour chaque position $\mathbf{x}$ de l'image, le vecteur de caractéristique locale L est extrait. Ensuite, la saillance de chaque position $\mathbf{x}$ est représentée par $\frac{1}{p(L|G)}$. La pertinence au

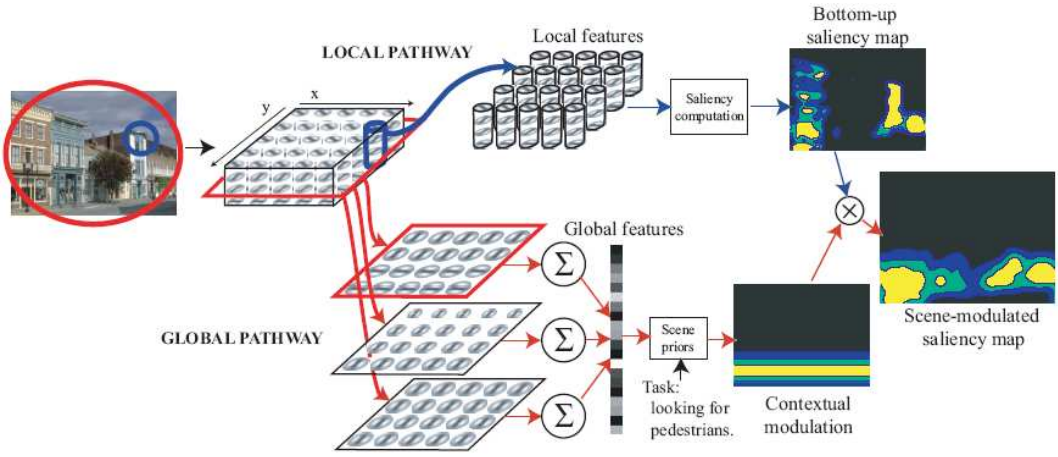


FIG. 1.9 – Le modèle d'attention visuelle de Torralba. La carte d'attention d'une image est créée en combinant deux voies séparées : une voie représentant l'information locale de l'image et l'autre l'information globale [Torralba et al., 2006].

contexte est calculée par $p(\mathbf{x}|O = 1, G)$. La carte d'attention visuelle finale peut donc être créée à partir des niveaux d'attention de toutes les positions de l'image selon l'équation 1.2.

Pour la tâche de recherche visuelle, le modèle de Torralba a montré que la carte de pertinence au contexte peut prédire les fixations mieux que la carte de saillance. De plus, la combinaison de ces deux cartes donne les meilleurs résultats.

3.3 Où se crée la carte d'attention visuelle ?

La question où dans le cortex se crée la carte d'attention visuelle n'a pas de réponse unanime. Dans les modèles de Koch et Ullman [Koch and Ullman, 1985] ou d'Itti [Itti et al., 1998], la carte de saillance est construite à partir des cartes de caractéristique sans toutefois préciser où ce processus a lieu. Quant aux recherches de Zhaoping [Zhaoping, 2002, 2005], toujours sur le modèle ascendant, la carte de saillance est supposée créée au niveau du cortex visuel primaire (V1) où les caractéristiques visuelles élémentaires sont analysées. De plus, d'autres études avancent aussi le traitement des informations visuelles dans d'autres parties du cortex comme V2 et leurs "feedback" vers V1 pour moduler des réponses des neurones dans V1 [Hansen et al., 2001; Stettler et al., 2002]. Néanmoins, dans [Stettler et al., 2002], les auteurs ont montré que le renforcement du contour est plus lié aux connexions dans V1 que le "feedback" de V2 vers V1.

Quand les facteurs de bas niveau et les facteurs de haut niveau sont à la fois pris en compte dans la création de la carte d'attention visuelle, cette combinaison est censée être effectuée dans les aires de plus "haut niveau" que V1. Dans [Mazer and Gallant, 2003], ils ont proposé une convergence des voies ascendante et descendante dans V4 qui guiderait la programmation oculomotrice lors d'une recherche visuelle. La même conclusion s'est avérée dans [Ogawa and Komatsu, 2004]. De plus,

d'autres aires du cortex sont supposées comme étant le lieu de la construction d'une carte d'attention visuelle : l'aire intrapariétale latérale [Gottlieb et al., 1998] ou le colliculus supérieur [Findlay and Walker, 1999].

4 Résumé

Dans un premier temps, nous avons regardé les mouvements oculaires lors de l'exploration de scènes visuelles. En étudiant ces mouvements enregistrés à partir des expériences psychophysiques, on a révélé certaines propriétés des mouvements oculaires et les facteurs les influençant.

Les sujets ont une tendance à regarder plus au centre qu'à la périphérie d'une image quelque soit la tâche. De plus, la distribution des amplitudes de saccades n'est pas uniforme mais contient majoritairement des petites saccades. Il y a également plus de saccades dans les directions verticale ou horizontale que dans une direction oblique. Plusieurs études ont confirmé deux modes d'exploration : "ambiant" et "focal". Le mode "ambiant" correspond à une exploration rapide dans un champ visuel large, caractérisée par des saccades longues et des fixations courtes. En revanche, le mode "focal" correspond à une exploration dans une petite zone en détails. Ce mode est souvent observé avec des saccades courtes et des fixations longues.

Les expériences psychophysiques permettent de diviser en deux groupes les facteurs influençant les mouvements oculaires. Le premier groupe concerne des facteurs de bas niveau dans une scène comme la luminance, la couleur, l'orientation et le contraste. Ces facteurs arrivent tôt dans le traitement du système visuel. Le deuxième groupe comporte les facteurs de haut niveau comme la sémantique, la mémoire ou la tâche. Contrairement au premier groupe, ces facteurs arrivent tard dans le traitement du système visuel. En général, ces deux facteurs coexistent et influencent la perception visuelle. Néanmoins, la contribution de chacun des facteurs dépend bien de l'expérience, sans tâche ou avec tâche, et de la durée d'exploration.

Les mouvements oculaires dans une scène statique se font grâce à des saccades. Comment se programment ces saccades ? Des études ont supposé la programmation de saccades en parallèle dans des expériences de recherche visuelle pour des stimuli simples. Cependant, pour des scènes naturelles où des stimuli sont beaucoup plus complexes, la question de programmation de saccade reste encore ouverte. La stratégie de programmation d'une seule saccade ne devrait pas être écartée dans ce cas-là.

Pour étudier les mouvements oculaires et la programmation de saccade par méthode computationnelle, un modèle d'attention visuelle est indispensable car il permet d'estimer les saccades ou fixations. Selon les deux groupes de facteurs qui peuvent influencer les mouvements oculaires, il y a également deux types de modèle d'attention visuelle : ascendant et descendant. Le modèle ascendant s'appuie sur les caractéristiques de bas niveau dans une image. Dans la littérature, les modèles de ce type suivent souvent l'architecture proposée par Koch et Ullman [Koch and Ullman, 1985] où la carte de saillance est la combinaison des caractéristiques de

bas niveau. Quant au modèle descendant, l'influence de la tâche ou du contexte est intégrée pour moduler la carte de saillance. Des résultats de la littérature ont montré une bonne performance des modèles descendants pour la recherche visuelle. Pourtant, dans une exploration libre de scènes naturelles, lorsqu'il n'y a pas de tâche spécifique, les facteurs de bas niveau peuvent jouer un rôle plus important, d'autant plus si on considère uniquement la prédiction des premières fixations. Dans ce cas, le modèle ascendant peut être suffisant pour modéliser l'attention visuelle.

Chapitre 2

Modèle d'attention visuelle proposé

1 Introduction

Lors de l'exploration d'une scène visuelle, les mouvements oculaires d'un sujet ne sont pas aléatoires mais dépendent, entre autres, des caractéristiques visuelles des stimuli, de la consigne donnée au sujet avant l'exploration, du sujet qui explore la scène, etc. Pour prédire ces mouvements oculaires, différents groupes de recherche proposent des modèles dits d'attention visuelle : des modèles dits "ascendants" ou *"Bottom-Up"* et des modèles dits "descendants" ou *"Top-Down"*.

Dans cette thèse, nous nous intéressons uniquement aux modèles ascendants, et nous proposons un modèle d'attention visuelle dans cette lignée. Ce modèle permet de prédire les zones saillantes (qui attirent le regard) dans des scènes naturelles. Ces zones prédites seront comparées aux régions fixées par un ensemble de sujets lors d'une expérience d'exploration libre de scènes utilisant l'oculométrie. L'exploration des scènes se fait sans consigne précise, pour ne pas contraindre les sujets et donc pour se rapprocher des modèles ascendants. De plus, le modèle que nous proposons a pour objectif de prédire uniquement les premières fixations : prédiction des mouvements oculaires pendant 1 à 2 s successivement à la présentation d'une scène. Des travaux [Tatler et al., 2005; Henderson et al., 1999] ont montré que les premières fixations sont majoritairement commandées par les caractéristiques des stimuli visuels de bas niveau et cela indépendamment de la tâche.

Comme nous l'avons décrit au chapitre 1, les modèles d'attention visuelle ascendants se basent sur le fonctionnement du système visuel humain pour prédire les zones d'attention (ou les zones saillantes) d'une scène visuelle [Koch and Ullman, 1985; Itti et al., 1998; Le Meur et al., 2006]. Dans ces modèles, la simulation du fonctionnement des cellules rétiniennes reste peu, voire pas abordée, alors qu'il a été montré que de nombreux traitements sont réalisés dès la rétine [Hérault, 2001] ; il peut donc être intéressant d'inclure, dans un modèle d'attention, la modélisation de ces traitements rétiniens.

Dans ce chapitre, nous présentons un modèle d'attention visuelle biologiquement

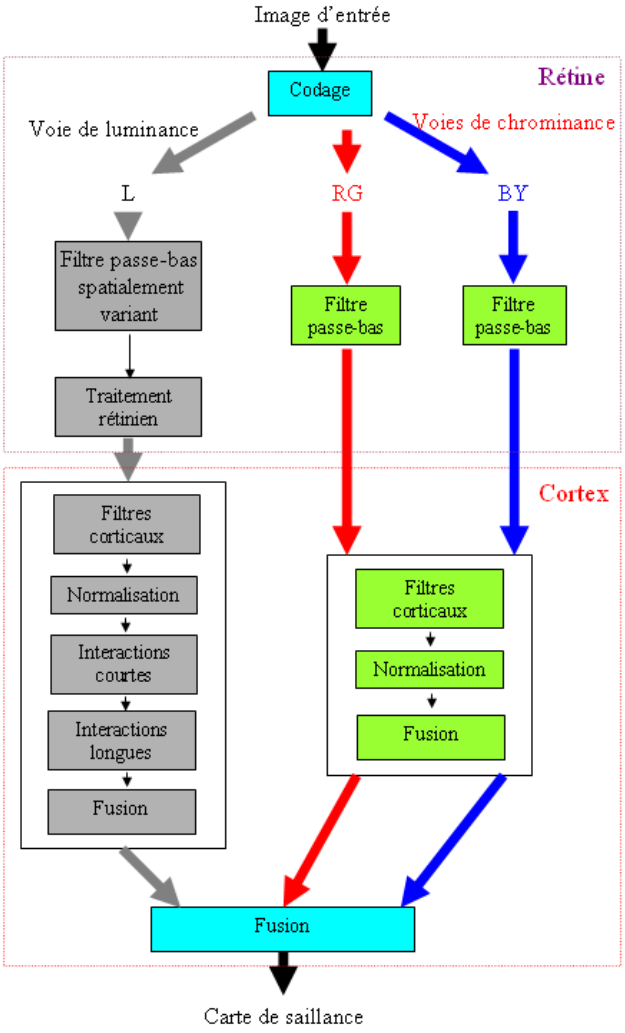


FIG. 2.1 – Schéma du modèle d'attention visuelle proposé. Ce modèle se décompose en deux étages principaux : le modèle de la rétine et le modèle du cortex visuel primaire. Ce modèle permet d'obtenir pour une image sa carte de saillance qui met en exergue les zones saillantes.

plausible qui imite en le simplifiant, le fonctionnement du système visuel depuis la rétine jusqu'au cortex visuel. Ce modèle est inspiré de l'architecture initialement proposée par Koch et Ullman [Koch and Ullman, 1985] et se distingue des autres modèles proposés dans la littérature par plusieurs aspects : l'importance accordée au traitement rétinien, la modélisation des filtres corticaux et des interactions entre ces filtres. Ce modèle schématisé à la figure 2.1 se décompose en deux grandes parties de traitement de l'information visuelle : une première qui concerne la modélisation du traitement réalisé par les cellules de la rétine et une deuxième celle du traitement réalisé par les cellules du cortex visuel primaire.

– Le modèle de la rétine permet tout d'abord de séparer l'information de luminance et l'information de chrominance (par opposition de couleurs). Ce modèle permet également un échantillonnage spatialement variant de l'information visuelle (haute résolution à l'endroit où l'œil se porte au niveau du point de fixa-

tion et résolution de plus en plus faible avec l'éloignement dans le champ visuel par rapport à ce point de fixation) ainsi qu'un réhaussement des contrastes en luminance et un blanchiment spectral.
– Le modèle cortical décompose l'information visuelle issue de la rétine en différents attributs élémentaires (ici orientations, fréquences spatiales) et par l'intermédiaire des interactions entre les filtres corticaux fait ressortir des zones de l'image différentes de leurs contextes (au sens des attributs).

Dans les paragraphes suivants, nous décrivons en détails les différents étages de ce modèle.

2 Rétine

2.1 Anatomie de la rétine

La rétine, tissu neuronal tapissant le fond de l'œil, est le premier étage du système visuel de traitement du flux de lumière entrant dans l'œil. Elle convertit les signaux lumineux en signaux nerveux transmis via le nerf optique aux cellules du cortex visuel. Cette conversion s'effectue à travers plusieurs couches neuronales (Fig.2.2). Ici, sans avoir l'ambition d'aller dans les détails de l'anatomie de la rétine, nous nous concentrons sur les principales cellules rétiniennes. C'est le fonctionnement des cellules décrites ci-dessous qui est simulé dans le modèle d'attention visuelle proposé.

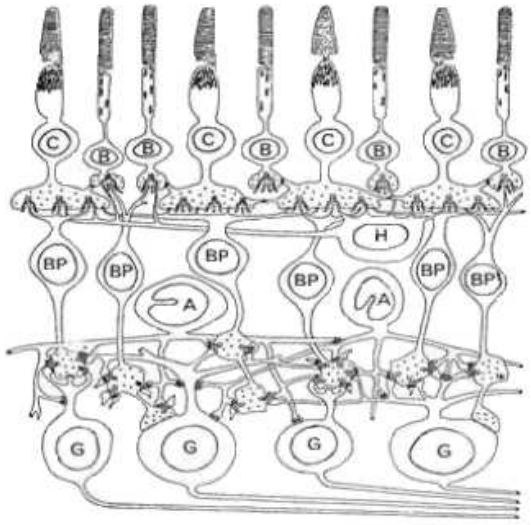


Fig. 2.2 – La structure de la rétine est caractérisée par différentes couches de cellules : Les photorécepteurs (cônes (C) et bâtonnets (B)), les cellules horizontales (H), les cellules bipolaires (BP), les cellules amacrines (A) et les cellules ganglionnaires (G) [Kowaliski, 1990].

2.1.1 Les photorécepteurs

Il existe deux types de photorécepteurs : les cônes et les bâtonnets. Les premiers, les *cônes*, sont impliqués dans la vision photopique (vision de jour) ; trois types de cônes se distinguent en fonction de leur sensibilité spectrale : ils correspondent grossièrement aux couleurs Rouge, Vert et Bleu. Les cônes se concentrent principalement au centre de la rétine (région appelée fovéa). Leur densité diminue lorsque l'excentricité (distance par rapport à la fovéa) augmente (Fig. 2.3). Les seconds, les *bâtonnets* sont responsables de la vision scotopique (vision de nuit) et sont majoritairement répartis à la périphérie de la rétine (Fig. 2.3).

Le rôle principal des photorécepteurs est de permettre le codage de l'information visuelle indépendamment de la luminosité ambiante (nous percevons les détails en plein jour mais également dans la pénombre). Ils adaptent pour cela leurs dynamiques de réponse en fonction de la luminosité ambiante moyenne.

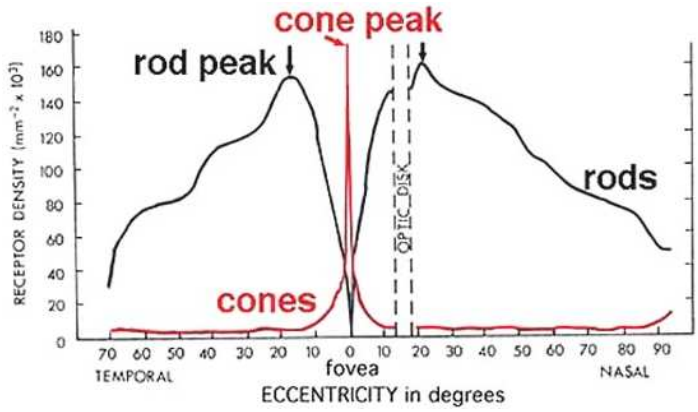


FIG. 2.3 – Evolution de la densité des photorécepteurs (cônes et bâtonnets) à la surface de la rétine en fonction de l'excentricité par rapport au centre de la rétine (fovéa) en degrés angulaires [Osterberg, 1935].

2.1.2 Les cellules horizontales

Les cellules horizontales tiennent leur nom de leur "forme" mais également de leur position dans la rétine. Elles connectent un ou plusieurs photorécepteurs et sont connectées entre elles. Elles effectuent le lissage de l'information transmise par les photorécepteurs [Hérault, 2001]. Ainsi, les cellules horizontales représentent l'information locale moyenne des stimuli et elles participent à l'adaptation des photorécepteurs à la luminosité ambiante.

2.1.3 Les cellules bipolaires

Avant de décrire les propriétés des cellules bipolaires, il est nécessaire d'aborder la notion de "champ récepteur". Cette notion de "champ récepteur" sera également

utilisée dans la suite pour décrire le fonctionnement des autres cellules rétiniennes ainsi que celui des cellules corticales. Le champ récepteur d'une cellule correspond à la région de l'espace visuel dans laquelle une stimulation entraîne une réponse de la cellule [Levine and Shefner, 1991]. Selon Hubel et Wiesel, le champ récepteur d'une cellule à un certain niveau dans le système visuel est construit à partir des champs récepteurs des cellules des niveaux précédents [Hubel and Wiesel, 1962].

Les cellules bipolaires reçoivent en signal d'entrée les sorties des photorécepteurs et des cellules horizontales. Leur champ récepteur comporte deux zones concentriques ; le centre est activé par les photorécepteurs et la périphérie inhibée par les cellules horizontales ou l'inverse. Selon la nature de leurs réponses, il y a deux types de bipolaires : les cellules "centre ON" et les cellules "centre OFF". Les cellules "centre ON" répondent quand un faisceau lumineux est présenté au centre de leur champ récepteur et les cellules "centre OFF" quand le faisceau est à la périphérie. Ainsi, les cellules bipolaires répondent au contraste du stimulus visuel.

2.1.4 Les cellules amacrines

Les cellules amacrines sont en contact avec les cellules bipolaires et les cellules ganglionnaires et jouent un rôle de lissage similaire à celui réalisé par les cellules horizontales. De plus, elles sont très sensibles aux variations temporelles et sont donc importantes dans la perception du mouvement.

2.1.5 Les cellules ganglionnaires

Les cellules ganglionnaires constituent l'étage de sortie de la rétine. Elles reçoivent les informations provenant des cellules bipolaires et des cellules amacrines et délivrent des réponses via le nerf optique au cortex visuel. Les cellules ganglionnaires ont des champs récepteurs similaires à ceux des cellules bipolaires avec deux zones concentriques. On distingue également les cellules "centre ON" et "centre OFF". De plus, il existe deux principaux types de cellules ganglionnaires : parvocellulaire (midget) et magnocellulaire (parasol) ; elles se distinguent par leurs réponses en fréquences spatiales et temporelles. Les informations de hautes fréquences spatiales de luminance et de faibles variations temporelles sont véhiculées par les midgets [Kandel et al., 2000]. De plus, ces cellules midget véhiculent l'information de couleur codée en opposition Rouge-Vert (RG) et Bleu-Jaune (BY) [Dacey, 1996; Dacey and Packer, 2003; Chatterjee and Callaway, 2003]. Quant aux parasols, elles sont sensibles à la luminance en basses fréquences spatiales mais en hautes fréquences temporelles.

2.2 Modélisation de la rétine

2.2.1 Introduction

Les connaissances en physiologie sur la rétine ont fourni les fondements pour modéliser son fonctionnement. Parmi les études sur la modélisation du traitement rétinien, nous pouvons citer les travaux effectués par Mead et son équipe [Mead and Mahowald, 1988; Mead, 1989; Mahowald and Mead, 1991] aux Etats-Unis à la fin des années 1980. Dans ces travaux, le fonctionnement de la rétine est modélisé par un circuit électrique avec des composants passifs (résistance, capacité). D'une part, le

modèle de Mead a été le précurseur de nombreux travaux sur les rétines artificielles implémentables sur silicium par des circuits VLSI (*"Very-large-scale integration"*). D'autre part, cela a montré tout l'avantage de formaliser les traitements rétiniens en utilisant les outils du traitement du signal. Ce modèle a inspiré plusieurs autres parmi lesquels figure celui de Beaudot [Beaudot et al., 1993; Beaudot, 1994; Hérault, 2001]. Dans son travail, Beaudot a modélisé en détails la rétine selon sa structure en couches. Le fonctionnement des cellules essentielles de la rétine a été modélisé par des filtrages spatio-temporels à l'aide de la transformée de Fourier pour la dimension temporelle continue et de la transformée en Z pour la dimension spatiale discrète dû à l'échantillonnage des photorécepteurs. L'avantage de ce modèle est de fournir pour la première fois une modélisation fonctionnelle assez complète de la rétine. Dans ses premiers travaux, seule la voie luminance a été modélisée. Ce modèle est complété avec les travaux de David Alleysson sur les voies chromatiques [Alleysson, 1999].

Dans ce chapitre, nous reprenons le modèle rétinien proposé par Beaudot. Dans un premier temps, puisque nous nous intéressons uniquement aux scènes statiques, nous nous focalisons uniquement sur la modélisation du filtrage spatial réalisé par la rétine. Néanmoins, dans un deuxième temps (cf. chapitre 6) nous utiliserons le modèle complet, spatio-temporel, pour étudier la dynamique temporelle des mouvements oculaires. Nous avons complété le modèle initial de Beaudot en ajoutant la modélisation de l'échantillonnage spatialement variant des photorécepteurs. Nous utilisons pour cela un filtrage passe-bas spatialement variant [Curcio et al., 1990; Goodchild et al., 1996]. A partir des travaux de David Alleysson [Alleysson, 1999], nous ajoutons également le traitement des voies chromatiques, mais dans une version simplifiée.

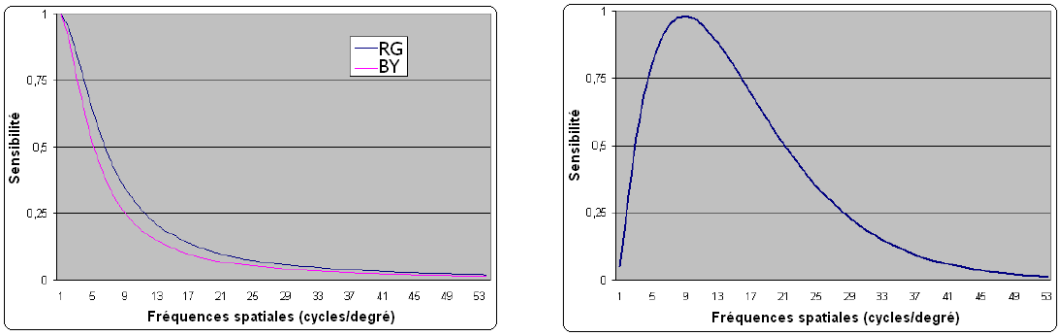
2.2.2 Codage de l'information en luminance et chrominance

Dans notre modèle, une image en couleur définie suivant les plans couleur rouge (R), vert (G) et bleu (B) est décomposée en une image de luminance et deux images de chrominance (Fig. 2.1). La luminance (L), représentée dans l'espace de couleur NTSC pour être adaptée au affichage, est calculée à partir des trois plans couleur R, G, B. Pour les voies chromatiques, leur codage dans le système visuel humain est simplifié en utilisant les oppositions de couleurs avec une voie Rouge - Vert (RG) et une voie Bleu - Jaune (BY) [Dacey, 1996; Dacey and Packer, 2003; Chatterjee and Callaway, 2003]. Ce codage est résumé par les équations suivantes :

$$\begin{array}{rcl} L & = & 0.2989R + 0.5870G + 0.1140B \\ RG & = & R - G \\ BY & = & B - \frac{R+G}{2} \end{array} \tag{2.1}$$

2.2.3 Filtrage et "traitement rétinien"

Dans le modèle de la rétine que nous proposons, le filtrage qui est appliqué à l'information de luminance et de chrominance est différent. En effet, il a été montré que les fonctions de sensibilité au contraste en fonction des fréquences spatiales diffèrent pour l'information de luminance et l'information de chrominance : opposition Rouge-Vert et Bleu-Jaune (Fig. 2.4). Ainsi pour les deux images de chrominance



(a) Chrominance (b) Luminance

FIG. 2.4 – Fonction normalisée de sensibilité au contraste : (a) pour les voies chromatiques [Le Callet, 2001] et (b) pour la luminance [Mannos and Sakrison, 1974].

le "traitement rétinien" consiste simplement en un filtrage passe-bas dont la fonction de transfert reproduit les courbes de sensibilité au contraste à la figure 2.4a[1]. La fréquence de coupure de ce filtre pour l'image Rouge - Vert est un peu plus grande que celle utilisée pour filtrer l'image Bleu - Jaune (respectivement 5.5 et 4.1 cycles par degré).

Pour l'image de luminance, le "traitement rétinien" modélisé est plus complexe. La courbe de sensibilité au contraste (Fig. 2.4b) illustre le comportement fréquentiel global de la luminance. Dans notre modèle, ce comportement sera obtenu sur les sorties par modélisation successive des différentes étapes de cellules de la rétine. Il va concerner la modélisation des deux sorties principales : la sortie correspondant aux cellules parvocellulaires et la sortie correspondant aux cellules magnocellulaires. La plus grande majorité de la modélisation concerne la sortie parvocellulaire. La sortie magnocellulaire correspond à un simple filtrage passe-bas de l'image d'entrée (proche de la modélisation des cellules horizontales). La figure 2.5 schématise les différentes étapes du "traitement rétinien" réalisé sur l'image de luminance pour former la sortie parvocellulaire de la rétine.

Filtrage passe-bas spatialement variant

Comme nous l'avons vu, la densité des cônes sur la rétine n'est pas uniforme mais varie en fonction de l'excentricité (Fig. 2.3). Ainsi, au centre de la rétine (fovéa) le nombre de cônes est maximum et ce nombre diminue fortement en périphérie [Osterberg, 1935]. Un phénomène similaire est observé avec la densité des cellules ganglionnaires alors que le rapport entre les cônes et les cellules ganglionnaires augmente avec l'excentricité [Goodchild et al., 1996]. En conséquence, la zone de l'image correspondant au point de vue (point de fixation ou l'endroit où se porte l'œil) est traitée en haute résolution et, plus on s'éloigne du point de vue, plus la résolution

[1]C'est une approximation car dans cette figure, les courbes de sensibilité au contraste ont été contruites pour les voies chromatiques qui sont codées de manière différente de notre codage, mais restent toujours en opposition de couleurs Rouge-Vert et Bleu-Jaune.

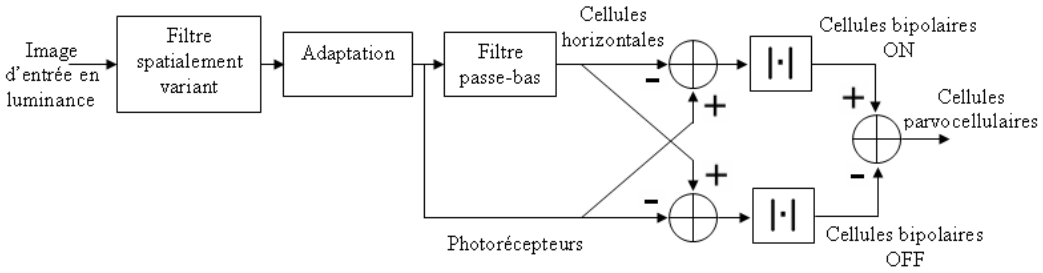


FIG. 2.5 – Modèle de traitement rétinien pour la luminance avec la sortie parvocellulaire.

est faible. Nous utilisons un filtre passe-bas spatialement variant afin de modéliser ces observations.

Au niveau des photorécepteurs, ce filtrage passe-bas spatialement variant correspond à un filtrage de type passe-bas dont la fréquence de coupure diminue au fur et à mesure que l'excentricité par rapport au point de vue augmente. Ce filtrage peut être implémenté de deux manières. Pour la première, l'image subit une décomposition pyramidale passe-bas à partir de laquelle la sortie du filtrage passe-bas spatialement variant s'obtient en interpolant les différents plans de la pyramide [Geisler and Perry, 1998; Perry and Geisler, 2002]. Pour la deuxième méthode, le filtrage passe-bas spatialement variant s'effectue par un filtrage récursif dont les paramètres sont variables avec l'excentricité. C'est cette implémentation que nous avons choisie. Une explication détaillée sera donnée au chapitre 5. Le traitement à ce niveau-là peut être résumé par l'équation :

$$L_{SV} = F_{SV}\{L\} \tag{2.2}$$

où L est la luminance de l'image en entrée, L_{SV} la luminance de l'image après le filtrage passe-bas spatialement variant et F_{SV} le filtre passe-bas spatialement variant.

Adaptation à la luminance

Les photorécepteurs de la rétine s'adaptent à une très grande dynamique de luminance. Cela nous permet de distinguer des objets dans des zones sombres en renforçant les faibles luminances sans saturer les zones de forte luminance. Ce phénomène, représenté par l'étape "Adaptation" du modèle de traitement rétinien à la figure 2.5, se modélise usuellement par une équation de type Naka-Rushton [Naka and Rushton, 1966] :

$$C = (L_{max} + L_0)\,\frac{L_{SV}}{L_{SV} + L_0} \tag{2.3}$$

avec L_{max} la luminance maximale et L_0 la luminance moyenne locale[2]. C représente

[2]Dans la simulation, nous avons choisi $L_{max} = 255$ et $L_0 = 0.1{+}410\frac{G}{G+105}$ avec $G = F_{PB}\{L_{SV}\}$. G est la sortie d'un filtre passe-bas (F_{PB}) de L_{SV} ; les constantes 0.1, 410 et 105 sont choisies empiriquement sur des images naturelles [Chauvin, 2003]. La valeur de 0.1 dans l'expression de L_0 a pour but d'éviter un dénominateur nul dans l'expression de C par exemple pour une zone noire (i.e. $L_{SV} = 0$).

la luminance en sortie des photorécepteurs.

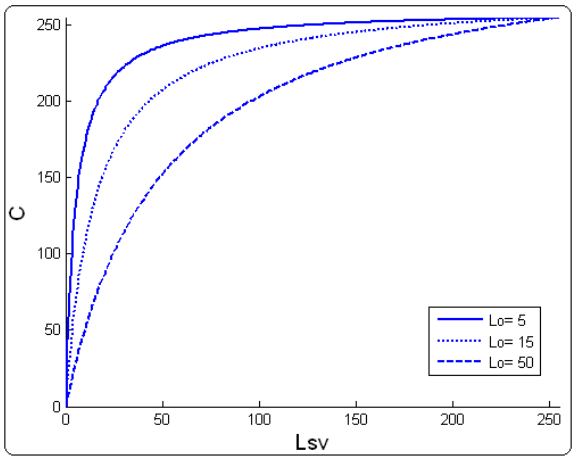


FIG. 2.6 – Fonction d'adaptation des photorécepteurs décrite à l'équation 2.3 pour différentes valeurs de la luminance moyenne locale L_0. La luminance en sortie des photorécepteurs C varie en fonction de la luminance L_{SV} après le filtrage passe-bas spatialement variant. Les dynamiques de C et L_{SV} correspondent à des dynamiques de luminance d'image (entre 0 et 255).

L'évolution de la sortie des photorécepteurs C est illustrée à la figure 2.6 pour différentes valeurs de L_0. L'adaptation à la luminance dépend de la moyenne locale L_0 ; plus L_0 est faible, plus la pente de courbe est importante. Cela permet d'éclaircir les zones sombres. Au contraire, pour une forte valeur de L_0, la dynamique de sortie des zones claires ou surexposées sera réduite. Nous remarquons également que la courbe de l'équation 2.3 a une forme approchée logarithmique par l'effet de saturation, mais restant bornée. Cette fonction d'adaptation à la luminance est souvent appelée "fonction de compression". Un exemple de l'adaptation des photorécepteurs est présenté à la figure 2.8.b. Elle se voit plus particulièrement au niveau du chapeau de Léna.

Réhaussement de contraste[3]

Les cellules horizontales (H) lissent l'information transmise par les photorécepteurs, elles sont ainsi modélisées par un filtrage passe-bas de la sortie C des photorécepteurs.

Les photorécepteurs et les cellules horizontales fournissent les signaux d'entrée des cellules bipolaires. Les photorécepteurs représentent l'excitation au centre du champ récepteur d'une cellule bipolaire et les cellules horizontales, l'inhibition à la périphérie. Le fonctionnement des cellules bipolaires est donc modélisé par la

[3]Nous considérons ici le contraste comme simplement la valeur de différence entre des niveaux spatialement voisins.

différence entre la sortie des photorécepteurs et la sortie des cellules horizontales ; ce qui correspond à un filtrage passe-bande (Fig. 2.7). Suivant ce mécanisme, les cellules bipolaires sont responsables du réhaussement de contraste des stimuli.

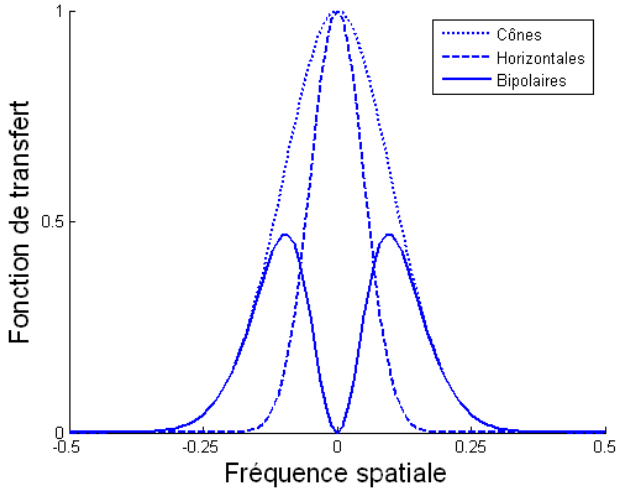


FIG. 2.7 – La fonction de transfert 1D des cellules bipolaires correspond à un filtre passe-bande. Il correspond à la différence de deux filtres passe-bas : celui des photorécepteurs et celui des cellules horizontales. L'axe des abscisses correspond aux fréquences spatiales réduites.

Les deux types de cellules bipolaires sont modélisés (Fig. 2.5). Les cellules bipolaires "centre-ON" (B_{ON}) donne une réponse positive si le signal au centre est supérieur à celui en périphérie. En revanche, les cellules bipolaires "centre-OFF" (B_{OFF}) répondent quand le signal à la périphérie est supérieur au signal au centre (Eq. 2.4). D'ailleurs, les cellules bipolaires ne répondent pas si le stimulus est constant dans tout le champ récepteur (i.e. région uniforme : fréquence spatiale nulle). Le modèle se décrit alors par les équations suivantes :

$$\begin{cases} B_{ON} &= |C - H|_+ \\ B_{OFF} &= |H - C|_+ \end{cases} \tag{2.4}$$

avec

$$|x|_+ = \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} .$$

Les cellules rétiniennes suivantes, les cellules amacrines, jouent un rôle essentiel pour le traitement du mouvement. Pour notre part, en considérant dans ce chapitre uniquement les scènes statiques, la modélisation de ces cellules ne sera pas intégrée. Mais, nous reviendrons sur cette modélisation au chapitre 6 dans l'implémentation d'un modèle spatio-temporel de la rétine.

L'étage de sortie de la rétine est formé par les cellules ganglionnaires : parvocellulaires et magnocellulaires. Alors que les cellules parvocellulaires sont sensibles aux basses fréquences temporelles et aux hautes fréquences spatiales, elles

sont modélisées par la différence entre les cellules bipolaires ON et OFF :

$$P = B_{ON} - B_{OFF}. \tag{2.5}$$

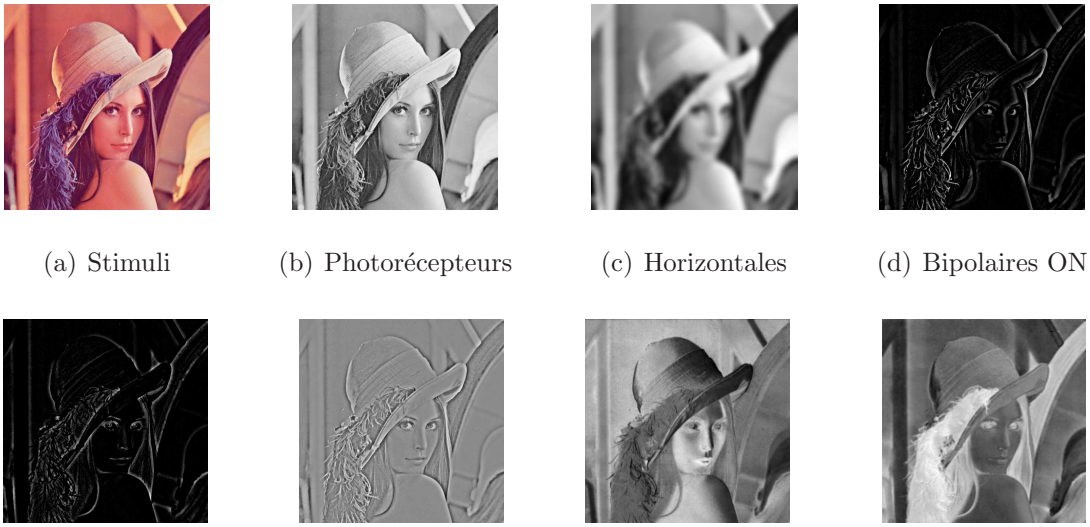
Dans le modèle complet (Fig.2.5), il y a également une étape d'adaptation non linéaire au niveau des cellules ganglionnaires. Elle est identique à celle implémentée au niveau des photorécepteurs (Eq. 2.3). Dans le cas de notre étude, cette deuxième non-linéarité n'a pas été implémentée car les résultats sur les simulations numériques ne montraient pas d'effets significatifs et permettaient de simplifier la modélisation de la voie parvocellulaire :

$$P = C - H \tag{2.6}$$

Ainsi, sans non-linéarité, cette équation résulte de l'équation 2.5 et de la formation de la voie parvocellulaire comme $P = B_{ON} - B_{OFF}$ (Fig.2.5).

Quant aux cellules magnocellulaires, elles répondent aux hautes fréquences temporelles et aux basses fréquences spatiales. Ainsi, pour un stimulus statique, la voie magnocellulaire représente uniquement ses basses fréquences spatiales. Ces cellules sont ici modélisées comme les cellules horizontales par un filtrage passe-bas.

La figure 2.8 illustre les sorties des différentes cellules rétiniennes pour la luminance et les sorties des voies chromatiques.



(a) Stimuli (b) Photorécepteurs (c) Horizontales (d) Bipolaires ON

(e) Bipolaires OFF (f) Parvocellulaires (g) RG (h) BY

FIG. 2.8 – Les sorties des différentes cellules rétiniennes pour la voie de luminance et les sorties des voies de chrominance. (a) Stimuli ; (b) Photorécepteurs ; (c) Cellules horizontales ; (d) Cellules bipolaires ON ; (e) Cellules bipolaires OFF ; (f) Cellules parvocellulaires ; (g) Opposition de couleurs Rouge-Vert ; (h) Opposition de couleurs Bleu-Jaune.

2.3 Modèle de la rétine : résumé

Nous avons modélisé les propriétés essentielles de la rétine au niveau spatial pour une image d'entrée en couleur. L'image est tout d'abord décomposée en 3 images : une de luminance et deux de chrominance avec les oppositions de couleurs, Rouge-Vert et Bleu-Jaune. Puis des traitements sont réalisés sur ces 3 images. Pour la luminance, le traitement rétinien vise à renforcer les contrastes (hautes fréquences spatiales) par une chaîne de traitements réalisés par les photorécepteurs, les cellules horizontales, les cellules bipolaires, les cellules parvocellulaires et les cellules magnocellulaires. Quant aux images de chrominance, le modèle de rétine consiste en un filtrage passe-bas spatial dont la fréquence de coupure pour la voie Rouge-Vert est un peu plus haute que celle pour la voie Bleu-Jaune. Enfin, les sorties de la rétine, comportant à la fois l'information de luminance et de chrominance, sont envoyées jusqu'au cortex visuel dans lequel plusieurs traitements seront appliqués pour construire la carte de saillance.

3 Cortex visuel primaire

3.1 Physiologie

Le cortex visuel primaire est une aire corticale située en arrière de la boîte crânienne. Il est également appelé V1 ou cortex strié. L'aire V1 est sans doute la partie la plus connue du cortex visuel [Hubel et al., 1977; Hubel, 1981; Tootell et al., 1981; Bullier, 2002; Delorme and Flückiger, 2003]. Parmi ces études figurent celles de Hubel et Wiesel récompensées par le prix Nobel en 1981. La structure de l'aire V1 est caractérisée par une organisation en couches horizontales et en colonnes verticales par rapport à la surface du cortex (Fig. 2.9). Le cortex visuel primaire est composé de 6 couches horizontales. Les sorties rétiniennes, relayées par les corps géniculés latéraux (CGL), se terminent principalement dans la couche 4. Les neurones dans une même colonne verticale sont sensibles à une même orientation et à différentes fréquences spatiales du stimulus visuel. Ainsi, lorsque l'on se déplace dans une colonne verticale, la fréquence spatiale, à laquelle les neurones sont sensibles, varie. Deux colonnes adjacentes présentent une variation de sensibilité à l'orientation d'environ 10°. Les champs visuels de l'œil gauche et droit sont alternativement traités par des colonnes entrelacées, cela s'appelle la "dominance oculaire". Les colonnes regroupant toutes les orientations dans 360° et les deux champs visuels constituent une hypercolonne. Ainsi, une hypercolonne est capable d'analyser complètement une portion du champ visuel en la décomposant en différentes fréquences spatiales et différentes orientations.

L'aire V1 comporte 3 types de cellules : les cellules simples, les cellules complexes et les cellules hypercomplexes. Les cellules simples et complexes répondent aux stimuli de type "barres de lumière orientées" (*"oriented gratings"*). Alors que les cellules simples sont sensibles aux barres à une certaine position spatiale dans le champ visuel, les cellules complexes répondent aux barres de lumière quelque soit leur position dans le champ visuel. D'après Hubel et Wiesel, les réponses des cellules complexes sont construites à partir d'une combinaison non linéaire des réponses des

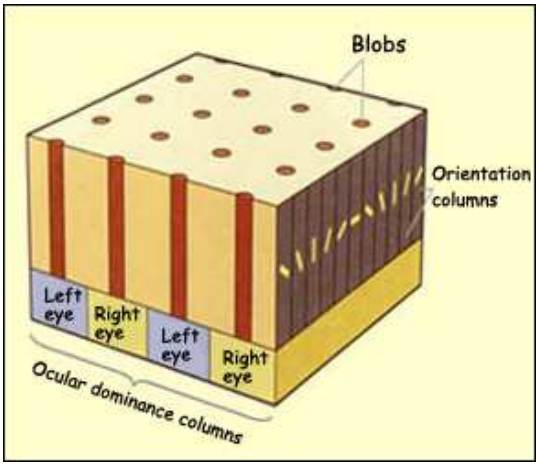


FIG. 2.9 – Organisation d'une hypercolonne de l'aire V1. (depuis http ://the-brain.mcgill.ca/flash/a/a_02/a_02_cl/a_02_cl_vis/a_02_cl_vis.html).

cellules simples. Quant au troisième type de cellules, les cellules hypercomplexes, elles sont sensibles à la longueur des stimuli [Hubel et al., 1977; Movshon, 1978]. Dans notre modèle nous ne modélisons que les cellules complexes, celles-ci étant majoritaires dans le cortex visuel. De plus, elles portent une information semblable à celle des cellules simples et cette information est suffisante pour notre modèle de saillance.

Les cellules corticales interagissent entre elles renforçant ou diminuant leurs réponses à un stimulus. Ici, sans ambition de décrire en détails toutes les interactions entre les cellules corticales, nous modélisons deux types d'interactions : les interactions courtes, *"short-range interactions"*, et les interactions longues, *"long-range interactions"*. Ces interactions se distinguent par leur portée de connexion.

Les interactions courtes ont lieu entre les cellules dont les champs récepteurs se recouvrent. Ce type d'interactions est lié aux connexions latérales locales dans le cortex. Ainsi, la réponse d'une cellule sensible à une orientation peut être supprimée par les réponses des cellules voisines répondant à d'autres orientations (*"cross-orientation suppression"*) [DeAngelis et al., 1992; Das and Gilbert, 1999].

Les interactions longues agissent entre les cellules de différentes colonnes reliées par de longues connexions horizontales ; ces cellules sont sensibles à une même orientation [Ts'o et al., 1986]. Ainsi, la réponse d'une cellule peut être modulée par une stimulation extérieure à son champ récepteur. Les expériences psychophysiques par exemple ont montré que le seuil de détection au contraste d'un stimulus dépend de la présence et de l'orientation de stimuli présentés autour [Zenger et al., 2000]. D'après Braun [Braun, 1999], une modulation excitatrice a lieu entre les segments colinéaires et une modulation inhibitrice entre les segments non-colinéaires et orthogonaux. Par conséquent, les interactions longues ont souvent été exploitées pour renforcer les contours des objets dans les scènes [Hansen et al., 2001; Stettler et al., 2002].

3.2 Modélisation du traitement cortical

L'extraction de caractéristiques visuelles élémentaires est effectuée au niveau de l'aire V1. Dans cette partie, nous nous concentrerons sur la modélisation du traitement en orientations et en fréquences spatiales réalisé par les cellules complexes. Les interactions courtes et longues seront modélisées pour renforcer ou inhiber les réponses de ces cellules. Nous présentons également les étapes de normalisation et de fusion des réponses des différentes cellules pour créer une carte de saillance.

3.2.1 Filtres corticaux

Comme nous l'avons présenté ci-dessus, les cellules corticales sont sensibles aux orientations et aux fréquences spatiales du stimulus visuel. Parmi les cellules corticales, les cellules complexes sont souvent choisies pour représenter le traitement en orientations et en fréquences réalisé au niveau du cortex visuel. La réponse fréquentielle de ces cellules peut être modélisée par des filtres de type "passe-bande orienté".

Dans la littérature, les filtres de Gabor sont couramment utilisés [Daugman, 1980]. Les filtres Gabor sont présents dans plusieurs modèles d'attention visuelle [Itti et al., 1998; Itti and Koch, 2001; Torralba et al., 2006]. Plus récemment certains auteurs ont proposé un autre filtre pour modéliser les cellules complexes : les filtres LogNormaux [Field, 1987; Knutsson et al., 1994; Guyader, 2004; Massot and Hérault, 2008]. Les fonctions de transfert de ces filtres s'apparentent à une distribution normale de la variable fréquentielle en échelle logarithmique. Ce qui explique le nom de ce filtre[4]. Nous avons choisi de modéliser le fonctionnement des cellules corticales en utilisant les filtres LogNormaux.

Filtre LogNormal

Le filtre LogNormal $F_{\theta_0,f_0}(\theta, f)$ comme illustré à la figure 2.10c est à variables séparables en fréquence f et en orientation θ dans le plan fréquentiel :

$$F_{\theta_0,f_0}(\theta, f) = \Phi_{\theta_0}(\theta) \cdot R_{f_0}(f) \tag{2.7}$$

avec

$$R_{f_0}(f) = \exp\left\{ -0.5 \frac{\log\left(\frac{f}{f_0}\right)^2}{\sigma_f^2} \right\}$$

$$\Phi_{\theta_0}(\theta) = \exp\left\{ -0.5 \frac{(\theta - \theta_0)^2}{\sigma_\theta^2} \right\}$$

où θ_0 est l'orientation du filtre, f_0 la fréquence centrale du filtre (voir Fig. 2.10b,c). σ_f et σ_θ déterminent respectivement sa bande passante radiale BW_f et sa bande passante transversale BW_θ. La séparation des variables facilite le calcul des bandes

[4]Le filtre LogNormal est aussi appelé Log-Gabor.

passantes. Ainsi, la bande passante radiale est définie à partir de $R_{f_0}(f)$:

$$\begin{aligned} BW_f &= \log_2\left(\frac{f_{c2}}{f_{c1}}\right) \\ &= 2\sigma_f\sqrt{2\log_2 e} \text{ (octave)} \end{aligned} \tag{2.8}$$

avec f_{c1}, f_{c2}[5] les fréquences de coupure inférieure et supérieure du filtre R_{f_0}.

La bande passante transversale à partir de $\Phi_{\theta_0}(\theta)$ est calculée :

$$\begin{aligned} BW_\theta &= \theta_{c2} - \theta_{c1} \\ &= 2\sigma_\theta\sqrt{2\ln 2} \end{aligned} \tag{2.9}$$

avec θ_{c1}, θ_{c2} les fréquences de coupure inférieure et supérieure du filtre Φ_{θ_0}.

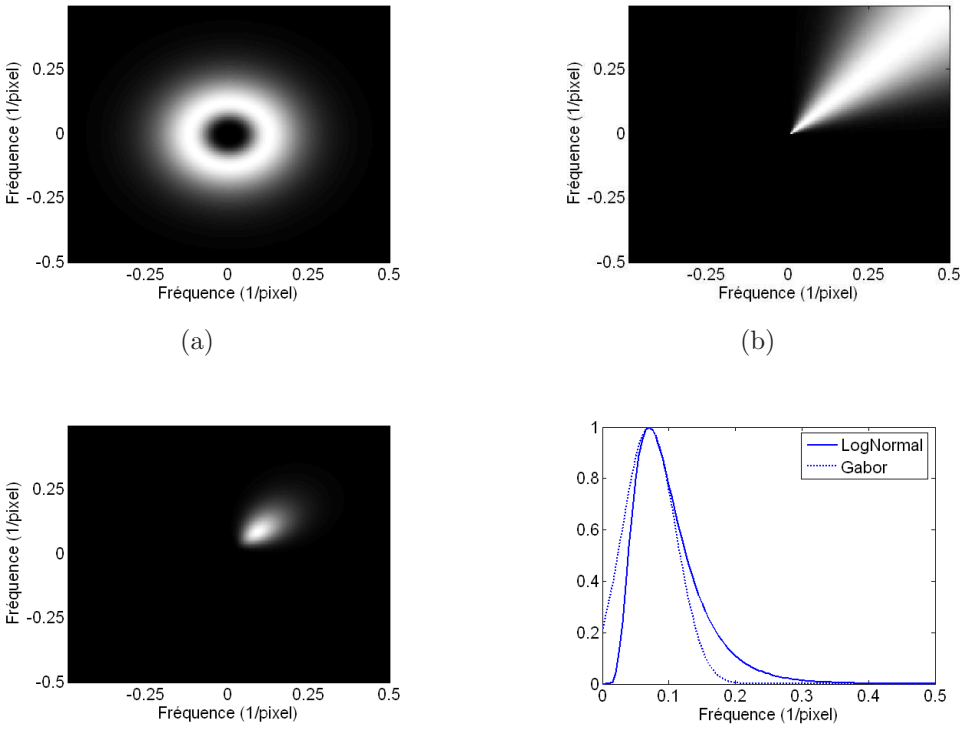


FIG. 2.10 – Illustration d'un filtre LogNormal : (a) fonction radiale $R_{0.12}(f)$ avec $\sigma_f = 0.44$; (b) fonction orientée $\Phi_{45°}(\theta)$ avec $\sigma_\theta = 14$; (c) un filtre LogNormal 2D $F_{45°,0.12}(\theta, f)$ construit à l'aide du produit de la fonction radiale avec la fonction orientée ; (d) Exemple d'un filtre LogNormal 1D. Le filtre LogNormal s'annule à la fréquence nulle ; contrairement au filtre Gabor.

Nous remarquons à la figure 2.10d que lorsque la fréquence centrale d'un filtre de Gabor est proche de zéro, ce dernier couvre toujours la partie à fréquence nulle

[5]Les fréquences de coupure inférieure et supérieure sont déterminées à la mi-hauteur de la fonction de transfert en amplitude du filtre.

quelque soit son écart-type, ce n'est pas le cas pour le filtre LogNormal. Cet avantage du filtre LogNormal est utile pour la construction d'un banc de filtres où l'on veut que les filtres à différentes orientations ne se recouvrent pas à la fréquence nulle. De plus, pour une image naturelle, l'énergie à la fréquence nulle est souvent très importante et n'est pas spécifique d'une orientation particulière. Cette énergie à la fréquence nulle pourrait donc biaiser les sorties des filtres basses fréquences.

Banc de filtres LogNormaux

Une hypercolonne de cellules corticales est modélisée par un banc de filtres LogNormaux dans N orientations et M fréquences spatiales. Le banc de filtres comporte alors $N \times M$ filtres $F_{\theta_i,f_j}(\theta, f)$, $i = 1..N$, $j = 1..M$.

Les orientations choisies se répartissent régulièrement entre 0 et $\frac{N-1}{N}180°$ ($\theta_i = \frac{i-1}{N}180$). Quant aux fréquences, elles suivent une décomposition dyadique : les fréquences centrales diminuent d'un facteur 2 à partir d'une fréquence maximale[6]. Ainsi, $f_{j-1} = \frac{f_j}{2}$ et $f_M = f_{max}$. Le banc de filtres est choisi de manière à couvrir au mieux l'ensemble des orientations et des fréquences spatiales présentes dans les scènes naturelles. Nous choisissons ici une fréquence maximale $f_{max} = 0.25\ pixel^{-1}$.

Nous allons maintenant déterminer l'écart-type $\sigma_{f,j}$ pour les filtres LogNormaux à la fréquence f_j. A partir de l'équation 2.8, l'écart-type $\sigma_{f,j}$ peut être calculé en fonction de la bande passante radiale $BW_{f,j}$ selon l'équation suivante :

$$\sigma_{f,j} = \frac{BW_{f,j}}{2\sqrt{2\log_2(e)}} \tag{2.10}$$

Dans notre modèle, la bande passante radiale est déterminée à partir des données biologiques [DeValois et al., 1982]. La figure 2.11a présente la bande passante en octave d'un filtre cortical en fonction de la fréquence spatiale (en échelle logarithmique). Cette courbe peut être modélisée selon une loi linéaire en $\log(f)$ à la figure 2.11b. Ainsi,

$$BW_f = a\log_2(f_0) + b \tag{2.11}$$

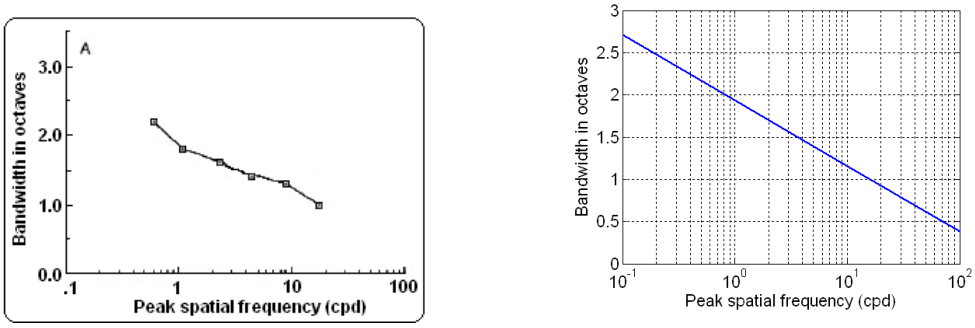
avec $a = -0.2333$ et $b = 1.9334$.

D'où, nous pouvons calculer les écart-types $\sigma_{f,j}$ des filtres LogNormaux en fonction des fréquences centrales f_j :

$$\sigma_{f,j} = \frac{a\log_2(f_j) + b}{2\sqrt{2\log_2(e)}} = a_1\log_2(f_j) + b_1 \tag{2.12}$$

avec $a_1 = \frac{a}{2\sqrt{2\log_2(e)}}$ et $b_1 = \frac{b}{2\sqrt{2\log_2(e)}}$.

[6]Le facteur 2 est également utilisé pour la décomposition dyadique en ondelettes [Daubechies, 1992; Strang and Nguyen, 1996]



(a) (b)

FIG. 2.11 – Evolution de la bande passante des cellules corticales en fonction de leur sensibilité en fréquences spatiales : (a) Données psychophysiques [DeValois et al., 1982] ; (b) Interpolation par une loi linéaire en $\log(f)$.

Ainsi,

$$\begin{aligned}\frac{\sigma_{f,j}}{\sigma_{f,j-1}} &= \frac{a_1 \log_2(f_j) + b_1}{a_1 \log_2(f_{j-1}) + b_1} \\ &= \frac{a_1 \log_2(f_{max} 2^{j-M}) + b_1}{a_1 \log_2(f_{max} 2^{j-1-M}) + b_1} \\ &= \frac{a_1 \cdot j + b_2}{a_1 \cdot (j-1) + b_2}\end{aligned} \tag{2.13}$$

avec $b_2 = a_1(\log_2 f_{max} - M) + b_1$.

La figure 2.12 illustre le recouvrement des filtres LogNormaux dans la direction radiale.

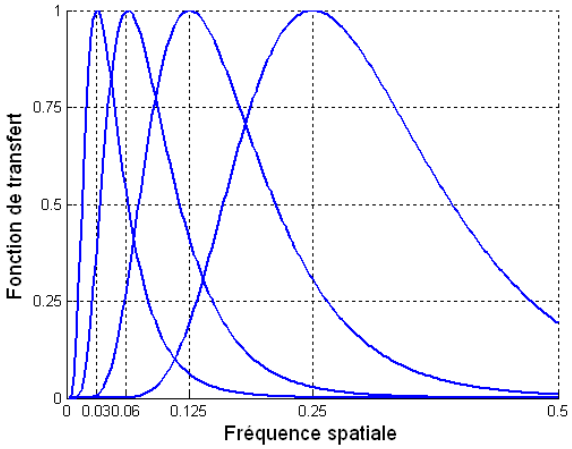


FIG. 2.12 – Le recouvrement des filtres LogNormaux en une dimension.

L'écart-type σ_θ qui intervient dans le calcul de la bande passante transversale BW_θ a été choisi pour que le banc de filtres couvre toutes les orientations du spectre.

A partir de l'équation 2.9, on obtient :

$$\sigma_\theta = \frac{BW_\theta}{2\sqrt{2\ln(2)}} \tag{2.14}$$

La bande passante transversale BW_θ est choisie constante dans toutes les orientations et $BW_\theta = k \cdot \frac{180}{N}$ où k détermine le niveau de recouvrement entre les deux filtres de deux orientations voisines. Avec $k = 1$, les contours à mi-hauteur des fonctions de transfert de deux filtres LogNormaux de deux orientations voisines sont tangents. Plus k est grand, plus le recouvrement est important. Comme BW_θ est constant, σ_θ est aussi constant pour tous les filtres[7]. La figure 2.13 présente le banc de filtres LogNormaux dans le plan fréquentiel qui sera utilisé pour la décomposition de la voie de luminance.

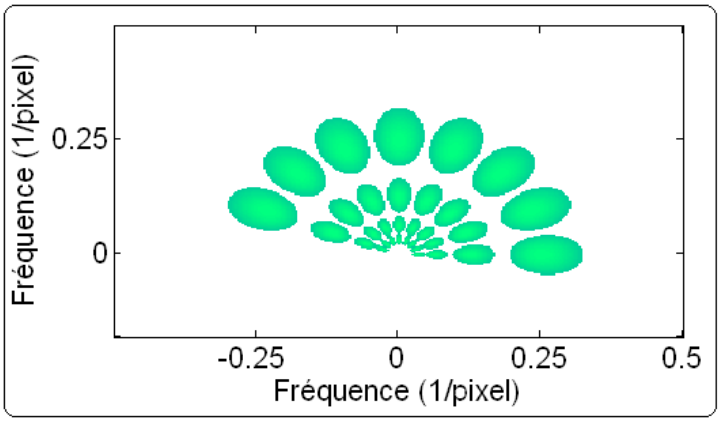


FIG. 2.13 – Un banc de filtres LogNormaux dans 8 orientations et 4 fréquences spatiales dans le plan fréquentiel. La figure représente la partie des fonctions de transfert des filtres LogNormaux qui est au dessus de 80% de leur hauteur maximale.

Au niveau de l'implémentation, pour des raisons de simplicité algorithmique, le filtrage des images par le banc de filtres est réalisé dans le domaine fréquentiel (Eq. 2.15).

$$\begin{cases} E_{ij}(\theta, f) &= F_{ij}(\theta, f) \cdot \mathrm{FFT}\{I_r\} \\ e_{ij} &= \left(\mathrm{FFT}^{-1}\{E_{ij}\}\right)^2 \end{cases} \tag{2.15}$$

avec I_r une sortie de la rétine qui peut être la voie parvocellulaire pour la luminance ou les voies chromatiques.

Ainsi, après le banc de filtres LogNormaux, une image est décomposée en $N \times M$ *cartes d'énergie* e_{ij} dont chacune représente l'énergie de l'image dans une orientation et une fréquence spécifique.

La décomposition par le banc de filtres est appliquée à la voie de luminance et aux voies chromatiques. Cependant, le nombre de filtres utilisés est différent pour

[7]Dans l'implémentation, nous avons choisi $k = 1.5$ (et donc $\sigma_\theta = 14$).

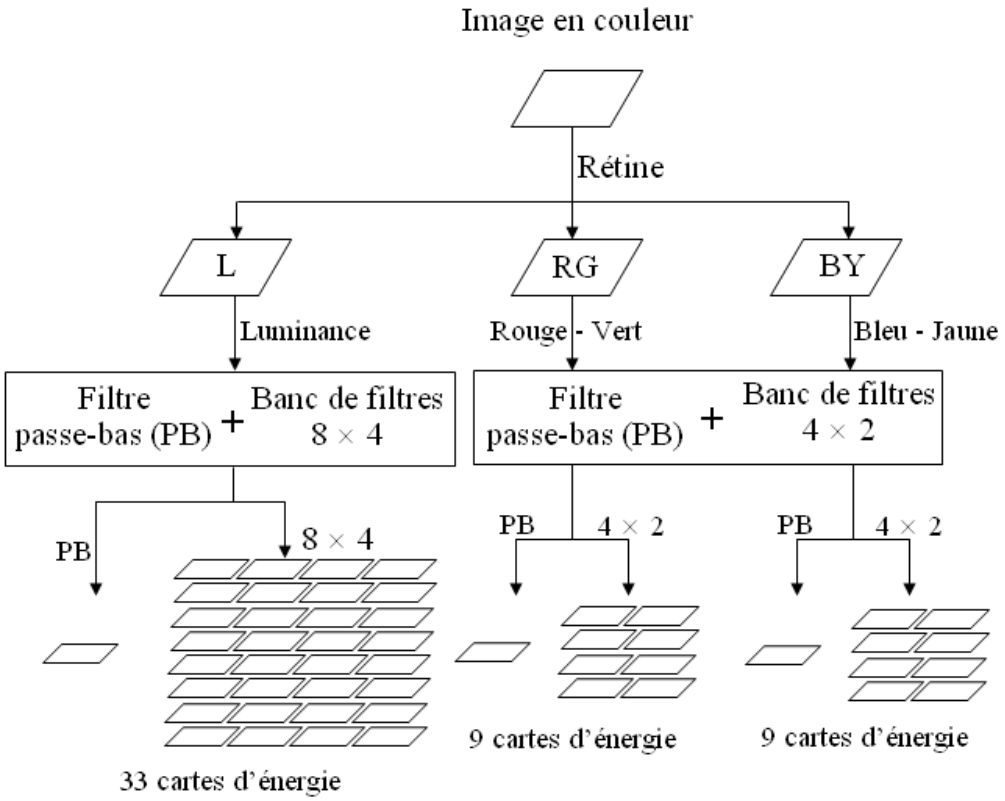


FIG. 2.14 – Schéma représentant la décomposition des voies de luminance et de chrominance en différentes orientations et différentes fréquences.

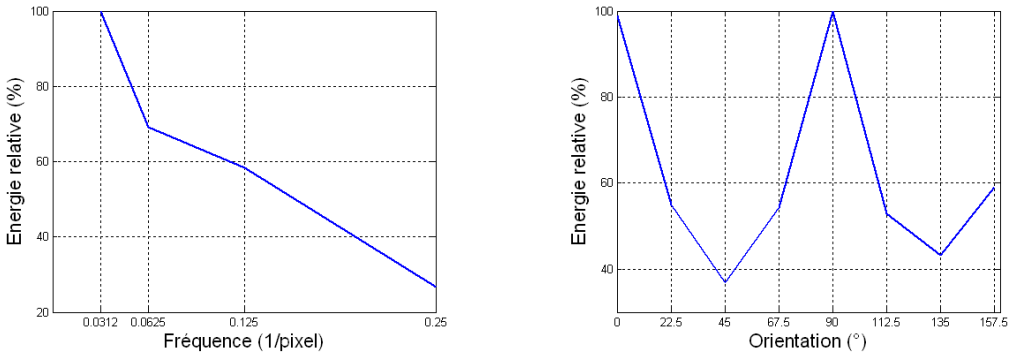
ces dernières. En effet, les voies chromatiques sont moins sensibles à l'orientation que la voie de luminance [McIlhagga and Mullen, 1996; Beaudot and Mullen, 2005]. De plus, les voies chromatiques contiennent plutôt des basses fréquences ; tandis que la voie de luminance contient des hautes fréquences. Ainsi, dans l'implémentation, nous avons choisi 8 orientations et 4 bandes de fréquence pour la luminance (Fig. 2.13). Ces paramètres réalisent une optimisation "heuristique" entre le nombre de filtres et le temps de calcul. Quant aux voies chromatiques, elles sont analysées par des filtres aux 4 orientations (0°, 45°, 90° et 135°) et 2 bandes de fréquences correspondant aux 2 bandes les plus basses fréquences du banc de filtres à la figure 2.13.

De plus, un filtre passe-bas centré sur la fréquence nulle est ajouté pour chaque voie chromatique et la voie magnocellulaire de la luminance afin de capter les informations autour de la fréquence nulle.

La figure 2.14 résume les cartes d'énergie sortant du banc de filtres pour la voie de luminance et pour les voies de chrominance. Ainsi, il y a 33 cartes d'énergie pour la voie de luminance et 9 cartes d'énergie pour chacune des voies chromatiques.

3.2.2 Normalisation

L'amplitude du spectre des scènes naturelles se caractérise par une décroissance en "$1/f$" [Field, 1987, 1994; Tolhurst et al., 1992]. Ainsi, l'énergie des scènes naturelles se situe majoritairement dans les basses fréquences. De plus, les orientations horizontale et verticale comportent plus d'énergie que les orientations obliques [Baddeley, 1997; Oliva and Torralba, 2001; Switkes et al., 1978]. La figure 2.15 représente la distribution d'énergie selon différentes orientations et différentes fréquences pour la voie de luminance des 17 images Kodak (`http://www.cipr.rpi.edu/resource/stills/kodak.html`). Cette distribution est calculée de la manière suivante : La luminance de chaque image est décomposée par le banc de filtres LogNormaux décrit ci-dessus selon 8 orientations et 4 fréquences. L'énergie d'une orientation est la somme de tous les pixels des 4 cartes d'énergie dans cette orientation. De même, l'énergie d'une fréquence est la somme de tous les pixels des 8 cartes d'énergie à cette fréquence. Les résultats sont ensuite moyennés sur les 17 images de la base. Ces résultats confirment la distribution non-uniforme de l'énergie des scènes naturelles : elle se concentre aux basses fréquences spatiales et aux orientations horizontale et verticale.



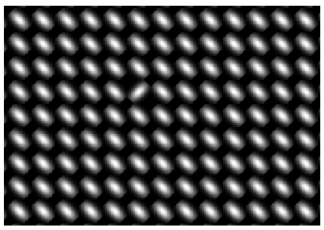
(a) (b)

FIG. 2.15 – La distribution des énergies en sortie d'un banc de filtres LogNormaux pour 17 scènes naturelles extraites du site *http ://www.cipr.rpi.edu/resource/stills/kodak.html*. (a) Selon 4 bandes de fréquence ; (b) Selon 8 orientations.

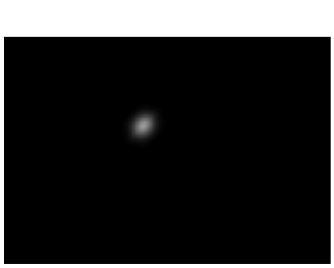
Ici, nous proposons de normaliser les cartes d'énergie pour que l'énergie des différentes orientations (après la normalisation) ait le même ordre de grandeur. Nous appelons l'*énergie totale* d'une carte d'énergie la somme de ses pixels et l'*énergie totale maximale* d'une orientation la valeur maximale de l'*énergie totale* de toutes les cartes dans cette orientation. L'*énergie totale maximale* $e_{i,max}$ de l'orientation i se calcule selon l'équation suivante :

$$e_{i,max} = \max_{j=1..M} \left\{ e^*_{ij} \right\} \tag{2.16}$$

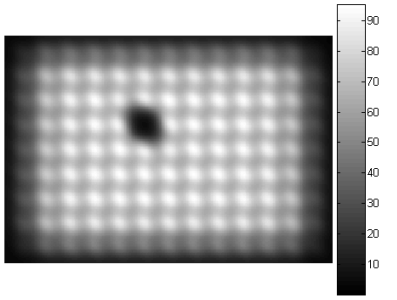
avec e^*_{ij} l'*énergie totale* de la carte e_{ij} et $e^*_{ij} = \sum_{x,y} e_{ij}(x,y)$.



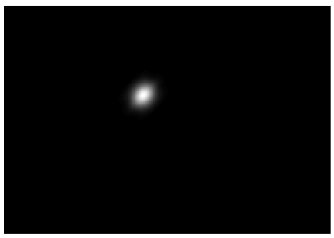
(a) Image initiale



(b) 45°



(c) 135°



(d) 45°

(e) 135°

FIG. 2.16 – Exemple de la normalisation par l'*énergie totale maximale* de chaque orientation. Première ligne : image initiale ; deuxième ligne : cartes d'énergie avant la normalisation ; troisième ligne : cartes d'énergie après la normalisation. Les cartes d'une même ligne sont affichées dans un même intervalle de luminosité (voir texte).

Puis, dans chaque orientation les cartes d'énergie sont normalisées par l'*énergie totale maximale* de cette orientation :

$$e_{ij}^{norm} = \frac{e_{ij}}{e_{i,max}} \tag{2.17}$$

où e_{ij}^{norm} est une carte normalisée.

Ainsi, après la normalisation, l'*énergie totale maximale* de chaque orientation est égale à 1. Par conséquent, elle peut renforcer les zones saillantes par rapport aux zones non-saillantes lorsque les zones saillantes et les zones non-saillantes appartiennent à deux orientations différentes. En effet, le nombre de pixels des zones saillantes est beaucoup moins que celui des zones non-saillantes (par exemple, le fond). Alors que l'*énergie totale maximale* est la même, les énergies des pixels des

zones saillantes devraient être plus fortes.

La figure 2.16 illustre l'effet de la normalisation sur une image artificielle. L'image est filtrée par le banc de filtres LogNormaux selon 8 orientations et 4 fréquences. Avant la normalisation, à la fréquence 0.06 $pixel^{-1}$, la carte d'énergie à l'orientation 135° (Fig. 2.16c) est très forte et elle risque de masquer le stimulus "pop-out" provenant d'une carte à 45° (Fig. 2.16b). Cependant, l'*énergie totale maximale* à l'orientation 135° est beaucoup plus importante que l'*énergie totale maximale* à l'orientation 45°. Par conséquent, après la normalisation par l'*énergie totale maximale*, l'énergie du stimulus "pop-out" devient beaucoup plus forte que celle du fond (Fig. 2.16d,e).

Dans notre modèle d'attention visuelle, les normalisations sont effectuées pour la voie de luminance et les deux voies chromatiques.

3.2.3 Interactions

La modélisation des interactions entre cellules corticales permet de moduler la réponse des filtres dans certaines orientations et est implémentée par les interactions courtes et longues. Dans la littérature, les interactions courtes ont été modélisées en prenant en compte les influences des cellules d'orientations voisines. Elles sont souvent de type "divisive" : les modulations des cellules voisines correspondent au dénominateur et la cellule intéressée au numérateur [Lee et al., 1999; Peters et al., 2005]. Quant aux interactions longues, elles s'effectuent souvent par la convolution avec un masque dit "papillon" (au vue de sa forme) dans une orientation particulière [Hansen et al., 2001; Peters et al., 2005].

Dans notre modèle d'attention visuelle, nous modélisons également les interactions courtes et longues. Alors que les interactions consistent à renforcer la saillance en orientation, elles sont réservées uniquement à la voie de luminance (sauf la partie centrée sur la fréquence nulle). Les voies chromatiques qui sont moins sensibles à l'orientation ne sont pas soumises à ces interactions.

Interactions courtes

Les interactions courtes sont réalisées d'une manière linéaire pour renforcer la saillance d'une orientation spécifique. Ces interactions interviennent à la même position dans différentes cartes. Selon l'équation 2.18, chaque carte d'énergie e_{ij}^{norm} est excitée par les cartes de même orientation (mais de fréquences voisines) et inhibée par les cartes aux orientations voisines (mais de même fréquence). La figure 2.17c illustre l'implémentation des interactions courtes dans notre modèle.

$$e_{ij}^{s} = e_{ij}^{norm} + 0.5e_{i,j-1}^{norm} + 0.5e_{i,j+1}^{norm} - 0.5e_{i-1,j}^{norm} - 0.5e_{i+1,j}^{norm} \tag{2.18}$$

où e_{ij}^{s} est une carte d'énergie après les interactions courtes.

Ainsi, l'énergie d'une orientation peut être inhibée si l'énergie des orientations voisines est plus forte. De plus, l'énergie d'une fréquence est renforcée par l'énergie

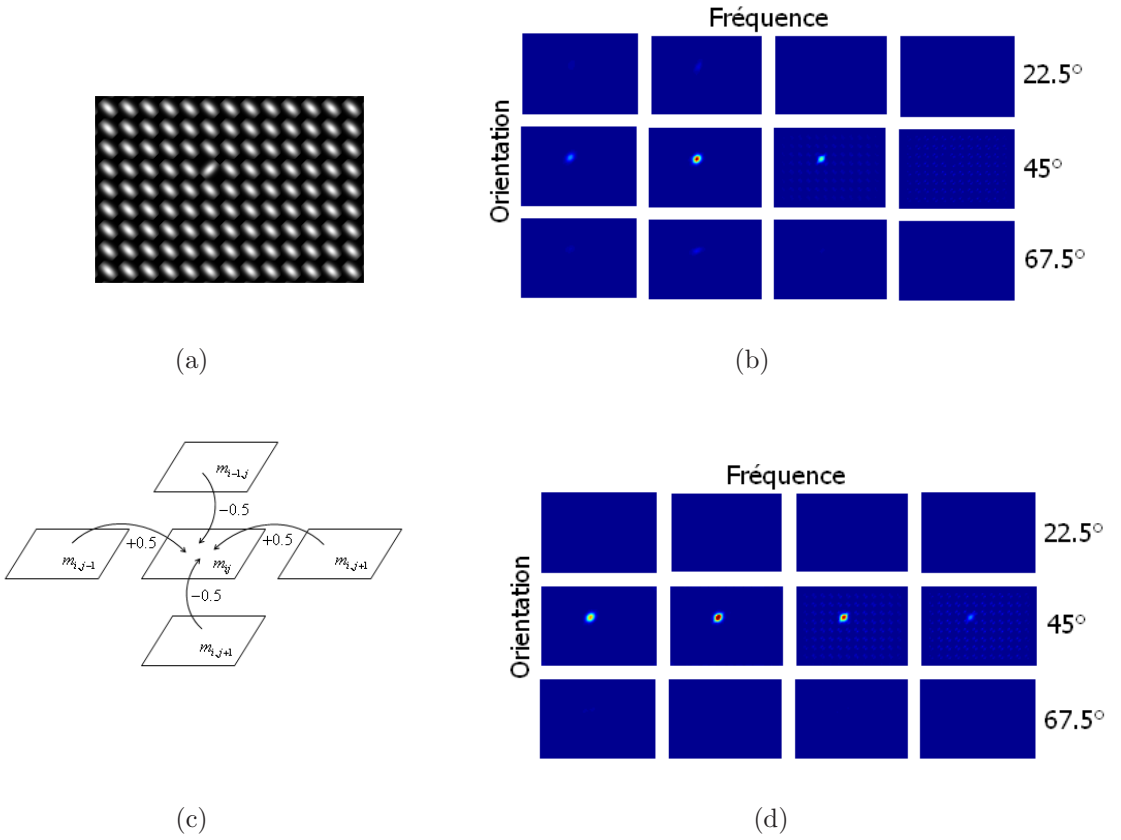


FIG. 2.17 – Exemple des interactions courtes : (a) Image initiale ; (b) Les cartes d'énergie avant les interactions courtes ; (c) Les interactions courtes entre les cartes voisines ; (d) Les cartes d'énergie après les interactions courtes.

des fréquences voisines et de même orientation. L'effet des interactions courtes est illustré à la figure 2.17. Comme ci-dessus, l'image initiale est décomposée par le banc de filtres LogNormaux en 32 cartes d'énergie selon 8 orientations et 4 fréquences. Avant les interactions courtes, le stimulus "pop-out" à 45° (Fig. 2.17.a) est présent dans les cartes d'énergie à 45° ainsi que dans celles aux orientations voisines (22.5° et 67.5°) (Fig. 2.17.b). Après les interactions courtes, les réponses aux orientations voisines sont supprimées tandis que celles à 45° sont renforcées, notamment pour la première, troisième et quatrième bandes de fréquence (Fig. 2.17.d). Ainsi, les interactions courtes permettent de faire ressortir la saillance du stimulus dans sa propre orientation.

Interactions longues

Les interactions longues se font au sein de chaque carte d'énergie et concernent les pixels voisins. Nous utilisons la convolution avec un masque "papillon" pour ces interactions :

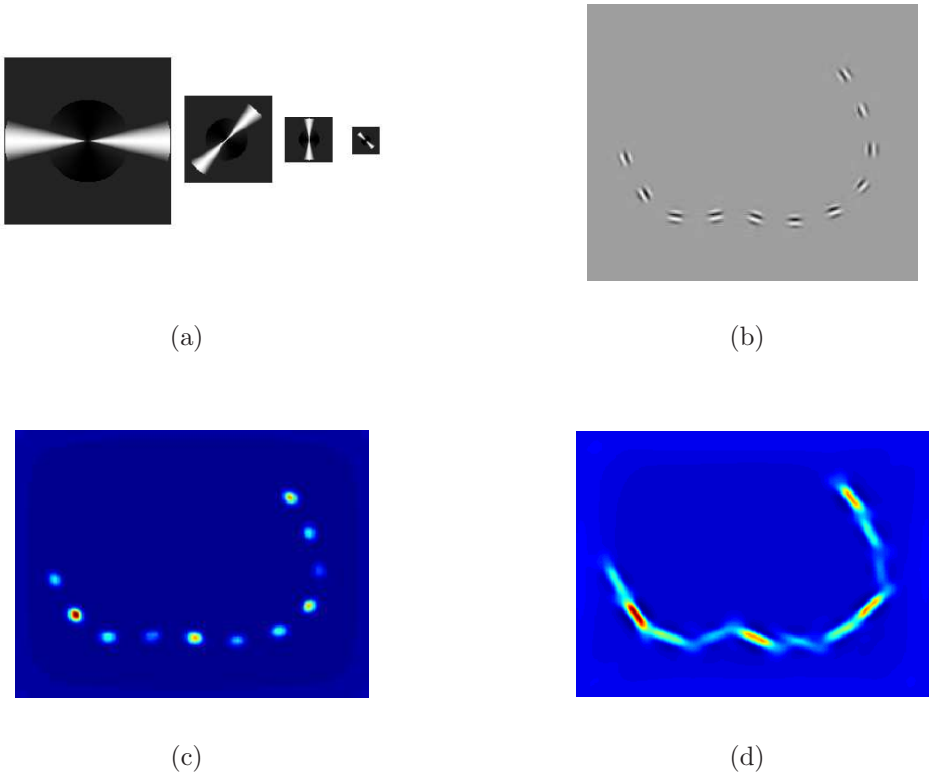
$$e^l_{ij} = e^s_{ij} \star B_{ij} \tag{2.19}$$

où e^l_{ij} est une carte d'énergie après les interactions longues, B_{ij} est le masque "papillon" pour l'orientation i et la fréquence j, et e^s_{ij} est une carte d'énergie après les

interactions courtes.

Un exemple des masque "papillon" à différentes tailles (fréquences) et orientations est présenté à la figure 2.18a. Les détails de la construction d'un masque "papillon" sont décrits à l'annexe A.

Chaque carte d'énergie est convoluée (Eq. 2.19) avec un masque "papillon" qui est adapté à cette carte à la fois en orientation et en taille. Ainsi, ce processus renforce les objets alignés et inhibe ceux qui ne le sont pas ; cela permet de faire ressortir les contours des stimuli. La figure 2.18 montre l'effet des interactions longues. Dans cet exemple, l'image initiale est aussi décomposée en 32 cartes d'énergie. Si on effectue la somme de ces cartes, on ne peut pas obtenir le contour formé par les stimuli (Fig. 2.18c). En revanche, si les interactions longues se font sur chacune des cartes, la somme des cartes d'énergie après ces interactions fait ressortir le contour (Fig. 2.18d).



(a) (b)

(c) (d)

FIG. 2.18 – Exemple des interactions longues : (a) Les masques "papillon" de différentes tailles et orientations ; (b) Stimuli ; (c) Somme des cartes d'énergie avant les interactions longues ; (d) Somme des cartes d'énergie après les interactions longues (voir texte).

3.2.4 Fusion

Dans les étapes précédentes, chacune des voies chromatiques et de luminance est décomposée en cartes d'énergie à différentes orientations et différentes fréquences.

L'étape de fusion est maintenant nécessaire pour créer une carte de saillance finale et unique. La fusion est réalisée à la fois en intra-voie (sommation des cartes d'énergie pour la voie de luminance et pour les voies de chrominance) et inter-voie (sommation entre les trois voies).

Pour chacune des voies, dans un premier temps, les cartes des différentes orientations dans une même bande de fréquence sont regroupées. Certaines études ont montré que le système visuel humain est plus sensible aux orientations horizontale et verticale qu'à celles obliques. Cela impliquerait une pondération plus importante pour les cartes provenant des orientations horizontale et verticale dans la fusion. Nous avons testé des fusions avec des pondérations différentes pour ces orientations mais n'avons pas observé au niveau de saillance de différence significative par rapport à une somme simple (sans pondération). Par conséquent, la carte de la bande de fréquence j est la somme simple des cartes dans toutes les orientations à cette bande :

$$e_j = \sum_{i=1}^{N} e_{ij}^l \tag{2.20}$$

La carte de chaque bande de fréquence e_j est ensuite segmentée par un seuillage dont le seuil est égal à la valeur moyenne $\bar{e_j}$ pour faire sortir clairement les zones saillantes :

$$e_j^t(x,y) = \begin{cases} 0 & \text{si } e_j(x,y) < \bar{e_j} \\ e_j(x,y) & \text{si } e_j(x,y) \geq \bar{e_j} \end{cases} \tag{2.21}$$

Les cartes provenant des différentes bandes de fréquence sont simplement sommées selon l'équation 2.22 pour générer une carte de caractéristique e_C et cela pour chaque voie : luminance et les deux de chrominance.

$$e_C = \sum_{j=1}^{M} e_j^t \tag{2.22}$$

Enfin, la carte d'énergie autour de la fréquence nulle, après avoir été normalisée par son *énergie totale*, est ajoutée à cette somme. Ainsi, nous avons trois cartes de caractéristique (une par voie) e_L, e_{RG}, e_{BY}.

Fusion des cartes de caractéristique

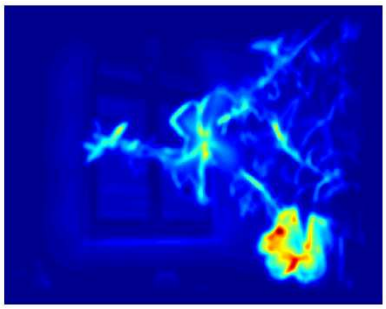
La fusion des cartes de caractéristique est une question délicate parce qu'elle implique des hypothèses sur les rôles spécifiques des voies dans la saillance. Plusieurs solutions ont été proposées pour combiner les différentes voies. Dans [Itti et al., 1998], Itti propose une fusion simple où les pondérations des trois voies sont égales. Au contraire, Le Meur [Le Meur et al., 2006] ne prend en compte que la voie de luminance pour la carte de saillance après avoir été renforcée par les voies chromatiques. Ici, dans un premier temps, la carte de saillance M_S est construite en fusionnant simplement les trois voies :

$$M_S = e_L + e_{RG} + e_{BY} \tag{2.23}$$

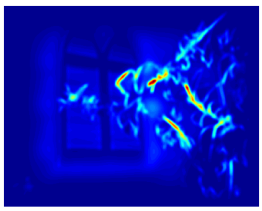
La figure 2.19 illustre la carte de saillance (M_S) et les cartes de caractéristique (e_L, e_{RG}, e_{BY}) pour une scène naturelle.

Nous reviendrons sur cette fusion simple en proposant une analyse détaillée de la contribution respective de chacune des cartes à la prédiction des fixations obtenues par une expérience d'oculométrie (cf. chapitre 4).

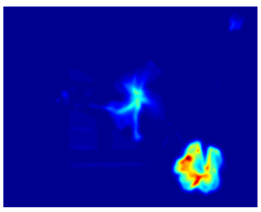
(a) Scène naturelle



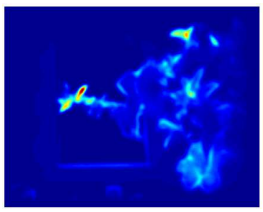
(b) Carte de saillance



(c) Carte de caractéristique e_L



(d) Carte de caractéristique e_{RG}



(e) Carte de caractéristique e_{BY}

FIG. 2.19 – Exemple de la carte de saillance. (a) Scène naturelle en couleur ; (b) Carte de saillance ; (c,d,e) Cartes de caractéristique. Sur les cartes de caractéristique et la carte de saillance, la couleur rouge représente la saillance la plus forte, la couleur bleu foncé la saillance la plus faible.

4 Conclusion

Dans ce chapitre, nous avons présenté toutes les étapes de notre modèle d'attention visuelle. Ce modèle inspiré de la biologie du système visuel permet le traitement de l'information visuelle depuis la rétine jusqu'au cortex visuel primaire. Ainsi, une image d'entrée est décomposée en une voie de luminance et deux voies chromatiques. Selon leurs caractéristiques psychophysiques, la voie de luminance est traitée différemment de celles chromatiques bien qu'elles se basent toutes les trois sur la même architecture. Le traitement de la voie de luminance est plus important que celui réalisé sur les voies chromatiques.

Notre modèle se distingue des précédents par un certain nombre de points. Pour la première fois, la rétine avec ses fonctions essentielles est modélisée. Nous avons simulé l'adaptation à la luminance et le réhaussement de contraste. Au niveau du cortex visuel une normalisation simple a été proposée en s'appuyant sur l'*énergie to-*

tale maximale de chaque orientation. De plus, les interactions courtes ont également été implémentées d'une manière linéaire pour renforcer une orientation spécifique des objets.

Le modèle permettant de calculer séparément la saillance de la voie de luminance et des voies chromatiques servira au chapitre 4 à tester l'influence respective de ces voies sur la prédiction des mouvements oculaires.

Chapitre 3

Méthodologie

1 Introduction

Ce chapitre a pour objectif de fournir les outils qui seront utilisés dans les chapitres suivants pour, d'une part, évaluer le modèle d'attention visuelle présenté au chapitre 2, et d'autre part, analyser les données expérimentales des mouvements oculaires. Dans un premier temps, nous allons présenter les critères permettant de comparer une carte de saillance et des zones fixées. Ces critères mesurent les correspondances entre les zones saillantes prédites par un modèle d'attention visuelle et celles réellement fixées par les sujets lors d'une exploration libre de scènes visuelles.

Dans un second temps, les données expérimentales des fixations oculaires seront étudiées à travers une modélisation statistique de leur distribution. Pour cela, nous utiliserons un modèle de mélange additif de fonctions estimé par l'algorithme "Expectation-Maximization". Deux versions seront proposées, une pour l'extraction des zones spatiales et l'autre pour l'extraction des facteurs possibles de guidage de l'attention visuelle.

2 Comparaison d'une carte de saillance et d'une carte de fixations

Un modèle d'attention visuelle crée la carte de saillance d'une scène ; celle-ci permet de prédire les zones ayant la plus forte probabilité d'attirer l'attention visuelle et donc les yeux des sujets. Pour évaluer la qualité d'une telle carte de saillance, il est nécessaire de comparer les zones mises en évidence par la carte avec les zones fixées par des sujets. Cette méthode a très souvent été utilisée dans la littérature [Itti et al., 1998; Parkhurst et al., 2002; Tatler et al., 2005; Peters et al., 2005; Torralba et al., 2006; Le Meur et al., 2006]. Ainsi, plusieurs critères ont été proposés parmi lesquels figurent le ROC (*"Receiver Operation Characteristic"*) [Tatler et al., 2005], le NSS (*"Normalized Scanpath Saliency"*), le "percentile", la divergence de Kullback-Leibler [Peters et al., 2005] et le taux de fixations correctes [Torralba et al., 2006]. Le principe de ces critères est de mesurer les correspondances entre les zones saillantes prédites par un modèle d'attention visuelle et celles fixées par des sujets humains. Plus la correspondance spatiale est grande, plus le modèle est dit perfor-

mant et donc, plus la carte de saillance qu'il fournit permettra une bonne prédiction des fixations oculaires.

Dans la suite, nous allons présenter les critères les plus souvent utilisés dans la littérature pour mesurer les correspondances entre les zones prédites par une carte de saillance et les zones fixées par des sujets pour une image.

2.1 Le "Percentile"

Ce critère est utilisé par exemple dans [Peters and Itti, 2008] pour mesurer les correspondances entre les zones saillantes d'une carte de saillance $M_S(\mathbf{x})$ d'une image et les zones de cette image effectivement fixées par des sujets. Les zones fixées sont représentées par les points de fixation $\{\mathbf{x}_i\}$ réalisées par des sujets sur cette image. Pour une carte de saillance donnée, le "'percentile" $P(\mathbf{x}_i)$ pour la fixation $\mathbf{x}_i$ est calculé par le pourcentage des pixels de l'image dont la saillance est inférieure à la saillance de cette fixation :

$$P(\mathbf{x}_i) = 100\frac{|\{\mathbf{x} \in X : M(\mathbf{x}) < M(\mathbf{x}_i)\}|}{|X|} \tag{3.1}$$

où $|.|$ représente le cardinal d'un ensemble et X est l'ensemble des pixels sur la carte de saillance.

Le "percentile" suppose que plus la saillance d'une zone est grande, plus cette zone correspond à une zone fixée par des sujets. Ainsi, un "percentile" plus important représente une meilleure correspondance entre la carte de saillance et les zones fixées. Ce critère obtient des valeurs comprises entre 0 et 100. Un "percentile" égal à 50 représente une correspondance équivalente au hasard. Nous notons que le "percentile" est invariant à une transformation monotone croissante de la carte de saillance. Pour mesurer le "percentile" d'une carte de saillance et d'un ensemble de fixations, ce critère est calculé pour chaque fixation, et ensuite moyenné sur toutes les fixations.

2.2 Le "Normalized Scanpath Saliency" (NSS)

Ce critère a été proposé par Peters [Peters et al., 2005; Peters and Itti, 2008]. Pour évaluer le NSS, une carte de saillance est d'abord normalisée (centrée-réduite) pour avoir une moyenne nulle et un écart-type unité. La valeur de NSS pour une fixation correspond à la valeur en cette fixation sur la carte de saillance normalisée :

$$NSS(\mathbf{x}_i) = \frac{1}{\sigma}\left(M_S(\mathbf{x}_i) - \mu\right) \tag{3.2}$$

où μ est la moyenne de la carte de saillance initiale : $\mu = \frac{1}{|X|}\sum_{\mathbf{x}\in X} M_S(\mathbf{x})$
σ est l'écart-type de la carte de saillance initiale : $\sigma = \sqrt{\frac{1}{|X|-1}\sum_{\mathbf{x}\in X}\left(M_S(\mathbf{x}) - \mu\right)^2}$.

La valeur de NSS égale à 0 indique qu'il n'y a pas de correspondance entre les zones prédites par la carte de saillance et les zones fixées. Plus la valeur de NSS est positive, plus cette correspondance est forte, tandis qu'une valeur de NSS

négative représente une "anti-correspondance". Le NSS n'est pas invariant à une transformation monotone croissante de la carte de saillance. Le NSS est calculé pour chaque fixation et moyenné ensuite pour toutes les fixations réalisées sur une image.

2.3 Le taux de fixations correctes

Ce critère a initialement été proposé par Torralba [Torralba et al., 2006]. La carte de saillance M_S est segmentée en deux types de zones : les zones saillantes et les zones non-saillantes. Cette segmentation est effectuée grâce à un seuil ξ qui est défini comme la valeur de saillance au dessus de laquelle il y a 20% des pixels de la carte (Fig. 3.1b). Ainsi, les pixels dont la saillance est supérieure à ce seuil appartiennent aux zones saillantes ; le reste des pixels appartient aux zones non-saillantes. Le taux de fixations correctes $T(M_S, F)$ mesure le pourcentage des fixations F qui sont localisées dans les zones saillantes d'une carte de saillance M_S. Ce critère est calculé selon l'équation 3.3 en projetant les fixations sur la carte de saillance segmentée.

$$T(M_S, F) = \frac{N_s}{N_t} 100\% \tag{3.3}$$

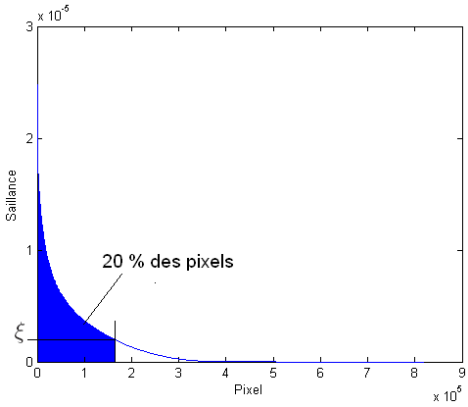
avec M_S la carte de saillance, F un ensemble de fixations, N_s le nombre de fixations localisées dans les zones saillantes précédemment définies et N_t le nombre total de fixations.

Le taux de fixations correctes est borné entre 0 et 100. Plus il est grand, plus la capacité de prédiction du modèle est bonne. La valeur égale à 20 correspond à une prédiction équivalente au hasard. De plus, le taux de fixations correctes est invariant à une transformation monotone croissante de la carte de saillance. La figure 3.1 illustre la manière de calculer le taux de fixations correctes. Après la segmentation de la carte en zones saillantes et en zones non-saillantes, la valeur de saillance est mise à 1 dans les zones saillantes et à 0 dans les zones non-saillantes. Ensuite, le taux de fixations correctes est calculé comme le pourcentage de fixations localisées dans les zones saillantes sur le nombre total de fixations.

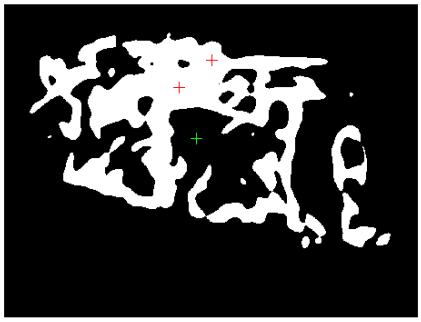
2.4 Le "Receiver Operating Characteristic" (ROC)

Le critère ROC est caractérisé par la capacité d'une carte de saillance à distinguer des fixations des points aléatoires. Soit ξ le seuil de décision tel que si la valeur de la saillance associée à une position spatiale est supérieure à ce seuil, alors cette position spatiale appartiendra à la zone saillante, sinon ce sera la zone non-saillante. En comparant avec les fixations expérimentales et les points aléatoires, on obtient un taux de classification des "vrais positifs" et un taux de classification des "faux positifs". Un "vrai positif" est obtenu lorsqu'une fixation est dans la zone saillante et un "faux positif" lorsqu'un point aléatoire est dans la zone saillante (Tab. 3.1). En faisant varier le seuil ξ de décision entre la valeur minimale et maximale de la carte de saillance, on obtient la courbe ROC représentant l'évolution du taux de classification des "vrais positifs" en fonction des "faux positifs".

(a)



(b)



(c)

FIG. 3.1 – Critère du taux de fixations correctes : (a) Carte de saillance ; (b) Les pixels sont classifiés selon leur saillance ; (c) Carte de saillance binarisée : les zones saillantes sont colorées en blanc. Les fixations sont projetées sur cette carte pour calculer le taux de fixations correctes.

Ce critère dispose de plusieurs avantages [Tatler et al., 2005] parmi lesquels figure la capacité à décrire la différence de saillance aux fixations et aux points aléatoires. Néanmoins, le choix des points aléatoires est critiqué car il ne prend pas forcément en compte les contraintes dues aux mouvements oculaires réels. Une des contraintes observées dans les mouvements oculaires est ce que l'on appelle "biais de centralité" ; on observe une tendance des sujets à regarder plus au centre d'une image quelque soit la tâche ou le contenu de l'image. Or si les points aléatoires sont choisis en utilisant une distribution uniforme, ils ne refléteront pas cette tendance. De plus, à partir d'une distribution aléatoire des fixations, on ne peut pas retrouver les tendances classiques sur les distributions des amplitudes de saccades, des directions, etc. Pour résoudre ce problème, les points ne sont pas générés aléatoirement pour une image mais correspondent aux fixations obtenues pour un même sujet mais pour une autre image [Reinagel and Zador, 1999; Parkhurst et al., 2002; Tatler et al., 2005; Frey et al., 2008].

Une carte de saillance est un bon prédicteur des fixations si sa courbe ROC est

TAB. 3.1 – Matrice de confusion obtenue lors de la classification des fixations et des points aléatoires en zones saillantes ou non.

	Fixations	Points aléatoires
Zones saillantes	**Vrais positifs**	**Faux positifs**
Zones non-saillantes	**Faux négatifs**	**Vrais négatifs**

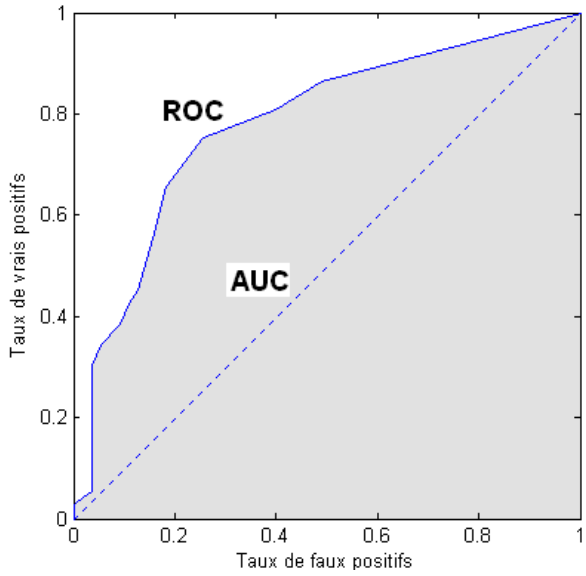


FIG. 3.2 – Exemple du critère ROC pour mesurer la capacité d'une carte de saillance à distinguer des fixations des points aléatoires. La carte de saillance est celle de l'image de la figure 3.1. Les points aléatoires utilisés dans le calcul du critère ROC sont choisis à partir des fixations réalisées sur une autre image. Le critère ROC est calculé par l'aire au-dessous de la courbe ROC.

au dessus de la diagonale (Fig. 3.2). La courbe ROC diagonale est obtenue pour une carte de saillance aléatoire ou autrement dit, pour une carte ne permettant pas de distinguer des fixations des points aléatoires. L'aire au-dessous de la courbe ROC (AUC, *"Area Under the Curve"*) est utilisée comme variable descriptive d'une courbe ROC, c'est le critère ROC. Plus la valeur d'AUC est grande, plus la qualité de la carte de saillance est bonne. La valeur maximale d'AUC est égale à 1 lorsque la courbe ROC passe par le point (0,1). Le critère ROC est invariant par rapport à une transformation monotone croissante de la carte de saillance.

2.5 La divergence de Kullback-Leibler

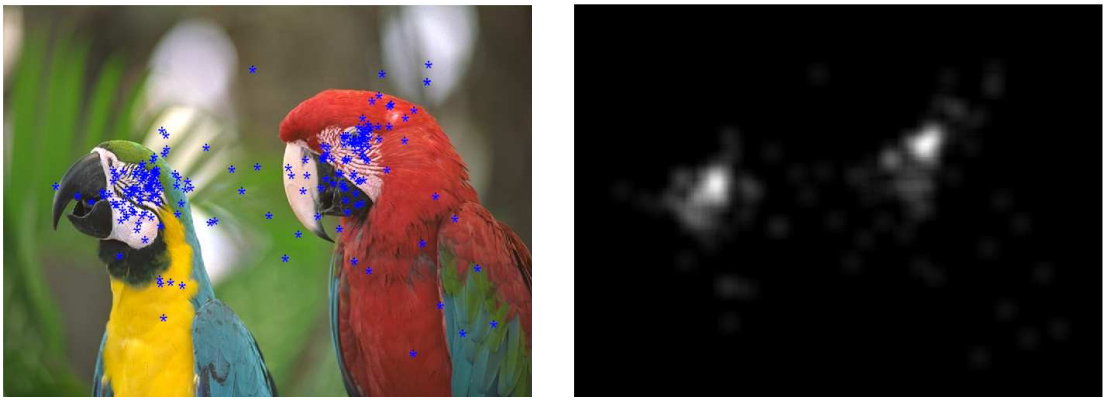
L'ensemble des zones fixées par des sujets pour une image crée une distribution spatiale ou une carte de densité des fixations (Fig. 3.3). Nous pouvons évaluer la correspondance entre une carte de saillance, considérée comme une carte de densité, et une carte de densité de fixations en comparant ces deux cartes de densité. Une méthode classique pour évaluer cette correspondance est la divergence de Kullback-Leibler [Kullback and Leibler, 1951]. La divergence de Kullback-Leibler de la carte de densité de fixations $M_F(\mathbf{x})$ par rapport à la carte de saillance $M_S(\mathbf{x})$ issue d'un modèle d'attention visuelle est définie par :

$$D_{KL}(M_S||M_F) = \sum_{\mathbf{x}\in X} M_S(\mathbf{x}) \log \frac{M_S(\mathbf{x})}{M_F(\mathbf{x})} \tag{3.4}$$

Cette divergence n'est pas symétrique. La divergence de Kullback-Leibler est donc utilisée sous la forme suivante pour qu'elle soit symétrique :

$$D_{KL}(M_S, M_F) = 0.5(D_{KL}(M_S||M_F) + D_{KL}(M_F||M_S)) \tag{3.5}$$

Plus la divergence Kullback-Leibler est faible, plus les correspondances entre les zones prédites par la carte de saillance et les zones fixées sont importantes et inversement. Néanmoins, la divergence de Kullback-Leibler n'est pas bornée. Elle n'est pas non plus invariante à une transformation monotone croissante de la carte de saillance. En pratique, comme les critères de ROC et de taux de fixations correctes, le calcul de la divergence de Kullback-Leibler est effectué pour l'ensemble des fixations réalisées sur une image.



(a) (b)

FIG. 3.3 – Exemple de la carte de densité des fixations : (a) Image avec les fixations (étoiles) réalisées par tous les sujets lors de l'exploration libre de l'image pendant 3 s ; (b) Carte de densité de fixations créée par la méthode des noyaux de Parzen (sommation des fonctions gaussiennes centrées à chacune des fixations).

2.6 Résumé

Nous avons présenté les critères souvent utilisés dans la littérature pour mesurer les correspondances entre la carte de saillance d'une image issue d'un modèle

TAB. 3.2 – Tableau de comparaison des critères.

Critères	Borné	Points aléatoires	Utilisation des fixations	Invariant par rapport à une transformation monotone croissante
“Percentile”	Oui	Non	Chaque fixation	Oui
NSS	Non	Non	Chaque fixation	Non
Kullback-Leibler	Non	Non	Plusieurs fixations	Non
Taux de fixations correctes	Oui	Non	Plusieurs fixations	Oui
ROC	Oui	Oui	Plusieurs fixations	Oui

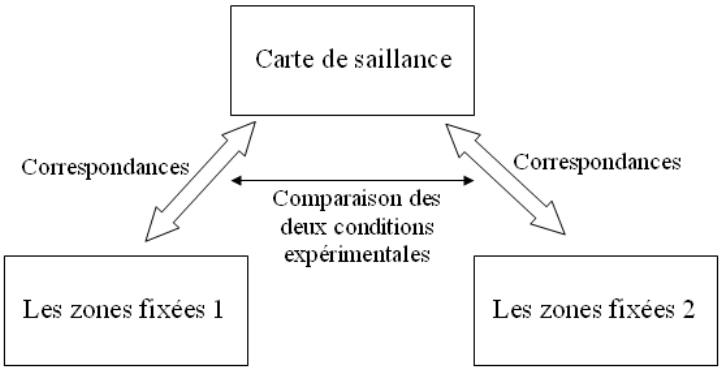
d’attention visuelle et les zones fixées sur cette image par des sujets. Ces critères se distinguent les uns des autres par plusieurs aspects. La première différence majeure concerne la dynamique des critères. Certains critères ont des valeurs appartenant à un intervalle fixé, souvent entre 0 et 100, comme le ROC, le “percentile” ou le taux de fixations correctes. D’autres critères ne sont pas bornés comme le NSS ou la divergence de Kullback-Leibler. La deuxième différence majeure est liée à la manière d’utiliser les fixations dans le calcul du critère. Plusieurs critères ne prennent en compte que les fixations comme le NSS, la divergence de Kullback-Leibler, le “percentile” ou le taux de fixations correctes. En revanche, le ROC se sert à la fois des fixations réalisées par des sujets et de points aléatoires. Ces critères se différencient également par le nombre de fixations prises en compte dans le calcul du critère. Le NSS et le “percentile” peuvent se calculer pour chaque fixation ; tandis que le taux de fixations correctes, le ROC ou la divergence de Kullback-Leibler prennent en compte toutes les fixations. Le tableau 3.2 résume ces différents aspects de ces critères.

Dans les chapitres suivants, afin d’évaluer les correspondances entre les zones prédites par une carte de saillance et les zones fixées, nous avons choisi de ne retenir que deux critères : le ROC et le taux de fixations correctes. Ces deux critères sont bornés et ne dépendent donc pas de la dynamique d’une carte de saillance ; ce qui convient à la comparaison de différentes cartes de saillance.

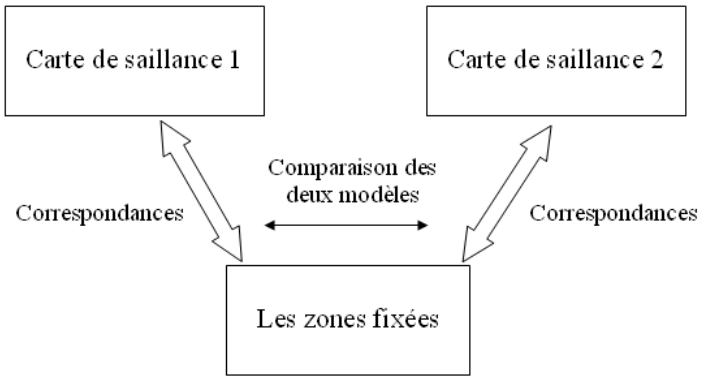
3 Les tests d'hypothèse

Nous venons de passer en revue quelques critères et nous en avons choisi deux pour mesurer les correspondances entres les zones prédites par une carte de saillance et celles fixées par des sujets. Dans les chapitres suivants, nous voulons tester ces correspondances dans des *scénarii* différents en fonction des conditions expérimentales (cf. chapitre 4) ou différents en terme de programmation de saccade (cf. chapitre 5) (Fig. 3.4). Au chapitre 4, la comparaison se fera entre deux expériences différentes : une pour les images en couleur et l'autre pour les mêmes images en niveau de gris. Ainsi, nous comparerons la même carte de saillance d'une image avec les zones fixées pour l'image en couleur et avec les zones fixées pour la même image en niveau de gris. Nous aurons donc deux valeurs de critères. Comment savoir si ces deux valeurs sont significativement différentes ? De la même façon, au chapitre 5, nous comparerons différents modèles de saillance pour une même image avec les mêmes zones fixées sur cette image.

Dans la suite, nous présentons rapidement les tests d'hypothèse qui nous serviront.



(a)



(b)

FIG. 3.4 – Comparaison des correspondances entre une carte de saillance et des zones fixées pour deux *scénarii* différents. (a) Deux expériences oculométriques ; (b) Deux modèles de saillance.

3.1 Le test de Student (t-test)

Le t-test permet de tester la différence entre les moyennes de deux échantillons en supposant *a priori* que ces deux échantillons proviennent de populations parentes qui suivent des lois normales et de variances homogènes. Ce test s'effectue à l'aide d'une loi "Student".

L'hypothèse nulle H_0 est "les moyennes de deux populations sont égales". Dans nos travaux, le t-test est réalisé avec un niveau de signification $\alpha = 0.05$ et de manière bilatérale, qui correspond à l'hypothèse alternative H_1 " les deux moyennes sont différentes". Le t-test peut être appliqué pour deux échantillons de tailles égales ou différentes.

3.2 Le test de Kolmogorov - Smirnov (KS-test)

Le t-test nécessite la "normalité" des échantillons. Pourtant, dans certains cas, cette condition n'est pas respectée. Par exemple, c'est le cas pour un critère qui se base sur une carte de saillance car les caractéristiques de bas niveau qui constituent la saillance ne suivent pas une distribution normale [Baddeley, 1996; Frey et al., 2008]. Ainsi, le test non-paramétrique de Kolmogorov-Smirnov (KS-test) est préféré car il ne demande *a priori* aucune condition préalable. Ce test repose sur la fonction de répartition empirique d'un échantillon.

L'hypothèse nulle H_0 du KS-test est "les deux échantillons suivent la même loi ou distribution". Comme le t-test, nous choisissons le KS-test avec un niveau de signification $\alpha = 0.05$ et de manière bilatérale. L'hypothèse H_1 est donc "les deux échantillons ne suivent pas la même loi". Le nombre d'observations de deux échantillons peut être différent.

3.3 L'intervalle de confiance

Les valeurs estimées sont souvent présentées avec un intervalle de confiance qui représente le degré de confiance de cette valeur. Lorsque l'on veut estimer un paramètre de la population qui suit une distribution déterminée, l'intervalle de confiance pour ce paramètre peut être trouvé pour cette distribution. Pourtant, dans le cas général, la distribution n'est pas connue *a priori*. L'intervalle de confiance peut être calculé dans ce cas-là par une technique de *"Bootstrap"* [Efron and Tibshirani, 1993]. Cette dernière permet de trouver l'intervalle de confiance d'un paramètre à partir d'un jeu de données en reéchantillonnant ce jeu et, en recalculant la valeur de ce paramètre sur ce nouvel échantillon. Cette démarche est appliquée N fois[1]. L'échantillonnage est effectué aléatoirement avec remplacement ("*random sampling with replacement*"), c'est-à-dire, qu'un même individu peut apparaître plus d'une fois. De plus, la taille du nouvel échantillon est la même que l'ancien. Ainsi, on obtient une distribution du paramètre à partir de laquelle l'intervalle de confiance $[a, b]$ à $\alpha\%$ est déterminé selon l'équation 3.6. La figure 3.5 illustre la distribution

[1]Dans cette thèse, la technique Bootstrap est utilisée avec $N = 10000$

d'un paramètre obtenue par la technique Bootstrap et son intervalle de confiance à 95%.

$$Pr(x < a) = \frac{100 - \alpha}{2} \text{ et } Pr(x > b) = \frac{100 - \alpha}{2} \tag{3.6}$$

avec x le paramètre à estimer.

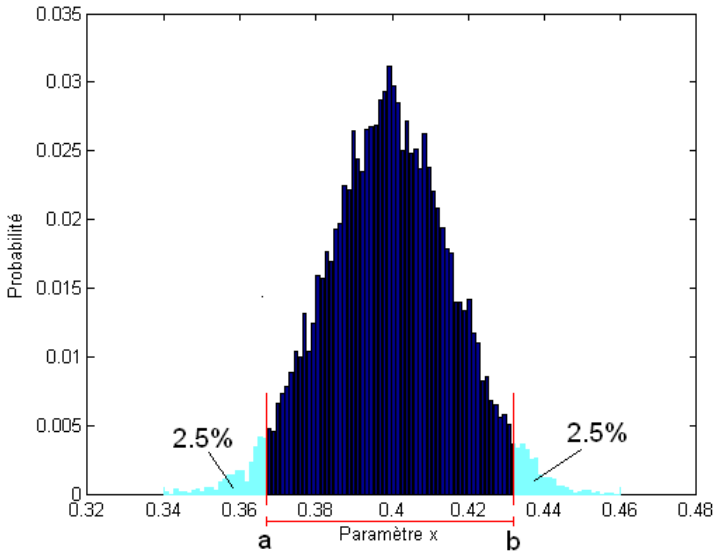


FIG. 3.5 – Intervalle de confiance à 95% calculé à partir d'une distribution d'un paramètre obtenue par reéchantillonnage "Bootstrap".

4 Modèle de mélange de fonctions gaussiennes

Le modèle de mélange de fonctions gaussiennes permet d'estimer la distribution d'une variable aléatoire en utilisant la somme de plusieurs fonctions gaussiennes. Ce modèle est utilisé dans la littérature par exemple pour l'indexation d'images en modélisant la distribution des descripteurs d'images comme la position spatiale, la chrominance [Guérin-Dugué et al., 2001; Biernacki and Mohr, 1999]. Ici, le modèle de mélange de fonctions gaussiennes est utilisé pour estimer et paramétrer la distribution spatiale d'un ensemble de points à deux dimensions, les fixations acquises lors d'une expérience oculométrique. Le principe de cette modélisation est de considérer la distribution empirique des points de fixation comme un mélange additif de fonctions gaussiennes élémentaires. Ces fonctions gaussiennes (ou modes gaussiens) devraient représenter chaque accumulation de points de fixation dans une zone spatiale localisée. On obtient ainsi pour une image vue par plusieurs sujets, l'ensemble des zones spatiales sur lesquelles les fixations oculaires se sont concentrées. Le modèle

de densité peut s'écrire suivant l'équation :

$$\begin{aligned} f(\mathbf{x}|\Theta) &= \sum_{k=1}^{K} p_k h_k(\mathbf{x}|\theta_k) \\ &= \sum_{k=1}^{K} p_k G_k(\mathbf{x}|\mu_k, \Sigma_k) \end{aligned} \tag{3.7}$$

où la première ligne correspond à un modèle additif de fonctions h_k quelconques paramétrées par θ_k et la deuxième ligne correspond à un modèle additif de fonctions gaussiennes, avec :
K le nombre de modes gaussiens
Θ l'ensemble des paramètres du modèle
G_k le mode gaussien k
p_k, μ_k et Σ_k la contribution et le paramétrage (moyenne et matrice de covariance) du mode gaussien k.

L'algorithme EM (*"Expectation-Maximization"*) proposé par Dempster [Dempster et al., 1977] permet d'estimer les paramètres du modèle de mélange de fonctions gaussiennes pour un nombre K de modes fixé *a priori*. Pour estimer ce nombre, on utilise une procédure de sélection de modèles. Elle sera appliquée après.

Cette modélisation présente une difficulté en présence de données bruitées pouvant affaiblir l'hypothèse de regroupement des observations suivant des modes gaussiens distincts. C'est peut-être le cas pour les données oculométriques qui peuvent représenter des mouvements oculaires aléatoires ou des mouvements oculaires non commandés par le contenu de l'image. Nous avons donc choisi de rajouter un mode supplémentaire pour modéliser les possibles fixations "bruitées" en prenant comme hypothèse une distribution uniforme. Ainsi, les fixations oculaires sont modélisées par un mélange de K modes gaussiens et un mode uniforme :

$$\begin{aligned} f(\mathbf{x}|\Theta) &= \sum_{k=1}^{K+1} p_k h_k(\mathbf{x}|\theta_k) \\ &= \sum_{k=1}^{K} p_k G_k(\mathbf{x}|\mu_k, \Sigma_k) + p_{K+1} U(\mathbf{x}) \end{aligned} \tag{3.8}$$

où $\Theta = (\theta_1, .., \theta_K, \theta_{K+1}) = (p_1, \mu_1, \Sigma_1, .., p_K, \mu_K, \Sigma_K, p_{K+1})$ les paramètres à estimer par l'algorithme EM.
$h_{K+1} = U$ la distribution uniforme sur le support spatial des points $\mathbf{x}$ (l'étendue de l'image) tel que : $\int_{-\infty}^{+\infty} U(\mathbf{x}) d\mathbf{x} = 1$.

Les paramètres Θ sont estimés itérativement par l'algorithme EM. Après initialisation avec $\Theta^{(0)}$, chaque itération m est constitué de 2 étapes :

Etape E (*Expectation*). Dans cette étape, le tableau $T^{(m)}$ de $t_{ik}^{(m)}$, $i = 1..N$,

$k = 1..K+1$ est calculé :

$$\begin{aligned} t_{ik}^{(m)} &= \text{Prob}\left\{\mathbf{x}_i \text{ vient de la source } k\right\} \\ &= \frac{p_k^{(m-1)} h_k(\mathbf{x}_i|\theta_k)}{\sum_{l=1}^{K+1} p_l^{(m-1)} h_l(\mathbf{x}_i|\theta_l)} \end{aligned} \tag{3.9}$$

Etape M (*Maximization*). La méthode de maximum de vraisemblance est utilisée pour estimer Θ en maximisant l'expression F dans l'équation 3.10 :

$$F(\Theta|\mathbf{x}_1, .., \mathbf{x}_N, T^{(m)}) = \sum_{i=1}^{N} \sum_{k=1}^{K+1} t_{ik}^{(m)} \ln\left(p_k h_k(\mathbf{x}_i|\theta_k)\right) \tag{3.10}$$

Finalement, les paramètres Θ sont mis à jour par les équations 3.11, 3.12, 3.13.

$$p_k^{(m)} = \frac{\sum_{i=1}^{N} t_{ik}^{(m)}}{N} \text{ avec } k = 1..K+1 \tag{3.11}$$

$$\mu_k^{(m)} = \frac{\sum_{i=1}^{N} t_{ik}^{(m)} \mathbf{x}_i}{\sum_{i=1}^{N} t_{ik}^{(m)}} \text{ avec } k = 1..K \tag{3.12}$$

$$\Sigma_k^{(m)} = \frac{\sum_{i=1}^{N} t_{ik}^{(m)} (\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)'}{\sum_{i=1}^{N} t_{ik}^{(m)}} \text{ avec } k = 1..K \tag{3.13}$$

A la convergence, nous obtenons tous les paramètres de Θ. Ainsi, la distribution des fixations est complètement décrite par ces paramètres selon l'équation (Eq. 3.8).

Les équations ci-dessus permettent d'estimer tous les paramètres du modèle pour un nombre K de modes gaussiens. Il est donc nécessaire de fixer au préalable ce nombre. Cela est effectué en amont par une étape de sélection de modèles. Cette sélection est effectuée en utilisant un critère d'information qui combine la vraisemblance et une mesure de complexité. En effet, lorsque le nombre de paramètres du modèle augmente, la vraisemblance augmente également. Néanmoins, un trop grand nombre de paramètres peut entraîner le surapprentissage qui risque de capturer du bruit. Ce problème est pris en compte pour choisir le meilleur modèle en utilisant le critère BIC (*"Bayesian Information Criterion"*) [Schwarz, 1978]. L'objectif est de minimiser le critère BIC :

$$BIC(\Theta|X) = -2L(\Theta|X) + \upsilon \ln(N) \tag{3.14}$$

avec Θ l'ensemble des paramètres
υ le nombre de paramètres
$X = \{\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_N\}$ l'ensemble des points de fixation
N le nombre de points de fixation
L la log-vraisemblance du modèle et $L(\Theta|X) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} p_k G_k(\mathbf{x}_i|\mu_k, \Sigma_k)\right)$.

Dans nos simulations, la valeur de BIC est multipliée par $-\frac{1}{2}$. Ainsi, le critère à maximiser est $\left\{L - \frac{1}{2}\upsilon \ln(N)\right\}$.

TAB. 3.3 – Les paramètres pour les distributions théoriques et les paramètres estimés par le mélange de fonctions gaussiennes (MG).

	D1	D2	D3	D4	D5
Centre théorique ($[x; y]$)	$[300; 200]$	$[800; 200]$	$[350; 500]$	$[750; 450]$	
Pondération théorique	0.1538	0.1538	0.3077	0.2308	0.1538
Centre estimé par MG ($[x; y]$)	$[297; 200]$	$[798; 206]$	$[350; 502]$	$[747; 451]$	
Pondération estimé par MG	0.1564	0.1566	0.3052	0.2302	0.1517

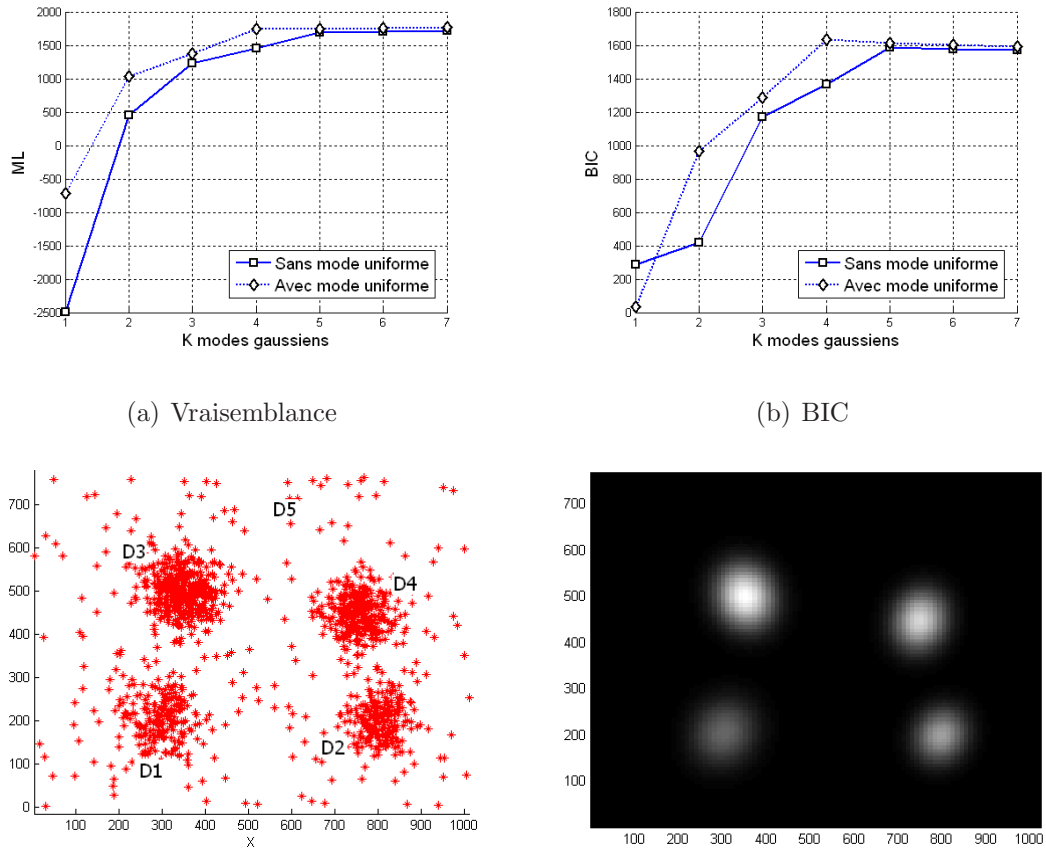
Test sur des données artificielles

Nous avons testé le modèle de mélange de fonctions gaussiennes pour un jeu de données artificielles qui contient des points engendrés par 4 distributions normales et une distribution uniforme dont les paramètres sont données au tableau 3.3. La distribution spatiale du jeu de données est affichée à la figure 3.6c. Le jeu de données a été testé avec plusieurs modèles comportant différents modes gaussiens et avec ou sans mode uniforme. A la figure 3.6a,b sont présents les critères de vraisemblance et BIC. Nous pouvons vérifier que la vraisemblance augmente avec le nombre de modes gaussiens. Le critère BIC permet de choisir le meilleur modèle : ici un mélange de 4 fonctions gaussiennes. Nous notons que pour un même nombre de modes gaussiens, le modèle avec mode uniforme donne bien ici un meilleur critère que celui sans mode uniforme, puisque dans ce jeu de données il y avait une contribution aléatoire uniforme. La fonction de densité estimée est illustrée à la figure 3.6d.

Dans le tableau 3.3, nous remarquons que les grandeurs estimées correspondent aux grandeurs attendues. Notons de plus, que pour tenir compte de la sensibilité de l'algorithme EM aux conditions initiales (et d'autant plus si les données sont bruitées), nous avons adopté la démarche suivante. Pour chaque valeur du nombre K de modes gaussiens, nous effectuons 5 simulations de l'algorithme EM pour des conditions initiales différentes et nous prenons le meilleur au sens de la vraisemblance. Ce nombre de 5 est un compromis tenant compte du temps de calcul.

Test sur des données oculométriques

Pour l'étude des fixations oculaires, cette modélisation nous permet de relever dans les images les zones sur lesquelles les modes gaussiens se sont placées et de les analyser relativement à leur position, leur variance et leur saillance en les associant aux valeurs de saillance fournies par le modèle de saillance. La figure 3.7 montre un exemple du mélange de fonctions gaussiennes pour modéliser la distribution spatiale des fixations oculaires effectuées sur une image par 26 sujets.



(a) Vraisemblance

(b) BIC

(c) Jeu de données

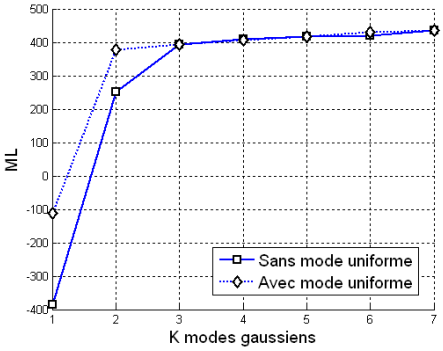
(d) Carte de densité

FIG. 3.6 – Modèle de mélange de fonctions gaussiennes pour un jeu de données artificielles : (a & b) Illustration de l'étape de sélection de modèles. Les critères de vraisemblance et BIC en fonction du nombre de modes gaussiens et en considérant des modèles avec ou sans mode uniforme. (c) Jeu de données. (d) Carte de densité représentée par les modes gaussiens du meilleur modèle.

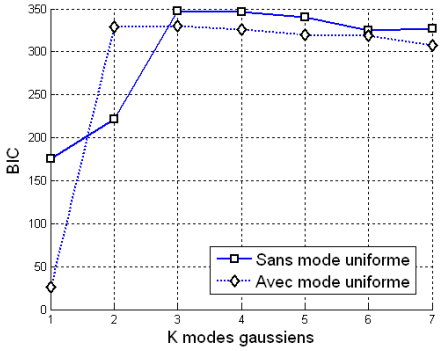
5 Modèle "EM carte"

La modélisation précédente permet d'analyser les fixations oculaires pour une condition donnée, image par image, pour tous les sujets. Les résultats de l'analyse seront donc dépendants de l'image. Si l'on s'intéresse maintenant à comprendre les facteurs qui guident l'attention visuelle, il est nécessaire de regrouper pour une condition donnée, toutes les images pour tous les sujets. Cette approche a été proposée dans [Vincent et al., 2009] ; nous allons la reprendre ici, en l'adaptant à notre cas d'étude.

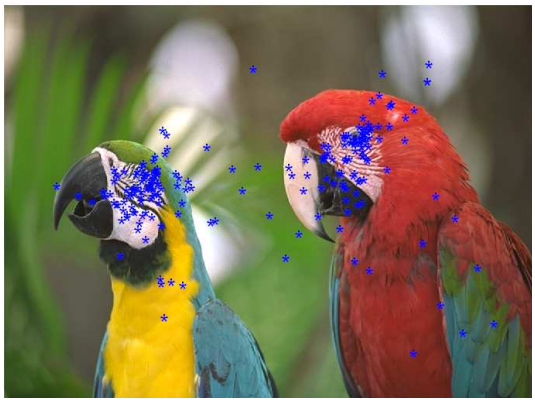
Le principe est le suivant. Nous recherchons à partir des fixations oculaires de tous les sujets, de toutes les images, pour une condition donnée, les facteurs de guidage $h_k(\mathbf{x})$ qui ont contribué à la fixation $\mathbf{x}$ observée. L'hypothèse est que l'interaction entre les différents facteurs est additive. Cette hypothèse peut paraître simple et réductrice dans un contexte très général d'interaction entre des facteurs de bas niveau et haut niveau. Dans notre cas, le contexte d'étude est la saillance visuelle *"Bottom-Up"* et les facteurs de guidage sont de type bas niveau et ils interagissent



(a) Vraisemblance



(b) BIC



(c) Jeu de données

(d) Carte de densité

FIG. 3.7 – Modèle de mélange de fonctions gaussiennes pour un jeu de données oculométriques : (a & b) Illustration de l'étape de sélection de modèles. Les critères de vraisemblance et BIC en fonction du nombre de modes gaussiens et en considérant des modèles avec ou sans mode uniforme. (c) Image superposée des fixations réalisées par tous les sujets. (d) Carte de densité représentée par les modes gaussiens du meilleur modèle.

suivant une fusion additive dans le modèle d'attention visuelle (cf. chapitre 2), ce qui justifie cette hypothèse dans le modèle statistique développé ici. Ainsi, ces facteurs de guidage sont communs pour toutes les images ; par exemple, la luminance passe-bande ou la chrominance Rouge-Vert basse fréquence. Pour chaque image, les facteurs de guidage sont extraits et représentés par des cartes de caractéristique ; cela explique le nom du modèle statistique "EM carte".

Dans le modèle "EM carte", les fixations sont donc expliquées par les facteurs de guidage selon un modèle additif. La densité pour la fixation $\mathbf{x}$ est donnée par l'équation suivante :

$$f(\mathbf{x}|\Theta) = \sum_{k=1}^{K} p_k h_k(\mathbf{x}) \tag{3.15}$$

avec $\Theta = (p_1, .., p_K)$ paramètres à estimer,
K le nombre de facteurs de guidage dans le modèle "EM carte", et

TAB. 3.4 – Les pondérations des distributions théoriques et les pondérations estimées par le modèle "EM carte".

	D1	D2	D3	D4	D5
Pondération théorique	0.1538	0.1538	0.3077	0.2308	0.1538
Pondération "EM carte"	0.1519	0.1490	0.3057	0.2347	0.1587

h_k $k$ième facteur de guidage.

Contrairement au modèle de mélange de fonctions gaussiennes, qui comporte les modes gaussiens élémentaires représentant les zones spatiales localisées et qui dépend donc de l'image, le modèle "EM carte" est indépendant de l'image. Ainsi, les observations utilisées dans le modèle "EM carte" regroupent les fixations de toutes les images et de tous les sujets. Dans ce modèle, les facteurs de guidage h_k sont connus *a priori*. Le facteur de guidage $h_k(\mathbf{x}_i)$ pour la fixation $\mathbf{x}_i$ est calculé dans l'image sur laquelle cette fixation a été réalisée. L'objectif est alors de trouver les pondérations p_k de ces facteurs (Eq. 3.15). Nous reprenons l'algorithme EM mais dans une version simplifiée pour estimer les paramètres de Θ. Alors que l'étape E ne change pas, l'étape M est plus simple car nous ne nous intéressons plus qu'au calcul des pondérations. Les pondérations p_k peuvent donc être estimées à l'aide de l'équation 3.11. A la sortie du modèle "EM carte", nous obtenons les pondérations des K facteurs de guidage. Ces pondérations représentent les contributions des facteurs aux fixations oculaires.

Pour tester le modèle "EM carte", nous reprenons le jeu de données artificielles ci-dessus et les 4 cartes pré-définies correspondant aux 4 modes gaussiens et une carte uniforme ; ces cartes ont engendré le jeu de données. Ces 5 cartes représentent les facteurs de guidage du modèle "EM carte". A la convergence du modèle "EM carte", nous obtenons donc uniquement les pondérations des facteurs représentées dans le tableau 3.4. Selon ces résultats, les pondérations des facteurs de guidage estimées par le modèle "EM carte" sont très proches de celles des distributions théoriques. Les résultats du modèle "EM carte" pour des données oculométriques seront présentés au chapitre 4.

6 Conclusion

Dans un premier temps, nous avons décrit les critères pour mesurer les correspondances entre les zones prédites par une carte de saillance et les zones fixées par des sujets. Deux critères ont été retenus : le ROC et le taux de fixations correctes. Le premier mesure la capacité d'une carte de saillance à distinguer des fixations, réalisées sur une image par des sujets, de points aléatoires. Le second représente la capacité d'une carte de saillance à prédire des fixations. Les tests d'hypothèse ont également été présentés pour tester si les valeurs des critères calculées sont significatives ou non.

Dans un deuxième temps, nous avons présenté le modèle de mélange de fonctions gaussiennes permettant de modéliser la distribution spatiale des fixations. Alors que ce modèle modélise les zones des fixations sur chaque image, il est dépendant de l'image. Nous avons également présenté le modèle "EM carte", qui est une version simplifiée de l'algorithme EM, pour estimer les contributions des facteurs de guidage aux mouvements oculaires. Ces facteurs sont des caractéristiques de bas niveau associées aux images. Le modèle "EM carte" peut être appliqué pour toutes les fixations de toutes les images et de tous les sujets. Il est indépendant de l'image.

En résumé, ce chapitre présente les outils permettant de comprendre les chapitres suivants. Ces outils permettent d'évaluer le modèle d'attention visuelle en utilisant les données oculométriques. Nous allons voir dans les chapitres suivants comment les outils décrits dans ce chapitre seront utilisés pour comparer les mouvements oculaires lors d'une exploration d'une image en couleur et lors d'une exploration d'une image en niveau de gris (cf. chapitre 4) ou pour comparer différents modèles de programmation de saccade (cf. chapitre 5).

Chapitre 4

Contributions des caractéristiques visuelles aux mouvements oculaires

1 Introduction

Dans le chapitre 1, nous avons passé en revue les propriétés des mouvements oculaires (saccades et fixations) et nous avons vu que ces mouvements sont influencés par des facteurs de bas et haut niveau. Comme nous l'avons déjà précisé auparavant, nous nous concentrons sur les caractéristiques visuelles de bas niveau de scènes naturelles en évaluant leurs rôles dans l'attention visuelle lors d'une exploration libre. Parmi ces caractéristiques visuelles figure le contraste local, notamment celui de luminance, qui est souvent considéré comme le facteur prépondérant pour l'attention visuelle. Les travaux que nous avons menés dans ce chapitre concernent des images en couleur, nous nous intéressons aussi au rôle de la couleur pour les mouvements oculaires. Dans la littérature, la question sur l'influence de la couleur ou plus concrètement, sur la contribution de la couleur par rapport au contraste de luminance sur les mouvements oculaires reste encore ouverte. Jusqu'à maintenant il y a peu d'études qui abordent cette question bien que la couleur représente un facteur saillant selon la *"Feature integration theory"* de Treisman [Treisman and Gelade, 1980] pour la recherche visuelle.

Dans ce chapitre, nous étudions les mouvements oculaires lors d'une exploration libre de scènes naturelles en répondant aux deux questions suivantes. Premièrement, les sujets humains explorent-ils de la même façon les images en couleur et en niveau de gris ? Deuxièmement, comment peut-on quantifier les contributions des caractéristiques visuelles de bas niveau aux mouvements oculaires ? Dans un premier temps, nous avons mené une expérience d'exploration libre de scènes naturelles en couleur et de mêmes scènes en niveau de gris. Nous étudions ensuite les mouvements oculaires par une méthode non-paramétrique dans laquelle nous examinons un certain nombre de statistiques calculées sur les zones fixées. L'influence de la couleur sera étudiée en comparant les mouvements oculaires des sujets en fonction du type de scènes : couleur versus niveau de gris. Par simplification de langage, nous appellerons "fixations couleur" et "fixations niveau de gris", les fixations réalisées

respectivement sur les images en couleur et sur celles en niveau de gris.

Dans un deuxième temps, les mouvements oculaires sont étudiés par une méthode paramétrique. La distribution spatiale des fixations est d'abord examinée par une modélisation de densité. A partir de cette modélisation, nous pouvons aussi comparer les mouvements oculaires effectués sur les images en couleur et en niveau de gris. Ensuite, nous abordons la quantification des contributions des caractéristiques visuelles sur les mouvements oculaires. Ces caractéristiques sont explicitement calculées par le modèle d'attention visuelle proposé. Il sera aussi intéressant de quantifier la contribution de ces caractéristiques visuelles dans la prédiction du modèle, par modèle dit "EM carte" présenté au chapitre 3.

2 Expérience d'exploration libre de scènes naturelles

2.1 Description de l'expérience

2.1.1 Stimuli

Les stimuli visuels sont composés de 34 images de scènes naturelles : 17 scènes en couleur et les 17 mêmes scènes en niveau de gris. Les images sont en format "paysage" (taille 768 × 1024 pixels) ou "portrait" (1024 × 768 pixels). Ces images proviennent de la base d'images naturelles de Kodak (`http://www.cipr.rpi.edu/resource/stills/kodak.html`).

Les images appartiennent à 5 catégories : bâtiment (2 images), paysage (6), objet (6), personne (2), animal (1). La figure 4.1 représente une image de paysage en couleur et la même image en niveau de gris. Toutes les images utilisées dans l'expérience sont présentées à l'annexe D.

(a) (b)

FIG. 4.1 – Exemple d'une scène naturelle utilisée dans l'expérience : (a) en couleur ; (b) en niveau de gris. (Extrait de la base Kodak *http ://www.cipr.rpi.edu/resource/stills/kodak.html*)

2.1.2 Sujets

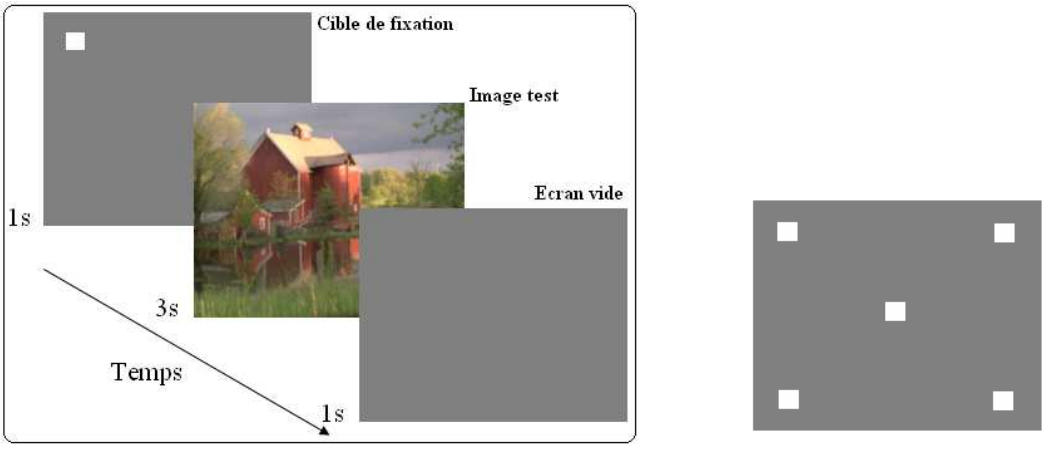
52 sujets (37 hommes et 15 femmes) ont participé à l'expérience. Les sujets sont en majorité des étudiants et aussi du personnel de l'administration de l'école d'ingénieur Polytech'Grenoble. Leur âge est compris entre 22 et 62 ans (moyenne : 29.2 ans et écart-type : 11.39 ans). Tous les sujets ont une acuité visuelle normale ou corrigée. Des consignes inscrites sur feuille papier ont été données à chaque sujet avant l'expérience et à la fin de l'expérience, chaque sujet a rempli une fiche d'information (concernant son âge, son œil directeur, le port ou non de lunettes, etc.). Les sujets sont divisés en 2 groupes : un groupe visionne les images en couleur et l'autre groupe visionne les mêmes images en niveau de gris. Ainsi, nous analysons les données de 26 sujets sur 17 images en couleurs et les données de 26 sujets sur 17 images en niveau de gris.

2.1.3 Démarche

L'expérience a eu lieu dans une pièce assombrie pour que les sujets puissent uniquement se concentrer sur les stimuli présentés à l'écran. L'expérimentateur est également présent dans la pièce. Un sujet, avec la tête maintenue par une mentonnière, est assis à une distance de 57 cm face à l'écran sur lequel sont présentées les images. Le casque de l'oculomètre Eyelink II (SR Research) est placé sur sa tête. Un stimulus est présenté sur un écran LCD d'Apple occupant $27° \times 42°$ d'angle visuel (`http://www.apple.com/fr/displays/`). Les images sont présentées au centre de l'écran en format "paysage" ou "portrait". Ainsi, une image occupe $20° \times 34°$ d'angle visuel en format "paysage" et $27° \times 26°$ en "portrait". Durant l'expérience, nous demandons aux sujets de bouger le moins possible. Une phase de calibration de l'oculomètre en 9 points est réalisée au début de l'expérience et au cours de l'expérience si le sujet bouge.

Chaque essai se déroule de la même manière : un écran avec une cible de fixation, un écran avec l'image et un écran gris moyen (Fig. 4.2a). La cible de fixation est un carré (de taille 1° x 1° d'angle visuel) pour stabiliser le regard d'un sujet. Le carré est placé aléatoirement dans une des 5 positions sur l'écran (aux 4 coins avec coordonnées $(100, 100)$, $(100, 924)$, $(1180, 100)$, $(1180, 924)$ et au centre $(640, 512)$) (Fig. 4.2b). Le sujet doit stabiliser son regard sur cette cible carrée durant 60 ms pour déclencher l'apparition de l'image à explorer librement. Si le sujet n'arrive pas à stabiliser son regard sur le carré, l'image apparaîtra aussi mais après un délai (10 s). Cependant, cet essai n'est pas gardé pour l'analyse. La position de la cible carrée varie d'un sujet à l'autre, mais elle est la même pour un sujet pour tous les essais. Une image est ensuite présentée au centre de l'écran pendant 3 s. Puis, durant l'écran gris moyen, les sujets peuvent se reposer mais ne pas bouger la tête. En outre, pour assurer l'exactitude des positions oculaires enregistrées, une calibration rapide de recadrage (*"drift"*) est effectuée après chaque essai.

Dans l'expérience, il y a deux images d'apprentissage suivies par 17 images de test. Les images d'apprentissage ont pour objectif de familiariser les sujets à l'expérience et ne seront pas utilisées pour l'analyse. L'ordre d'affichage des images pour chaque sujet est choisi aléatoirement.



(a) (b)

FIG. 4.2 – (a) Déroulement d'un essai : la flèche indique l'ordre de déroulement avec le temps de présentation de chaque écran ; (b) Les cinq positions possibles de la cible carrée de fixation.

2.1.4 Données expérimentales

Pour chaque couple "image × sujet", nous n'avons conservé les données que si le sujet avait bien stabilisé son regard sur la cible carrée avant l'apparition du stimulus (le pourcentage de couples "image × sujet" éliminés est 14% et 17% respectivement pour les images en couleur et en niveau de gris). Nous étudions, pour chaque image, les fixations et les saccades des sujets, c'est-à-dire les positions (x,y) de fixations, les durées de fixations et les amplitudes de saccades. La figure 4.3 illustre un exemple d'une scène avec les fixations des 17 sujets réalisées pendant les 3 s.

2.2 Propriétés observées sur les mouvements oculaires

A partir des données expérimentales, nous analysons des propriétés des mouvements oculaires. Nous choisissons les propriétés les plus importantes qui serviront par la suite à nos analyses. Ce sont les distributions des durées de fixations et les distributions des amplitudes de saccades obtenues ainsi que les positions de fixations sur les images en couleur et en niveau de gris. Ces propriétés ont déjà été étudiées dans des études précédentes dans des conditions d'exploration variées comme les stimuli présentés sur l'écran d'un ordinateur à l'intérieur ou lorsque les sujets se promènent à l'extérieur [Bahill et al., 1975; Andrews and Coppola, 1999; Antes, 1974; Pannasch et al., 2008; Velichkovsky et al., 2005; Tatler and Vincent, 2008]. Nous souhaitons observer des données similaires à celles de la littérature afin de valider notre protocole expérimental et nos données.

FIG. 4.3 – Exemple d'une scène avec les fixations des 17 sujets réalisées pendant 3s.

Distribution des durées de fixations

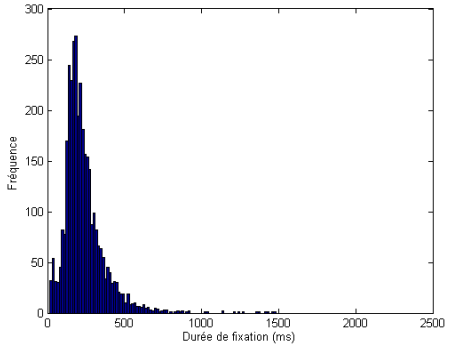
Nous avons relevé toutes les fixations de tous les sujets et puis tracé les histogrammes de leurs durées. A la figure 4.4a,b, ces histogrammes peuvent s'apparenter à des lois de Poisson. La valeur moyenne des durées de fixations se situe entre 200 et 250 ms comme observé par Andrews et Coppola [Andrews and Coppola, 1999]. Il est intéressant de noter que les histogrammes des durées de fixations couleur et les histogrammes des durées de fixations niveau de gris sont très similaires.

Distribution des amplitudes de saccades

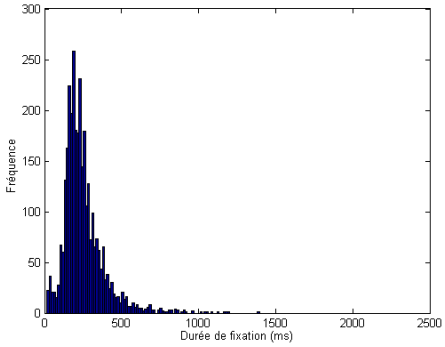
Les histogrammes des amplitudes de saccades sont également étudiées à la figure 4.4c,d de même que pour les histogrammes des durées de fixations. Les histogrammes pour les amplitudes de saccades couleur (Fig. 4.4c) ont la même forme que ceux niveau de gris (Fig. 4.4d) ; les deux histogrammes ont une forme exponentielle. De plus, la plupart des saccades ont une petite amplitude, inférieure à 15° comme dans l'étude de Bahill [Bahill et al., 1975] bien que les saccades puissent être plus grandes car la taille des images à explorer va jusqu'à 34°.

Distribution conjointe des amplitudes de saccades et des durées de fixations

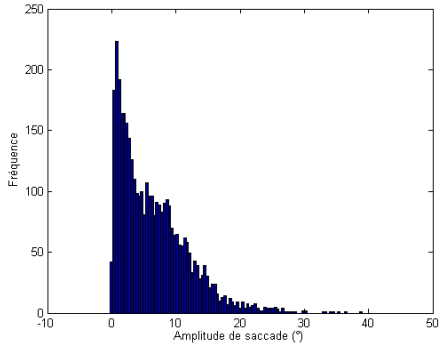
La figure 4.5a,b représente la distribution conjointe des durées de fixations et



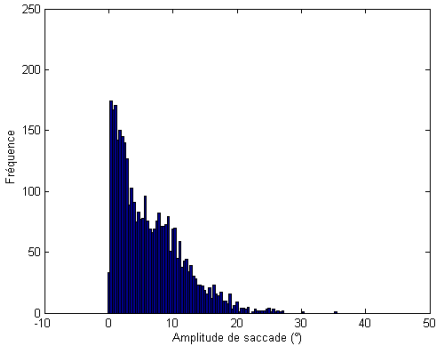
(a) Images en couleur



(b) Images en niveau de gris



(c) Images en couleur

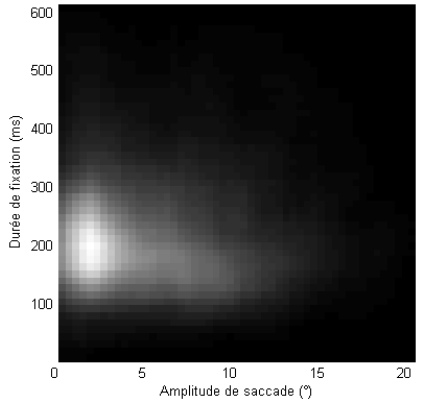


(d) Images en niveau de gris

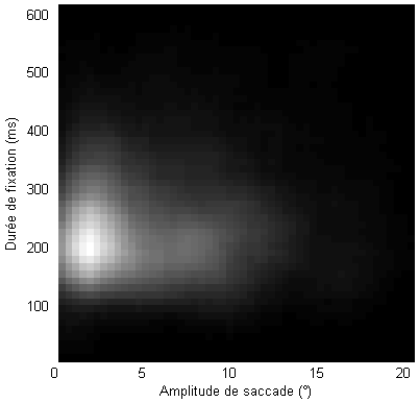
FIG. 4.4 – Distributions des durées de fixations et des amplitudes de saccades lors de l'exploration libre de 34 scènes naturelles en couleur et en niveau de gris, réalisées par deux groupes de 26 sujets pendant 3 s. (a & b) Durées de fixations ; (c & d) Amplitudes de saccades.

des amplitudes de saccades suivantes (la saccade qui suit la fixation courante) pour les images en couleur et les images en niveau de gris. La distribution, qui est similaire dans les deux jeux de données, montre que lors d'une exploration, les sujets humains effectuent majoritairement des fixations de durée moyenne (autour de 200 ms) suivies par des saccades courtes ou moyennes ($< 10°$). En revanche, il y a peu de fixations très courtes ou très longues ainsi que de saccades très longues. Ces résultats sont similaires à ceux observés par Tatler [Tatler and Vincent, 2008].

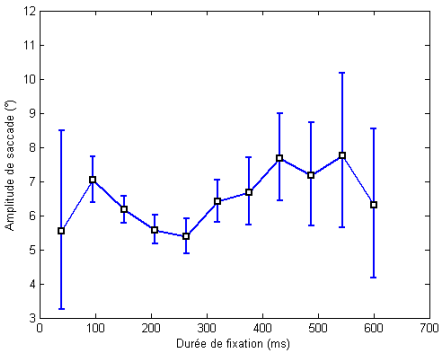
Nous avons également tracé la courbe de l'amplitude de saccade suivante en fonction de la durée de fixation courante (Fig. 4.5c,d). Les résultats obtenus à partir des images en couleur et de celles en niveau de gris se ressemblent. Bien que la variation de la courbe ne soit pas significative (taille de l'échantillon trop petite), nous pouvons observer un lien entre l'amplitude de saccade suivante et la durée de fixation courante. La saccade suivante est petite si la fixation courante est courte (< 80 ms) ou longue (> 200 ms). La saccade la plus grande est précédée par une fixation entre 80 ms et 150 ms. Nous remarquons que pour les fixations plus longues que 400 ms, les résultats ne sont plus stables car il y a peu de données. Les observations des am-



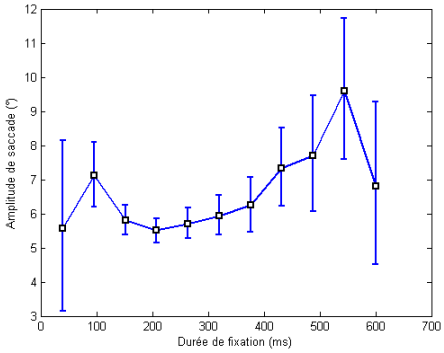
(a) Images en couleur



(b) Images en niveau de gris



(c) Images en couleur



(d) Images en niveau de gris

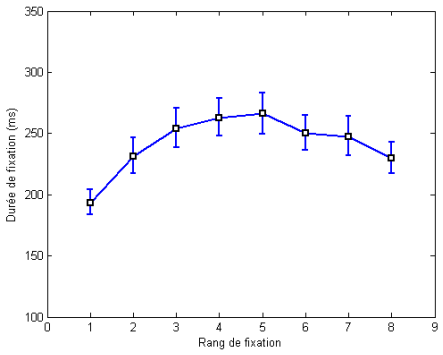
FIG. 4.5 – Relation entre les durées de fixations courantes et les amplitudes de saccades suivantes. (a & b) La distribution conjointe des durées de fixations et des amplitudes de saccades ; (c & d) L'amplitude de saccade en fonction de la durée de fixation. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap.

plitudes de saccades et des durées de fixations sur notre expérience sont conformes à celles obtenues dans [Velichkovsky et al., 2005; Tatler and Vincent, 2008].

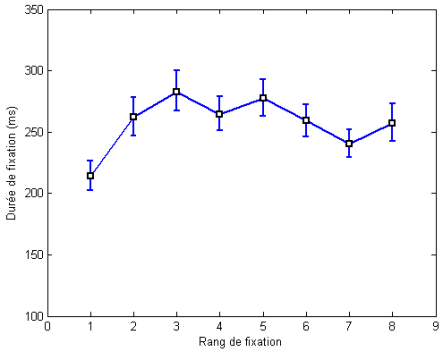
Evolution temporelle des durées de fixations et des amplitudes des saccades

Nous avons également tracé au cours du temps les valeurs moyennes des durées de fixations et des amplitudes de saccades afin d'examiner l'évolution temporelle de ces grandeurs (Fig. 4.6). Les 8 premières fixations (ou saccades) ont été choisies pour l'affichage car c'est le nombre de fixations qu'une majorité de sujets a effectué sur les images (75% des couples "image × sujet" pour les images en couleur et 76% pour les images en niveau de gris). Nous pouvons observer que ces propriétés pour les images en couleur et en niveau de gris ont encore les mêmes tendances. La première saccade après l'apparition de l'image est la plus grande saccade ; cette saccade concerne un mouvement oculaire vers le centre de l'image ; ce phénomène

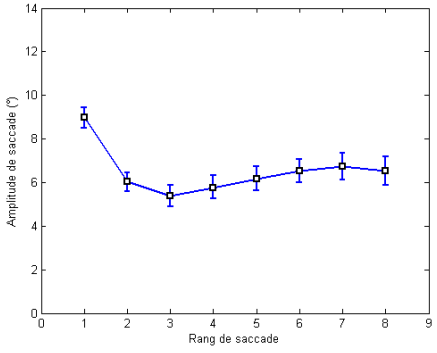
est appelé “biais de centralité” (cf. chapitre 1). Il s’explique par le fait que dans la vie quotidienne, les sujets sont familiers aux scènes avec une zone d’intérêt au centre. De plus, il représente une stratégie efficace pour explorer des scènes[1] [Tatler, 2007]. Ensuite, la deuxième saccade est plus petite et les amplitudes des saccades suivantes se stabilisent autour de 6°. Une tendance similaire mais dans le sens inverse est observée avec les durées de fixations. La première fixation est la plus courte et suivie par des fixations plus longues. Dans la littérature, une tendance similaire a été révélée : une décroissance des amplitudes de saccades et une croissance des durées de fixations au cours du temps [Antes, 1974; Pannasch et al., 2008] bien que ces observations aient été faites pour des explorations plus longues (de 7 s à 20 s au lieu de 3 s dans notre cas).



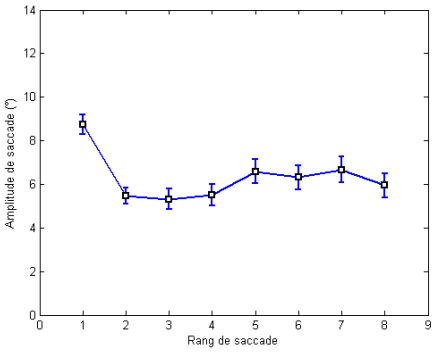
(a) Images en couleur



(b) Images en niveau de gris



(c) Images en couleur



(d) Images en niveau de gris

FIG. 4.6 – Durées moyennes de fixations et amplitudes moyennes de saccades au cours du temps pour les images en couleur et en niveau de gris. L’intervalle de confiance à 95% est calculé par la technique de Bootstrap. (a & b) Durées de fixations ; (c & d) Amplitudes de saccades.

[1]Nous verrons à la section §4.2 la quantification de la contribution de ce phénomène aux mouvements oculaires.

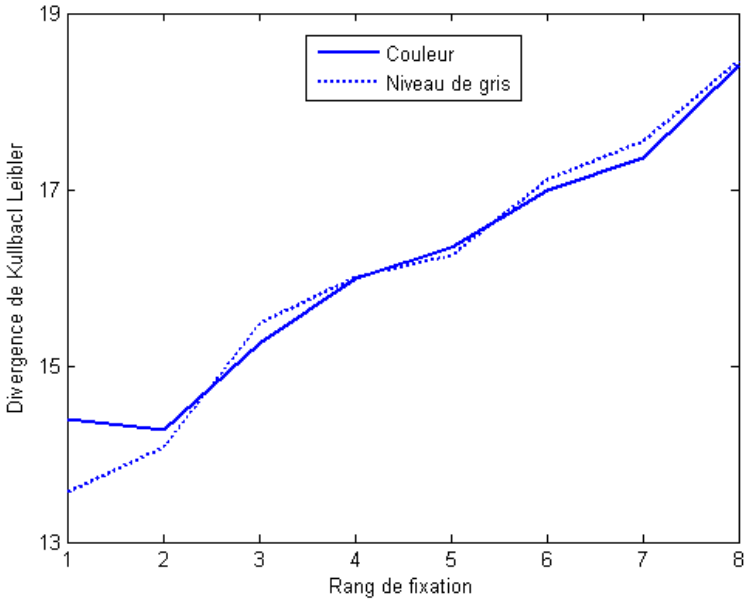


FIG. 4.7 – Cohérence inter-sujet pour les images en couleur (courbe en trait plein) et pour celles en niveau de gris (courbe en pointillé). On trace la divergence de Kullback-Leibler en fonction du rang de la fixation.

Cohérence inter-sujet

Nous avons étudié une autre propriété des mouvements oculaires : la cohérence inter-sujet qui mesure la similarité des positions des fixations entre les sujets au cours du temps. A chaque rang de fixation k, pour un sujet i, ses fixations pour toutes les images sont regroupées pour créer une carte de densité (appelée C_i^k). La même procédure est réalisée pour regrouper les fixations de tous les autres sujets pour créer une autre carte de densité (appelée $C_{i,autre}^k$). Ainsi, la cohérence inter-sujet pour le sujet i est calculée par la divergence de Kullback-Leibler [Kullback and Leibler, 1951] entre les deux cartes de densité C_i^k et $C_{i,autre}^k$. La cohérence inter-sujet à chaque rang de fixation est la moyenne des cohérences inter-sujet calculées pour tous les sujets. Sur la figure 4.7, une divergence de Kullback-Leibler faible correspond à une forte cohérence inter-sujet et inversement, une divergence importante une cohérence faible. Selon la figure, la croissance de la divergence Kullback-Leibler est observée pour les deux types de scènes. Les fixations des sujets sont proches au début de l'exploration et s'éloignent avec le temps. L'augmentation de la divergence peut être liée aux différents ordres de fixations des différents sujets. Les sujets peuvent explorer les mêmes zones mais dans les ordres différents. Ainsi, lors du calcul de la divergence à chaque rang de fixation, la divergence peut augmenter. Ce problème sera étudié à la section §4.1. Notre résultat ici est proche de celui de Tatler [Tatler et al., 2005].

En résumé, les similarités entre les résultats obtenus avec notre expérience et ceux de la littérature ainsi qu'entre les images en couleur et en niveau de gris montrent la fiabilité de nos données expérimentales et confirment les études précédentes.

3 Analyse non-paramétrique des mouvements oculaires

3.1 Méthode

Dans cette partie, les mouvements oculaires sont analysés en effectuant directement les statistiques des fixations calculées sur des caractéristiques visuelles de bas niveau. Dans la littérature, ces caractéristiques sont souvent utilisées pour discriminer les zones saillantes ou les régions d'intérêt (*"Region Of Interest"* - ROI) dans une image selon la méthode *"Bottom-Up"*. Les caractéristiques de bas niveau sont nombreuses. Dans [Privitera and Stark, 2000], les auteurs ont proposé une liste importante de caractéristiques allant de la symétrie, au contraste, à la différence *"center-surround"*, aux coefficients DCT (*"Discrete Cosine Transform"*), aux coefficients d'ondelette. Ici, sans tenir compte de toutes ces caractéristiques, nous nous concentrons sur les plus essentielles qui sont reliées à notre modèle de saillance. Ce sont les contrastes calculés sur les trois voies visuelles : luminance (L) et chrominance Rouge-Vert (RG), Bleu-Jaune (BY)[2]. Ces choix sont souvent utilisés dans la littérature [Reinagel and Zador, 1999; Tatler et al., 2005; Frey et al., 2008].

Le contraste des trois cartes (L, RG et BY) est calculé de la manière suivante. Nous sous-échantillonnons chaque carte par un facteur de 4 et nous réalisons la convolution de la carte résultante avec un filtre DoG (*"Difference Of Gaussians"*) pour obtenir une carte de contraste. Cette dernière est ensuite ramenée à la taille initiale (la taille de la scène). L'échantillonnage régulier combiné avec le filtrage DoG a pour but de calculer le contraste en supprimant les hautes fréquences pour éviter le bruit. Enfin, nous gardons la valeur absolue de chaque carte de contraste après avoir enlevé sa valeur moyenne [Tatler et al., 2005]. Ainsi, chaque image en couleur est décomposée en trois cartes de contraste : une carte de contraste de luminance (M^L), une carte de contraste RG (M^{RG}) et une carte de contraste BY (M^{BY}) (Fig. 4.8). A l'aide de ces caractéristiques visuelles, les fixations nous permettent d'examiner les statistiques des régions fixées avec deux objectifs :

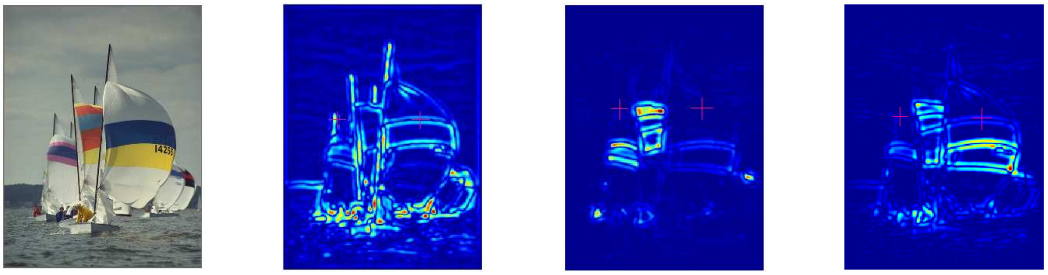
- Nous étudions si les statistiques des régions fixées diffèrent des statistiques des autres régions de l'image.
- Nous comparons les fixations effectuées sur les scènes en couleur et en niveau de gris afin d'évaluer le rôle de la couleur dans l'attention visuelle.

3.2 Statistiques des régions fixées

Après une étude globale des propriétés des mouvements oculaires (cf. section §2.2), nous voulons également vérifier quantitativement que les fixations se font sur des régions particulières de l'image et non pas sur des régions aléatoires.

La méthode souvent utilisée pour représenter la relation entre les caractéristiques visuelles et les mouvements oculaires est de comparer les statistiques des fixations avec celles obtenues sur des points aléatoires. Bien que ces derniers puissent être

[2]Ces voies sont extraites comme dans la section §2.2.2.



(a) Image originale (b) Carte de contraste M^L (c) Carte de contraste M^{RG} (d) Carte de contraste M^{BY}

FIG. 4.8 – Exemple des 3 cartes de contraste. (a) Image originale en couleur ; (b) Carte de contraste de luminance ; (c) Carte de contraste d'opposition RG ; (d) Carte de contraste d'opposition BY. Les fixations (croix) sont projetées sur ces cartes pour calculer les statistiques des différentes cartes au niveau de ces fixations.

choisis complètement aléatoirement dans une image, ce choix est critiqué parce qu'il ne prend pas en compte les paramètres classiques des mouvements oculaires. La méthode proposée pour le choix des points aléatoires est de prendre les fixations d'un sujet mais obtenues sur d'autres images [Reinagel and Zador, 1999; Tatler et al., 2005; Frey et al., 2008]. Ce choix-là permet de conserver des amplitudes de saccades "physiologiques".

3.2.1 Fixations couleur

Pour chaque image en couleur, les 3 cartes de caractéristique (contraste de luminance, contraste RG et BY) sont extraites. Ensuite, pour chaque couple "image × sujet", toutes les fixations sont collectées. Parallèlement, les fixations aléatoires pour ce couple sont générées. La valeur $C^l_{i,j}$ pour un couple "image i × sujet j" et pour la carte de contraste M^l, $l \in \{L, RG, BY\}$, est la moyenne des contrastes de toutes les fixations[3] :

$$C^l_{ij} = \frac{1}{N} \sum_{k=1}^{N} M^l(\mathbf{x}_k) \tag{4.1}$$

$M^l(\mathbf{x}_k)$: contraste à la position $\mathbf{x}_k$ extraite de la carte de contraste M^l
N : le nombre de fixations.

Le contraste moyen pour tous les sujets et toutes les images est représenté à la figure 4.9 selon le type de contraste. D'après cette figure, il existe une différence significative entre le contraste moyen obtenu pour les fixations et le contraste moyen obtenu pour des points aléatoires. Cela est confirmé par un test Kolmogorov - Smirnov (KS-test) (cf. chapitre 3) ($p \approx 0$ pour les trois contrastes).

[3]Le contraste d'une fixation est la valeur moyenne dans un carré de 1° x 1° d'angle visuel (correspondant à 30 × 30 pixels) autour de la fixation. La taille de ce carré a pour but de tenir compte de l'incertitude des équipements. De plus, cette taille correspond également à la taille du centre de la fovéa.

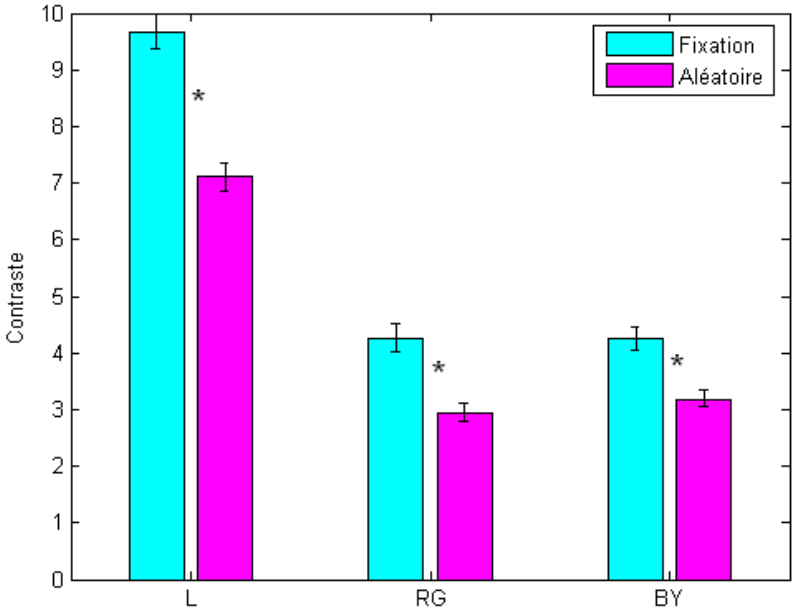


FIG. 4.9 – Comparaison entre les contrastes moyens calculés sur les fixations et sur les points aléatoires pour les trois cartes de contraste : contraste de luminance, contraste RG et contraste BY. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap. Le signe "étoile" (*) signifie une différence significative entre les deux valeurs de contrastes.

Nous pouvons donc conclure que les fixations réalisées par les sujets lors de l'exploration libre des scènes se situent sur des régions à fort contraste de luminance et de chrominance RG, BY. Ces résultats confirment la conclusion que les sujets fixent des régions qui ne sont pas aléatoires et ont tendance à regarder des zones à fort contraste de luminance mais également d'opposition de couleurs.

3.2.2 Fixations niveau de gris

Nous testons les fixations obtenues sur les images en niveau de gris. Nous pouvons noter que pour ces images une seule carte de contraste est calculée : le contraste de luminance.

Nous obtenons le même résultat que précédemment dans le cas des scènes en couleur. Les statistiques des fixations sont bien différentes de celles des points aléatoires (KS-test, $p \approx 0$).

En conclusion, les jeux de données des fixations couleur et niveau de gris ne sont pas dues au hasard. Nous avons montré qu'en ces points, les contrastes de luminance et de couleur sont supérieurs à ceux aux points aléatoires. Les mouvements oculaires peuvent donc être, en partie du moins, expliqués par des mesures de contraste.

3.3 Influence de la couleur

Nous comparons ici les mouvements oculaires enregistrés lors de l'exploration d'une image en couleur avec ceux enregistrés lors de l'exploration d'une image en niveau de gris. Comme ci-dessus, à partir d'une image en couleur sont extraites trois cartes de contraste. Ensuite, les fixations couleur obtenues pour les 17 sujets sont projetées sur ces cartes. Les fixations niveau de gris y sont également projetées (Fig. 4.10).

Il est à noter que les fixations niveau de gris ont été obtenues sur des images en niveau de gris et donc les contrastes de couleur n'existent pas pour ces images. Pourtant, comme nous souhaitons évaluer le rôle de la couleur, nous projetons aussi les fixations niveau de gris sur les cartes de contraste de couleur. Si la couleur apporte des informations supplémentaires par rapport à la luminance seule, les statistiques des fixations couleur sur les cartes de contraste de couleur devraient être différentes des statistiques des fixations niveau de gris sur ces mêmes cartes.

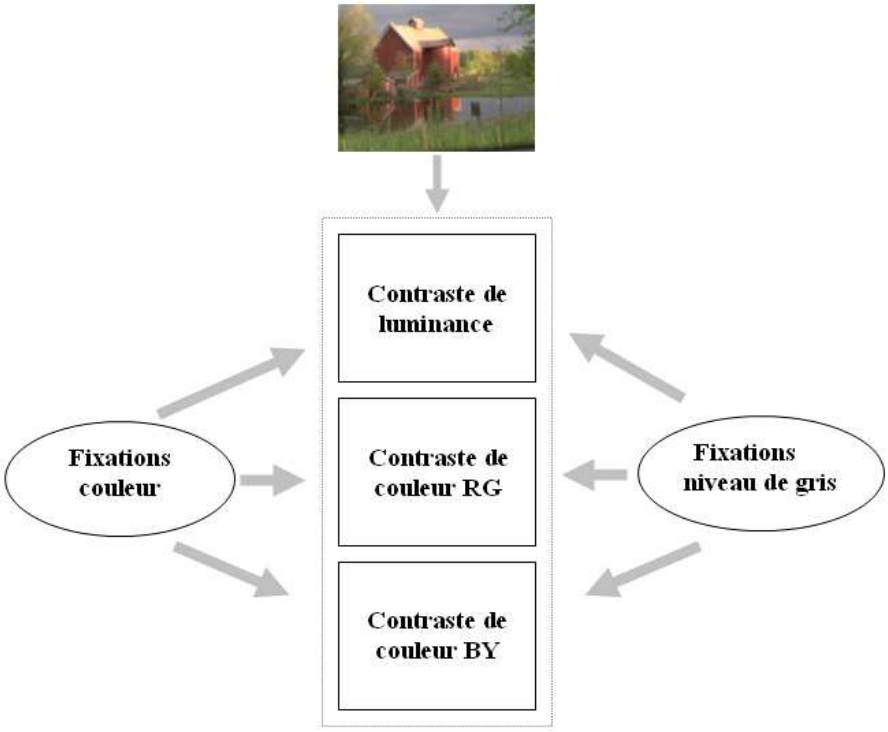


FIG. 4.10 – Les fixations couleur et niveau de gris sont "projetées" sur les cartes de contraste pour calculer les statistiques de ces fixations sur ces trois types de contraste.

3.3.1 Evaluation pour l'ensemble des images

Dans un premier temps, nous comparons les fixations couleur et niveau de gris pour l'ensemble des images en utilisant le critère ROC (cf. chapitre 3). Nous partons de l'hypothèse que si la couleur apporte une contribution aux mouvements oculaires,

il existera une différence significative entre l'aire au-dessous de la courbe ROC (AUC) pour les fixations couleur et l'AUC pour les fixations niveau de gris. Pour chaque type de fixations, nous calculons l'AUC pour chaque couple "image × sujet" et puis nous prenons la moyenne des valeurs obtenues sur tous les couples. La figure 4.11 présente les valeurs moyennes d'AUC pour les fixations couleur et niveau de gris et pour les trois cartes de contraste.

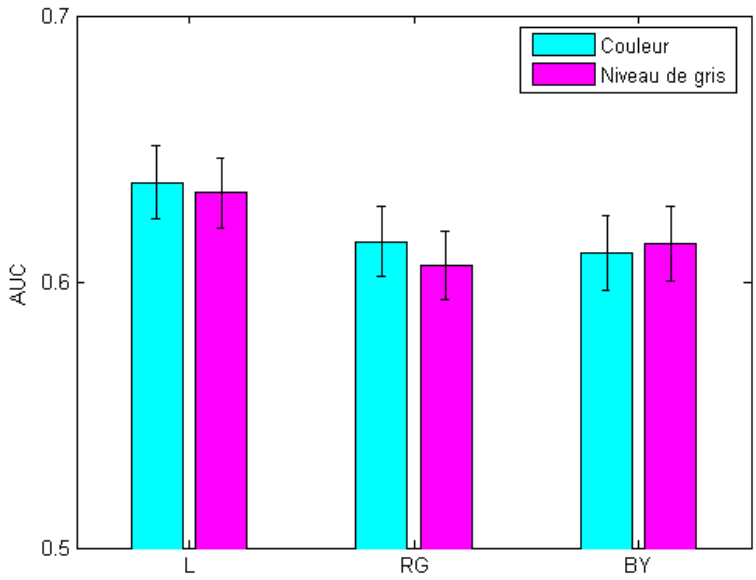


FIG. 4.11 – La moyenne de l'aire sous la courbe ROC (AUC) pour les fixations couleur et niveau de gris et pour les trois caractéristiques : contrastes de luminance et de deux voies chromatiques. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap.

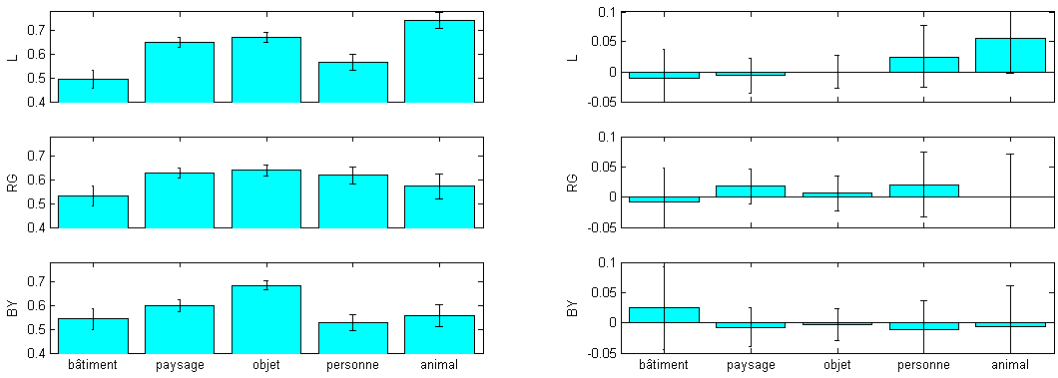
Dans la figure 4.11, nous voyons que pour les trois cartes de contraste (L, RG et BY) la valeur d'AUC pour les fixations couleur et celle pour les fixations niveau de gris ne diffèrent pas (KS-test, $p = 0.5026$, $p = 0.6543$ et $p = 0.1964$ respectivement pour le contraste de luminance, RG et BY).

Alors, pour l'ensemble des images, l'ajout de l'information chromatique dans des scènes naturelles ne modifie pas les mouvements oculaires par rapport au cas où seule la luminance est présente.

Ici, nous avons comparé les fixations couleur et les fixations niveau de gris pour les images de toutes les catégories confondues. Le résultat est qu'il n'y a pas de différence significative entre les deux jeux de fixations. Dans la suite, nous divisons la base d'images en catégories pour tester si la différence entre ces deux jeux de fixations dépend du type d'images.

3.3.2 Evaluation selon la catégorie sémantique

Dans [Frey et al., 2008], les auteurs ont comparé les mouvements oculaires sur des images en couleur avec les mouvements oculaires sur les mêmes images en niveau de gris. Cette comparaison a été faite pour plusieurs caractéristiques comme le contraste de luminance, la texture de luminance, le contraste de deux voies chromatiques Rouge-Vert et Bleu-Jaune, et la saturation dans l'espace couleur DKL [Derrington et al., 1984]. Les stimuli visuels comportent 191 images réparties dans 7 catégories : visage (26 images), fleur et animal (30), forêt (30), fractal (25), paysage (19), *"man-made"* (32) et *"rainforest"* (29). Les résultats ont montré que la différence entre les fixations couleur et les fixations niveau de gris est faible et dépendante de la catégorie sémantique de l'image. Parmi toutes les catégories et toutes les caractéristiques visuelles, il existe une différence pour la catégorie *"rainforest"* et pour le contraste Rouge-Vert. Ici, en reprenant cette approche, nous comparons les deux types de fixations en fonction de la catégorie de l'image. Notons cependant une différence entre l'expérience dans [Frey et al., 2008] et la nôtre. Dans [Frey et al., 2008], les mêmes sujets ont regardé les images en couleur et les mêmes images en niveau de gris. Pour chaque sujet, les deux expériences pour deux types de scènes sont espacées de 24 jours en moyenne. Dans notre cas, nous avons deux groupes de sujets : l'un pour les images en couleur, l'autre pour celles en niveau de gris (cf. section §2.1). Nous avons voulu aborder cette étude également sous l'angle des catégories, en restant néanmoins très prudents car dans notre base d'images, le nombre d'images par catégorie est très faible. En effet, notre base d'images peut se catégoriser en 5 groupes : bâtiment/building (2 images), paysage (6), objet (6), personne (2), animal (1).



(a) (b)

Fig. 4.12 – La moyenne de l'aire sous la courbe ROC (AUC) pour les trois types de contraste en fonction des catégories sémantiques : (a) pour les fixations couleur ; (b) différence de l'AUC pour les fixations couleur et niveau de gris. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap.

Le calcul d'AUC est fait comme précédemment mais moyenné sur chaque catégorie. La figure 4.12a représente l'AUC moyenne pour les fixations couleur selon différentes catégories et types de contraste. La différence de cette grandeur entre les fixations

couleur et niveau de gris est représentée dans la figure 4.12b. Il n'existe pas de différence significative entre les deux jeux de fixations (tous les tests de KS donnent $p > 0.05$). Ainsi, au vu de cette analyse et considérant ses limites par le faible nombre d'images par catégorie, nous n'avons pas mis en évidence de différence entre les mouvements oculaires lors de l'exploration libre de scènes en couleur et ceux obtenus lors de l'exploration libre de scènes en niveau de gris.

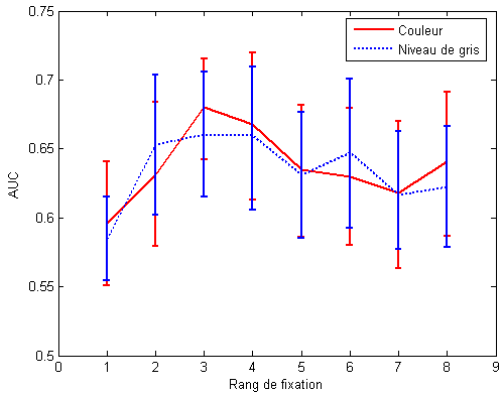
3.3.3 Evolution temporelle des statistiques des fixations

Dans cette partie, nous allons comparer les mouvements oculaires obtenus pour des scènes en couleur et ceux pour des scènes en niveau de gris en tenant compte de l'évolution temporelle des statistiques des fixations. En général, on reconnaît que la voie ascendante intervient tôt dans la perception visuelle et puis la voie descendante intervient. Alors que le rôle de la voie descendante augmente avec le temps, la contribution de la voie ascendante reste encore controversée. La première hypothèse est que la voie ascendante est dominante au début du processus de traitement de l'information visuelle et ensuite diminue au cours du temps [Parkhurst et al., 2002]. Dans la deuxième hypothèse, la contribution de la voie ascendante reste la même malgré l'augmentation de la voie descendante. Alors, le rôle de la voie ascendante diminue au cours du temps du fait de la dominance de la voie descendante, et non pas par sa décroissance elle-même [Tatler et al., 2005]. De plus, dans [Tatler et al., 2005], les auteurs ont montré que la différence des deux hypothèses concernant la voie ascendante est liée à la manière d'extraire la saillance des fixations. D'après Tatler, la saillance des fixations (selon le critère ROC) devrait être mesurée en comparaison avec la saillance des points aléatoires qui sont choisis comme les fixations du même sujet mais dans d'autres images pour prendre en compte le phénomène de biais de centralité des mouvements oculaires. En revanche, dans [Parkhurst et al., 2002] le calcul de la saillance des fixations ne concerne que des fixations. Cela pourrait, selon Tatler, créer des artefacts et faire diminuer la saillance des fixations au cours du temps.

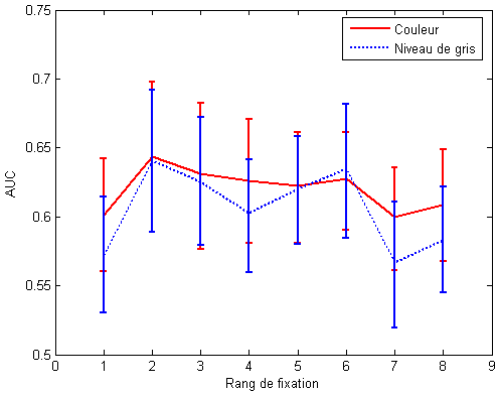
Ici, nous allons donc examiner l'évolution temporelle des statistiques des fixations au niveau des caractéristiques de bas niveau (constituant la voie ascendante) afin de pouvoir infirmer ou confirmer les deux hypothèses décrites ci-dessus. En même temps, nous testons si au cours du temps les sujets ont regardé de la même façon les images en couleur et celles en niveau de gris. Le raisonnement est que si la couleur apporte des contributions, la courbe temporelle pour les fixations couleur et celle pour les fixations niveau de gris divergeront avec le temps.

Dans la figure 4.13, la valeur d'AUC moyenne est tracée en fonction de l'ordre de la fixation pour les fixations couleur et niveau de gris et toujours pour les trois mêmes caractéristiques. Les courbes sont tracées pour les 8 premières fixations ; ce nombre de fixations est celui atteint par 75% des couples "image × sujet" pour les images en couleur et 76% pour les images en niveau de gris.

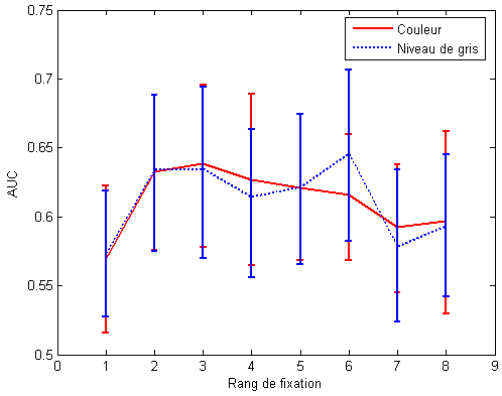
En regardant la figure 4.13, nous observons que l'influence des caractéristiques de bas niveau semblent se maintenir lors de l'exploration. Ce résultat confirme la conclusion de Tatler [Tatler et al., 2005] selon laquelle la contribution de la voie



(a) Contraste de luminance



(b) Contraste RG



(c) Contraste BY

FIG. 4.13 – Evolution temporelle de la moyenne de l'aire sous la courbe ROC (AUC) pour les fixations couleur et niveau de gris pour les trois caractéristiques : (a) Contraste de luminance ; (b) Contraste de couleur Rouge-Vert ; (c) Contraste de couleur Bleu-Jaune. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap.

ascendante serait quasi constante pour l'attention visuelle.

De plus, nous pouvons constater que les évolutions temporelles des valeurs AUC moyennes des fixations couleur et niveau de gris sont proches. La différence existe pour certaines fixations mais n'est pas significative. Nous pouvons conclure, en combinant les résultats ci-dessus, que pendant l'exploration libre de scènes naturelles, la luminance apporte une information suffisante pour les mouvements oculaires.

Dans la suite, les deux jeux de fixations sont encore comparés mais par une méthode paramétrique et puis, les contributions relatives des caractéristiques visuelles sont quantitativement évaluées.

4 Analyse paramétrique des mouvements oculaires

4.1 Analyse selon un modèle de densité spatiale par image

Maintenant, nous n'utilisons plus les statistiques calculées sur les fixations pour la comparaison des fixations couleur et niveau de gris. En revanche, nous essayons de comparer directement ces deux jeux de fixations en utilisant leurs positions et donc uniquement les données de l'expérience. Selon cette méthode, la distribution spatiale d'un ensemble des fixations est modélisée par un mélange de fonctions gaussiennes (cf. chapitre 3). Ainsi, comparer les fixations couleur et niveau de gris revient à comparer les distributions spatiales pour ces deux types de fixations.

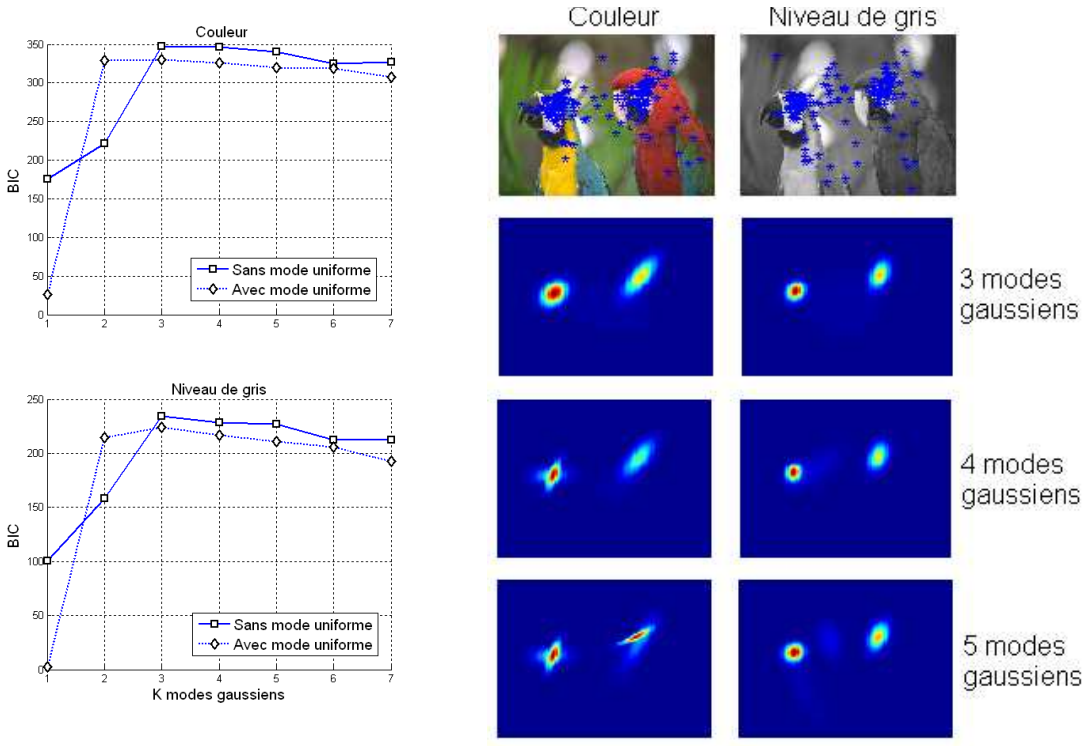
4.1.1 Evaluation qualitative

La modélisation des fixations par un mélange de fonctions gaussiennes part de l'hypothèse que les fixations sont engendrées par des sources qui sont liées aux zones saillantes de la scène. Ainsi, le modèle de mélange de fonctions gaussiennes dépend de l'image. Un modèle est donc calculé pour chaque image. Dans notre cas, pour une image, toutes les fixations de tous les sujets qui l'ont visionnée sont regroupées. Ces fixations constituent le jeu de données pour l'entrée d'un modèle de mélange de fonctions gaussiennes. Selon le critère BIC (*"Bayesian Information Criterion"*, cf. chapitre 3), nous pouvons choisir le meilleur modèle pour chaque jeu de données. En réalité, les critères BIC des meilleurs modèles peuvent être proches. Ainsi, nous décidons d'afficher les trois meilleurs modèles de mélange de fonctions gaussiennes pour chaque jeu de données. Ces trois modèles sont choisis à partir de 7 modèles sans mode uniforme et 7 modèles avec mode uniforme. Les 7 modèles correspondent à des modèles comportant de 1 à 7 modes gaussiens. Nous remarquons également que pour les fixations obtenues, le modèle sans mode uniforme est meilleur. Si les trois meilleurs modèles sont pris pour chaque image, le modèle sans mode uniforme est sélectionné dans 94% pour les images en couleur et 90% pour les images en niveau de gris. Cela signifie que les fixations sont liées majoritairement au contenu de l'image et peu au bruit.

En examinant visuellement les cartes de densité des fixations, qui proviennent des mélanges de fonctions gaussiennes, pour chaque scène en couleur et en niveau de gris, nous trouvons que les deux distributions spatiales semblent similaires (Fig. 4.14). Un critère quantitative est nécessaire pour confirmer cette observation.

4.1.2 Evaluation quantitative

Le critère naturel pour cette comparaison est la vraisemblance moyenne (Eq. 4.2) obtenue à la convergence de l'algorithme "EM" (*"Expectation-Maximization"*) pour le meilleur modèle sélectionné par le critère d'information BIC. Pour chaque couple d'images (couleur et niveau de gris), le meilleur modèle de mélange de fonctions gaussiennes est sélectionné à partir des fixations couleur. Ensuite, les fixations niveau de gris et les fixations couleur sont progressivement projetées sur ce modèle pour calculer la log-vraisemblance moyenne. Le raisonnement est que si la couleur apporte une contribution différente de la luminance, la log-vraisemblance moyenne



(a) (b)

FIG. 4.14 – Modélisation des fixations par un mélange de fonctions gaussiennes. (a) Critère BIC pour l'image en couleur et pour l'image en niveau de gris. (b) Les meilleurs modèles selon le critère BIC. Première ligne : images en couleur et en niveau de gris superposées de fixations récupérées depuis l'expérience. De ligne 2 à la ligne 4 : les trois meilleurs modèles du mélange de fonctions gaussiennes pour l'image en couleur à gauche et pour l'image en niveau de gris à droite.

pour les fixations couleur devrait être plus grande que celle pour les fixations niveau de gris.

$$L_m = \frac{1}{N} \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} p_k G(\mathbf{x}_k | \mu_k, \Sigma_k) \right) \tag{4.2}$$

avec N nombre de fixations et K nombre de fonctions gaussiennes du modèle.

La figure 4.15 représente l'évolution temporelle de la log-vraisemblance moyenne L_m pour les fixations couleur et niveau de gris. Cette valeur est moyennée sur 17 images. De plus, nous calculons également les log-vraisemblances moyennes pour un jeu de données généré par une distribution uniforme et un autre généré par la distribution théorique : la somme de fonctions gaussiennes dont les paramètres sont ceux du meilleur modèle. Toutes les log-vraisemblances moyennes sont affichées sur la même figure (Fig. 4.15).

A partir de cette figure, la log-vraisemblance moyenne pour les fixations couleur et niveau de gris sont proches et elles sont tous les deux proches de celle pour la dis-

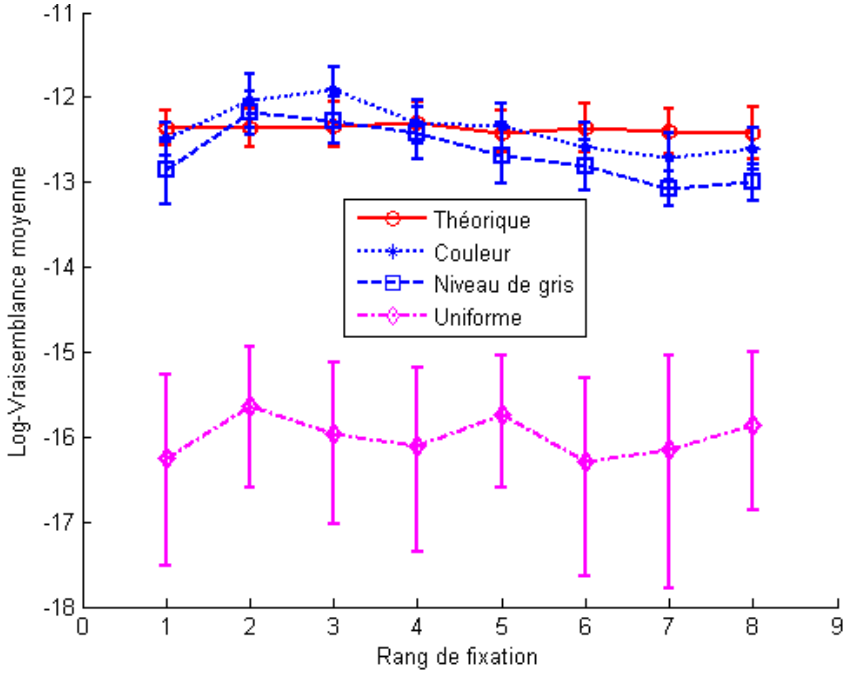


FIG. 4.15 – Log-vraisemblances moyennes pour les fixations couleur, les fixations niveau de gris, les données d'une distribution uniforme et les données d'une distribution théorique en fonction du rang de la fixation. Les log-vraisemblances moyennes sont calculées en se basant sur le meilleur modèle de mélange de fonctions gaussiennes pour les fixations couleur. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap.

tribution théorique. De plus, elles sont nettement supérieures à la log-vraisemblance moyenne des données uniformes. Ce résultat montre une similarité des mouvements oculaires obtenus pour des scènes en couleur et pour des scènes en niveau de gris. De plus, la log-vraisemblance moyenne pour les fixations couleur ou niveau de gris a tendance à diminuer selon l'ordre de fixations tandis que celle pour la distribution uniforme ou théorique reste constante. Cela s'explique par l'influence de la voie descendante. Selon le temps, les fixations sont de plus en plus dispersées et donc il est plus difficile pour le modèle de mélange de fonctions gaussiennes de modéliser tous les points.

En résumé, les résultats ci-dessus montrent que les mouvements oculaires effectués sur des images en couleur ne diffèrent pas de ceux effectués sur les mêmes images en niveau de gris lors de l'exploration libre de scènes naturelles. Cette conclusion est obtenue à partir des comparaisons entre des fixations couleur et niveau de gris par une méthode non-paramétrique, et puis par une méthode paramétrique. Alors, l'ajout de la couleur dans une scène naturelle ne modifie pas les mouvements oculaires par rapport à la luminance seule.

4.2 Quantification des contributions des caractéristiques visuelles

4.2.1 Méthode

L'évaluation du rôle de la couleur et de la luminance ci-dessus est qualitative. Nous savons que la couleur apporte peu à l'attention visuelle, mais nous ne connaissons pas encore quantitativement cette contribution. Cette question est étudiée dans cette section. Nous allons ici à nouveau utiliser le modèle d'attention visuelle proposé au chapitre 2 pour calculer les caractéristiques visuelles. En évaluant les contributions des caractéristiques aux mouvements oculaires, nous pouvons également étudier quels sont les facteurs intéressants à prendre en compte dans notre modèle d'attention visuelle.

Notre modèle d'attention visuelle décompose d'abord une image couleur en différentes cartes d'énergie. Ensuite, les cartes d'énergie sont regroupées suivant les 6 caractéristiques : luminance et deux voies chromatiques basse fréquence (BF), luminance et deux voies chromatiques passe-bande (PB) comme dans la figure 4.16. Ces cartes proviennent du schéma représentant la décomposition de la voie de luminance et des voies chromatiques au chapitre 2 (Fig. 2.14). Une carte basse fréquence est la sortie d'un filtre passe-bas tandis qu'une carte passe-bande est la somme des sorties des filtres LogNormaux à différentes orientations et différentes fréquences après avoir été normalisées et soumises aux interactions. Un exemple des 6 cartes de caractéristique est illustré à la figure 4.17. Il y a maintenant 6 cartes de caractéristique à fusionner pour créer la carte de saillance. Mais, quelles sont les pondérations de ces cartes dans la fusion ? Ici, nous répondons à cette question en utilisant le modèle "EM carte" (cf. chapitre 3) pour évaluer les contributions relatives de chacune des cartes de caractéristique à la carte de saillance, et autrement dit, à la prédiction des fixations.

Selon le modèle "EM carte", les 6 caractéristiques de bas niveau permettent d'expliquer les mouvements oculaires (i.e. les fixations dans ce cas) ; ces caractéristiques sont extraites pour toutes les images. Ainsi, les fixations de toutes les images sont regroupées pour le modèle "EM carte". Nous rappelons que les caractéristiques utilisées dans notre modèle d'attention visuelle sont liées aux images. Néanmoins, les mouvements oculaires peuvent aussi être expliqués par d'autres caractéristiques non-liées aux images. Ainsi, pour que le modèle "EM carte" soit plus complet, nous ajoutons une carte de biais de centralité[4] pour modéliser le fait que les sujets regardent souvent la zone centrale d'une image lors d'une exploration indépendamment du contenu de l'image. De plus, nous tenons compte d'une carte de distribution uniforme qui représente toutes les autres caractéristiques susceptibles d'expliquer les mouvements oculaires potentiels. Ainsi, il y a 8 cartes (6 cartes pour les caractéristiques visuelles, 1 carte pour le biais de centralité et 1 carte pour le bruit éventuel ou d'autres caractéristiques) qui peuvent jouer un rôle dans le guidage des mouvements oculaires pour le modèle "EM carte".

[4]La carte de biais de centralité est modélisée par une distribution gaussienne au centre de la scène avec un écart-type égal au huitième la taille de la scène dans deux dimensions.

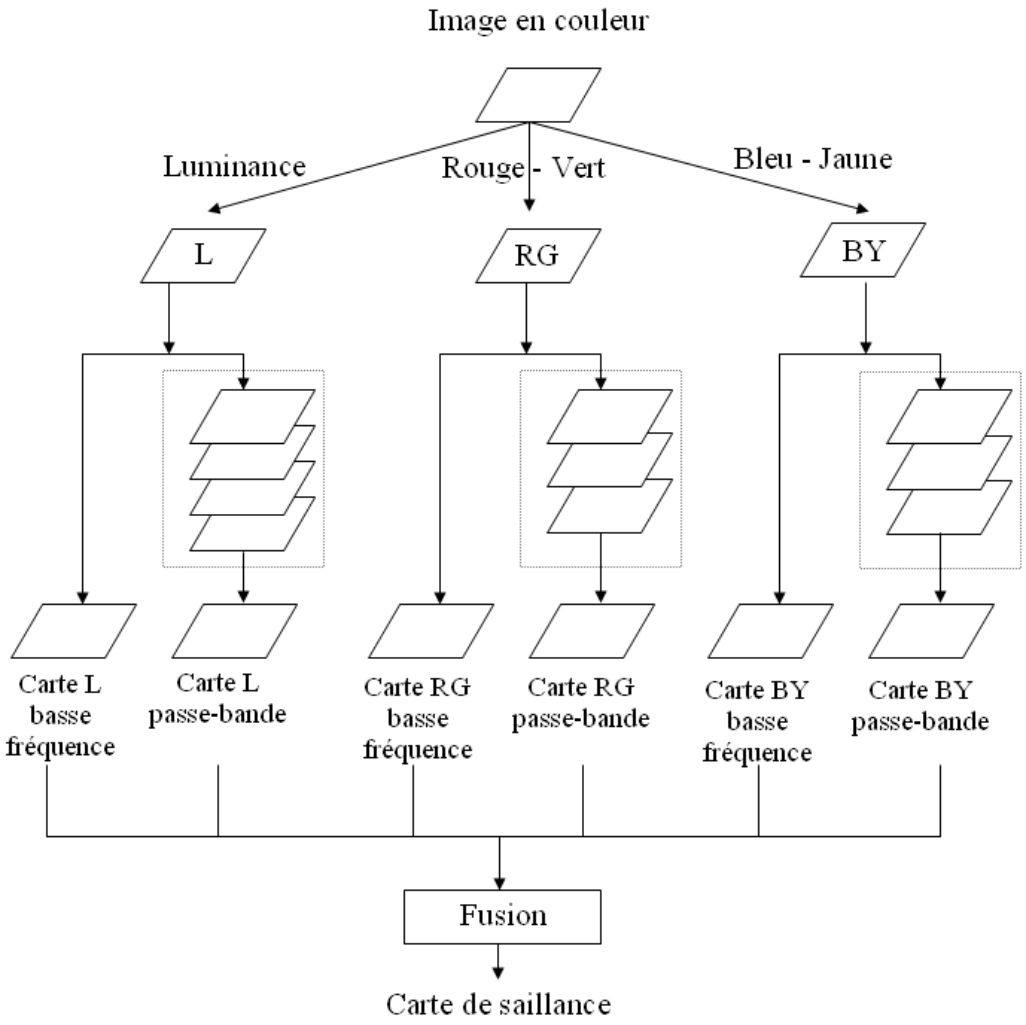


FIG. 4.16 – Les 6 cartes de caractéristique de bas niveau extraites d'une image en couleur par notre modèle d'attention visuelle. Ces caractéristiques constituent les facteurs de guidage utilisés dans le modèle "EM carte".

4.2.2 Evaluation globale des contributions

L'évaluation de cette section concerne l'ensemble des fixations couleur, sans tenir compte de leur ordre. Afin d'évaluer les contributions des caractéristiques que nous avons présentées ci-dessus, nous regroupons les fixations pour toutes les images en couleur. Pour chaque couple "image × sujet", seules les 8 premières fixations sont prises en compte. La probabilité d'apparition d'une fixation dépend maintenant des caractéristiques extraites pour toutes les images. A l'issue de l'algorithme "EM carte", nous obtenons les paramètres Θ, qui représentent les contributions de chacune des caractéristiques obtenues à la convergence de l'algorithme.

A la figure 4.18 sont illustrées les 8 pondérations des caractéristiques. Ces pondérations représentent la contribution que chaque caractéristique apporte à l'attention visuelle lors de l'exploration de scènes naturelles. Les résultats montrent une contribution importante de la luminance passe-bande. La contribution des caractéristiques

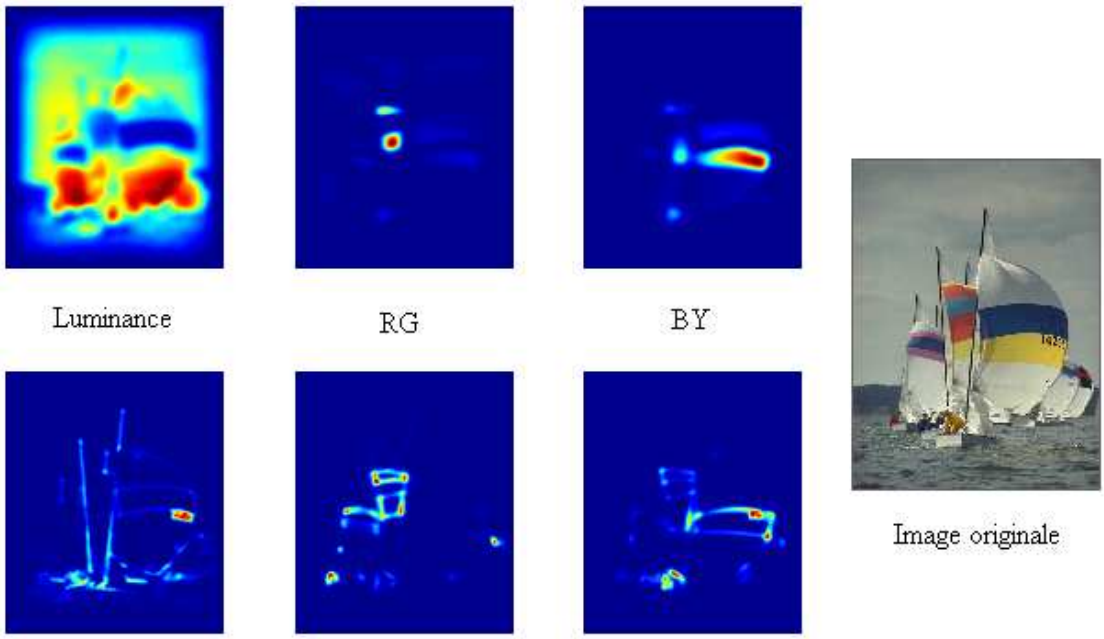


FIG. 4.17 – Exemple des 6 cartes de caractéristique de bas niveau d'une scène naturelle (à droite). Ces cartes sont calculées par le modèle d'attention visuelle décrit au chapitre 2. En haut : les cartes basse fréquence, en bas : les cartes passe-bande, de gauche à droite : les cartes de luminance, Rouge-Vert et Bleu-Jaune.

chromatiques que soit la partie basse fréquence ou la partie passe-bande est faible par rapport à celle de la luminance passe-bande. Ce résultat confirme la conclusion précédente que la couleur apporte peu à l'attention visuelle. Une autre propriété concerne la luminance basse fréquence qui n'est pas une caractéristique saillante comme cela a déjà été trouvé dans des études antérieures.

Une propriété intéressante qui peut être obtenue à partir de ce résultat concerne le biais de centralité. Plusieurs études ont observé ce phénomène quelque soit la tâche de l'expérience. Le modèle "EM carte" a permis de confirmer ce biais en quantifiant sa contribution aux mouvements oculaires. En effet, l'influence de cette caractéristique est importante, juste derrière la luminance passe-bande et bien au dessus des autres. Ainsi il est nécessaire de prendre en compte ce biais dans la modélisation de mouvements oculaires. Le biais de centralité peut s'expliquer de plusieurs manières. Premièrement, le centre d'une image contient souvent des zones saillantes qui peuvent attirer les yeux humains. Deuxièmement, ce biais reflète l'habitude des sujets humains qui déplacent les yeux vers le centre d'une image au début d'une exploration. Enfin, il constitue une stratégie pour explorer efficacement une scène et c'est également, la position de repos des yeux.

De plus, la distribution uniforme semble plus importante que les caractéristiques de couleur ou de luminance basse fréquence. Elle représente tous les autres facteurs qui peuvent influencer les mouvements oculaires mais dont nous ne connaissons pas

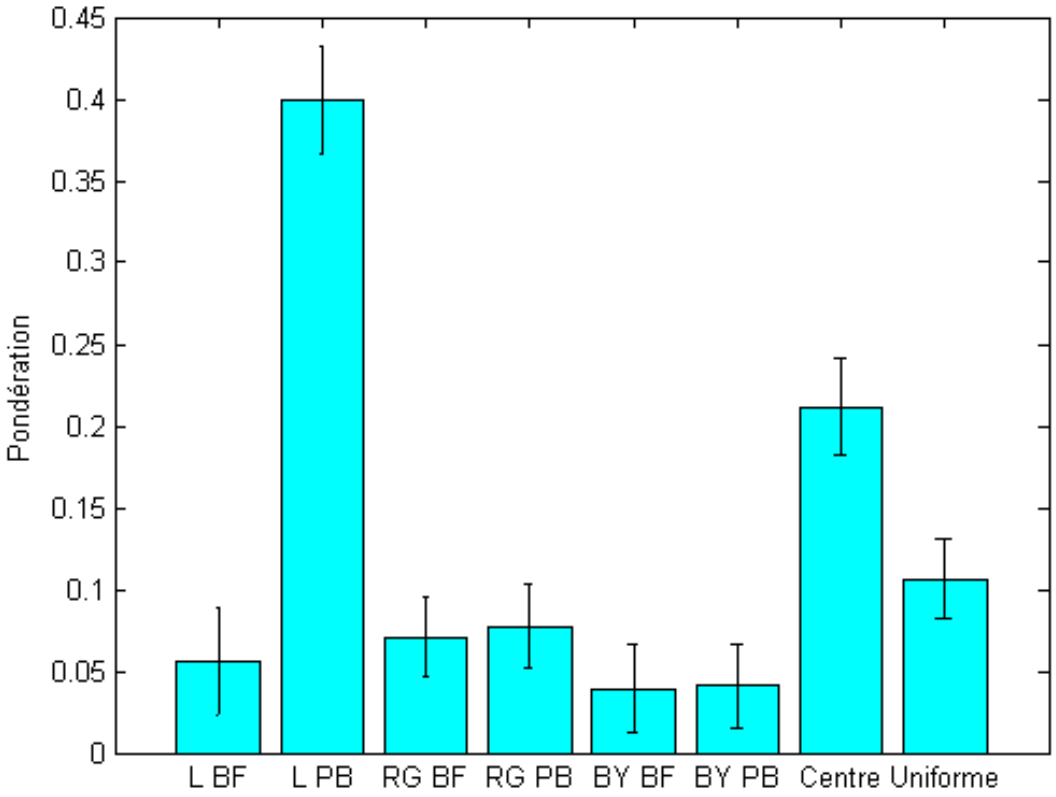


FIG. 4.18 – Contributions des 8 caractéristiques (voir texte) aux mouvements oculaires pour toutes les images et toutes les fixations. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap. BF signifie 'basse fréquence" et PB "passe-bande".

la distribution. Les résultats ont montré que lors d'une exploration, il existe une partie des mouvements oculaires commandés par des facteurs non identifiés, voire par le hasard.

Nous effectuons maintenant une deuxième modélisation, en supprimant les facteurs liés à la chrominance. Il ne reste plus que 4 facteurs. Selon la figure 4.19, les relations entre les pondérations des caractéristiques restantes changent peu. La luminance passe-bande est encore le facteur le plus important. Néanmoins, l'écart entre la contribution de la luminance basse fréquence et la contribution de la distribution uniforme est maintenant plus faible que dans le cas précédent. Cette situation pourrait s'expliquer par l'addition de la contribution des facteurs de chrominance basse fréquence à celle de la luminance basse fréquence puisqu'il y a de forte corrélation entre la luminance et la chrominance (cf. section §3.3).

Cette expérience a été conçue pour évaluer les contributions des caractéristiques de bas niveau, qui sont liées aux images, dans les mouvements oculaires lors de l'exploration libre de scènes naturelles. Les résultats ont montré que la luminance passe-bande apporte la contribution la plus importante. D'ailleurs, nous avons également

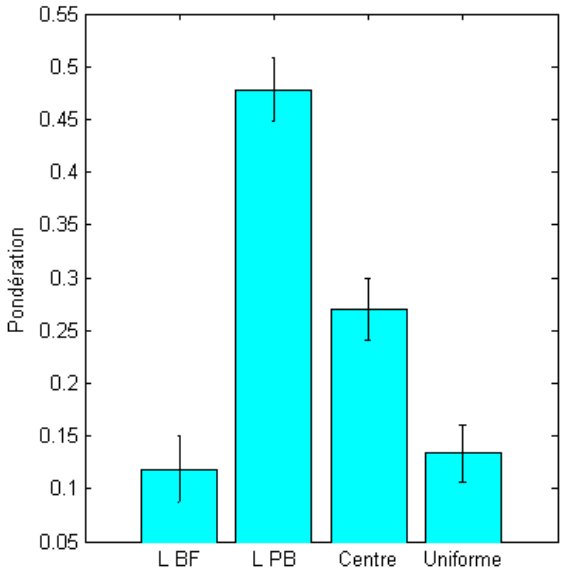


FIG. 4.19 – Contributions des 4 caractéristiques : luminance basse fréquence (LBF), luminance passe-bande (LPB), biais de centralité et la distribution uniforme aux mouvements oculaires pour toutes les images et toutes les fixations. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap.

observé qu'une partie des mouvements oculaires s'explique par les facteurs non liés aux images et représentés dans notre modèle par le biais de centralité et la distribution uniforme.

Dans la suite, nous évaluons les contributions des caractéristiques de la même manière que précédemment mais selon l'ordre des fixations.

4.2.3 Evolution temporelle des contributions

Jusqu'à présent, nous avons analysé globalement les contributions des facteurs durant les premières secondes de l'exploration libre de scènes. Nous abordons maintenant la question de l'évolution de ces contributions au cours de cette exploration. En effet, les recherches en neurophysiologie ont révélé la séparation de différentes voies visuelles à l'issue de la rétine en une voie de luminance et deux voies de chrominance mais la question concernant la perception de ces voies au niveau temporel reste encore ouverte. Ici, nous essayons d'étudier cette question en examinant la variation temporelle des contributions de ces caractéristiques par le modèle "EM carte".

Le modèle "EM carte" est appliqué sur les fixations sur l'ensemble des images de la même manière que précédemment. Cependant, il est effectué selon l'ordre des 8 premières fixations. Pour chaque rang de fixation, nous regroupons les fixations de tous les sujets pour toutes les images en couleur. La figure 4.20 représente l'évolution temporelle des contributions des caractéristiques. Les courbes sont tracées selon

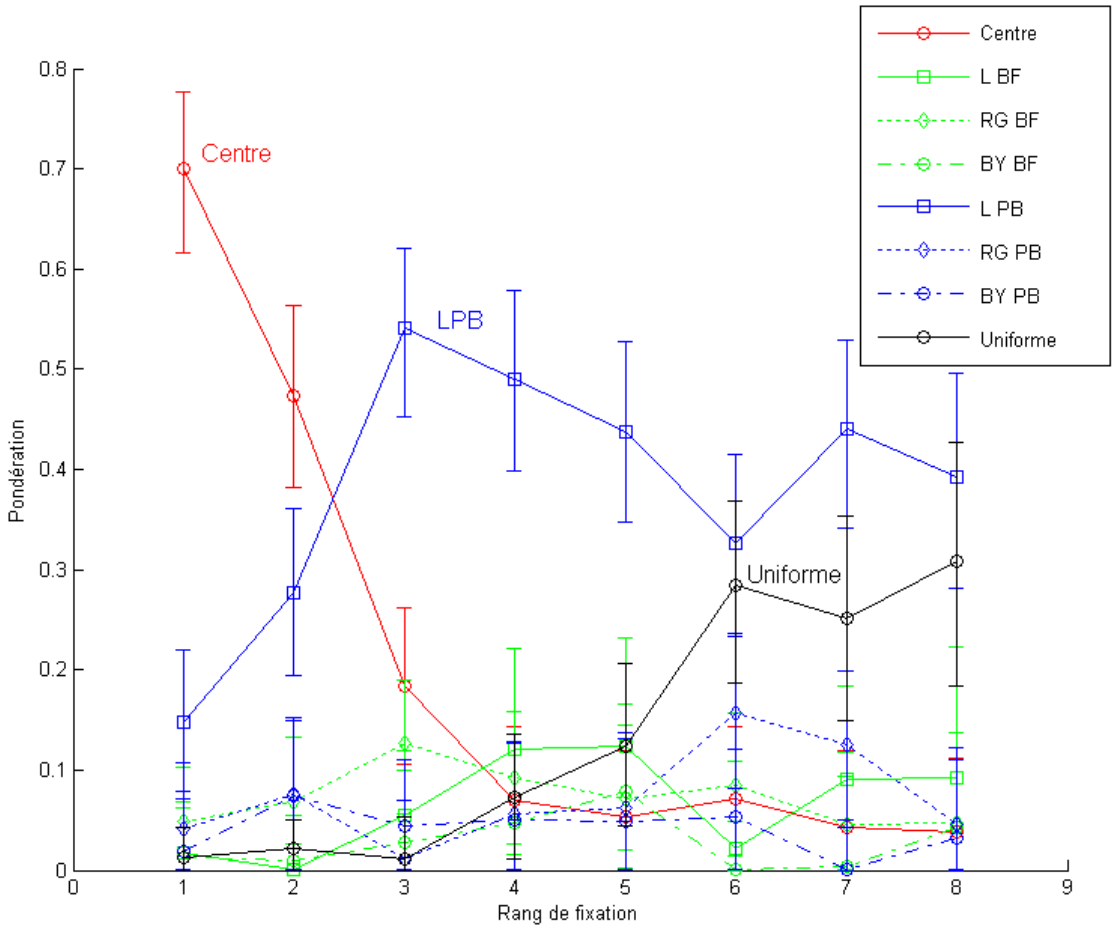


FIG. 4.20 – Evolution temporelle des contributions des caractéristiques pour l'ensemble des images en couleur en fonction du rang de la fixation. L'intervalle de confiance à 95% est calculé par la technique de Bootstrap.

l'ordre des 8 premières fixations. Nous observons que pour les 5 caractéristiques (les caractéristiques de couleur et la luminance basse fréquence) qui globalement contribuaient peu à l'attention, la variation temporelle de leur contribution reste également non significative. En revanche, les variations des contributions de la luminance passe-bande, du biais de centralité et de la distribution uniforme sont importantes.

Au début de l'exploration, le biais de centralité apporte une très grande contribution à l'attention visuelle. Ensuite, il chute très vite et après 3 fixations, sa contribution est de même ordre de grandeur que les voies chromatiques. Nous rappelons que la fixation initiale d'un sujet a été majoritairement choisie dans les coins de l'image. Pourtant, le biais de centralité est de suite important (70%) avant de diminuer.

Le rôle de la luminance passe-bande est toujours important à partir de la 3ème fixation. Au début de l'exploration, elle est dominée par le biais de centralité mais supérieure aux autres caractéristiques. Ensuite, la luminance passe-bande augmente rapidement, se stabilise, puis décroît lentement tout en restant le facteur essentiel expliquant les mouvements oculaires.

Nous analysons maintenant la contribution du mode de densité uniforme. Cette contribution est très faible au début puis augmente progressivement mais demeure bien au dessous de celle de la luminance passe-bande. A la fin, sa contribution représente 30% des fixations couleur. L'augmentation de ce poids reflète deux constatations liées : les fixations deviennent plus dispersées et les facteurs de bas niveau de l'attention analysés dans ce modèle (les 6 autres facteurs hormis le biais de centralité) voient leur pouvoir d'explication diminuer. On peut penser qu'après 3 premières fixations, l'attention visuelle est influencée par les facteurs de haut niveau, qui ont des influences très variables sur les différents sujets. De plus, et c'est une évidence, dans l'exploration libre, les sujets sont susceptibles de regarder partout dans une scène. Ces deux éléments entraînent la croissance de dispersion des fixations entre sujets ; ce qui est représentée par l'augmentation du rôle de la distribution uniforme.

Ici, les facteurs représentant le biais de centralité et la distribution uniforme qui sont dans le modèle les seuls deux facteurs non associés à des caractéristiques de bas niveau, ont des sens de variation opposés. La contribution du premier est importante au début de l'exploration et c'est le contraire pour la contribution du second. Cependant, la luminance passe-bande apporte la contribution la plus importante.

5 Conclusion

Dans ce chapitre, nous avons étudié les contributions des caractéristiques visuelles à l'attention. Dans un premier temps, nous avons comparé par l'analyse non-paramétrique les mouvements oculaires lors d'une exploration libre d'une image en couleur et en niveau de gris. Les fixations couleur et niveau de gris proviennent de deux groupes de sujets différents. La comparaison s'est faite en se basant sur les caractéristiques de bas niveau : le contraste de luminance et les contrastes de deux voies chromatiques, Rouge-Vert et Bleu - Jaune. Les résultats ont montré qu'il n'y avait pas de différence significative entre les fixations couleur et les fixations niveau de gris. Cette comparaison a été effectuée pour l'ensemble des images et pour différentes catégories sémantiques. Dans tous les cas, les trois caractéristiques n'ont pas révélé de différence entre ces deux types de fixations.

Dans un deuxième temps, l'analyse paramétrique a été utilisée pour étudier les mouvements oculaires. D'abord, nous avons comparé les fixations couleur et les fixations niveau de gris en modélisant la distribution de leurs positions par un modèle de densité spatiale. Les résultats ont montré la similarité entre les fixations couleur et niveau de gris par un critère quantitatif de vraisemblance moyenne. Ainsi, nous en avons conclu que la couleur apporte peu par rapport à la luminance aux mouvements oculaires lors de l'exploration libre de scènes naturelles.

Ensuite, nous avons essayé de quantifier les contributions de plusieurs caractéristiques aux mouvements oculaires par le modèle "EM carte". Deux méthodes ont été utilisées : l'une pour l'ensemble des fixations et l'autre en fonction de l'ordre de fixations. Les caractéristiques sont maintenant calculées par le modèle d'attention

visuelle que nous avons présenté au chapitre 2. Les résultats ont montré que la luminance passe-bande joue le rôle le plus important dans l'explication des mouvements oculaires. Les voies chromatiques ainsi que la luminance basse fréquence apportent une faible contribution.

Les études ont également révélé l'influence des facteurs non associés aux images : le biais de centralité et la distribution uniforme. Le rôle du premier, qui a souvent été observé dans la littérature, a été confirmé par le modèle "EM carte". Son influence est importante au début de la présentation d'une scène mais diminue au cours du temps. En parallèle, on note une augmentation de la dispersion des fixations modélisée par une distribution uniforme. Cependant, la contribution de la luminance passe-bande demeure la plus importante au cours du temps. Cela confirme le rôle des facteurs ascendants dans l'exploration libre de scènes naturelles.

Ainsi, en reprenant le modèle d'attention visuelle décrit au chapitre 2 (Fig. 2.14), nous pouvons garder seulement la voie passe bande de luminance afin de créer la carte de saillance d'une image pour prédire les zones fixées par des sujets.

Chapitre 5

Programmation de saccade

1 Introduction

Nous avons vu aux chapitres précédents que les gens explorent leur environnement visuel en bougeant constamment les yeux ; ces mouvements des yeux sont appelés des saccades. Nous bougeons nos yeux pour placer l'objet d'intérêt (la zone regardée) au centre de nos yeux. En effet comme nous l'avons vu au chapitre 2, la répartition des photorécepteurs à la surface de la rétine n'est pas uniforme et leur densité est maximum au centre. Ainsi, la zone d'intérêt est placée au centre de la rétine, là où la densité des photorécepteurs est maximale, et donc là où l'acuité visuelle ou la résolution spatiale est maximale [Wandell, 1995]. Cette densité non uniforme est représentée par la variation d'échelle spatiale d'une image. Dans la suite de ce chapitre nous appellerons la zone d'intérêt : le point de vue. Ainsi, l'échelle de l'image au point de vue est la plus fine tandis que l'échelle à la périphérie de ce point de vue est de moins en moins fine (Fig. 5.1). Nous allons dans ce chapitre modéliser cette échelle spatialement variante pour une image. Notre but premier étant de modéliser la stratégie de programmation de saccade lors de l'exploration libre de scènes.

Des études ont montré que dans certaines conditions nous programmions plusieurs saccades en parallèle. A partir d'un même point de vue, les deux saccades suivantes sont programmées [Becker and Jürgens, 1979; Hooge and Erkelens, 1998; McPeek et al., 2000]. Cette programmation de deux saccades en parallèle est appelée : *"concurrent saccade programming"*. Ces études montrent qu'en utilisant une telle stratégie nous sommes plus rapide dans des tâches de recherche de cible que si nous programmions les saccades les unes après les autres à partir de points de vue à chaque fois différents. Que se passe-t-il dans les situations expérimentales que nous avons jusqu'ici utilisées ? Les sujets programment-ils leurs saccades en parallèle ou séquentiellement lorsqu'ils explorent librement une scène naturelle ?

Dans ce chapitre, nous allons répondre à ces questions en utilisant d'une part une expérience oculométrique nous permettant d'obtenir des données réelles, et d'autre part le modèle de saillance décrit au chapitre 2. Nous supposons que les saccades sont programmées à l'aide de la carte de saillance. Autrement dit, le problème est étudié dans le cadre des processus ascendants (*"Bottom-Up"*) dans lequel l'attention

visuelle est stimulée par les caractéristiques visuelles de bas niveau. Comme nous avons montré au chapitre précédent que la couleur semble jouer un rôle faible dans la saillance, seules des images en niveau de gris sont utilisées dans ce chapitre.

L'organisation de ce chapitre est la suivante. Nous passerons tout d'abord rapidement en revue la décroissance de l'échelle par rapport au point de vue sur une image ; cela est représentée par un filtrage passe-bas spatialement variant. Nous présenterons ensuite le modèle de décroissance de l'échelle que nous avons choisi de développer, puis notre filtre passe-bas spatialement variant. Ensuite, en intégrant ce type de filtre au modèle d'attention visuelle décrit précédemment, nous testerons plusieurs stratégies de programmation de saccade. A partir de ces stratégies, nous allons extraire des données qui seront comparées aux données expérimentales issues de l'expérience en oculométrie ; ce qui nous permettra de voir quelle stratégie se rapproche le plus des données expérimentales.

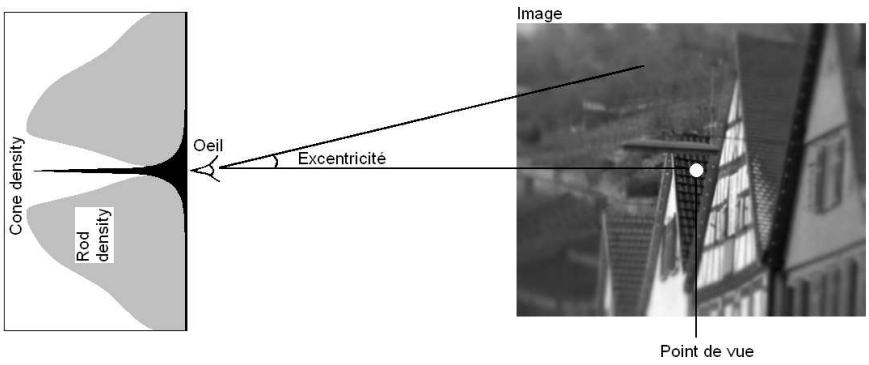


FIG. 5.1 – Le point de vue est perçu avec l'échelle la plus fine tandis que la périphérie de ce point est perçue avec une échelle de moins en moins fine. Le point de vue correspond à la fovéa sur la rétine où la densité des cônes est maximale.

2 Filtre passe-bas spatialement variant

La décroissance de l'échelle spatiale des stimuli par rapport au point de vue a déjà été abordée dans la littérature. Dans [Parkhurst et al., 2002], cette décroissance de l'échelle est simplement modélisée par une fonction gaussienne appliquée à la carte de saillance en sortie de son modèle d'attention visuelle. En faisant cela, il part du principe que le sujet porte son regard au centre et que la saillance diminue avec la décroissance de l'échelle. Malgré sa simplicité, cette implémentation améliore la prédiction du modèle.

Contrairement à Parkhurst, dans [Itti, 2006], la variation de l'échelle spatiale est implémentée en entrée de son modèle d'attention visuelle par un filtre passe-bas spatialement variant. Ce filtre correspond en fait à un filtrage passe-bas dont la fréquence de coupure varie en fonction de l'excentricité par rapport au point de vue. Ainsi, à la différence d'un filtre passe-bas unique qui conserve la même fréquence

de coupure, un filtre passe-bas spatialement variant dispose d'une fréquence de coupure qui décroît avec l'excentricité par rapport au point de vue. L'utilisation d'un tel filtrage augmente également les performances de prédiction de son modèle de d'attention visuelle.

Perry [Perry and Geisler, 2002] propose également une modélisation du filtrage passe-bas spatialement variant. Ce filtrage a initialement été utilisé pour la communication de vidéo à faible bande passante [Geisler and Perry, 1998]. Le filtrage passe-bas spatialement variant a ensuite servi à représenter des stimuli à échelle variable sur un écran lors d'une expérience de recherche visuelle [Geisler et al., 2006].

Dans ce chapitre, nous modéliserons également l'échelle spatialement variante par un filtre passe-bas spatialement variant. Nous allons décrire le modèle proposé par Perry, et ensuite, nous proposerons le nôtre. Ces deux modèles diffèrent à la fois au niveau de la courbe de décroissance de l'échelle en fonction de l'excentricité et au niveau de l'implémentation du filtre passe-bas spatialement variant.

2.1 Le modèle de Perry

2.1.1 Décroissance de l'échelle spatiale

Dans le modèle de Perry, la décroissance de l'échelle est décrite par une loi hyperbolique :

$$E(e) = A\frac{\alpha}{\alpha + e} \tag{5.1}$$

avec e l'excentricité (en degré) par rapport au point de vue. La constante A est inférieure ou égale à 1 et dépend de la distance entre le sujet et l'écran [Perry, 2002]. Le paramètre α représente l'excentricité à laquelle l'échelle est égale à la moitié de l'échelle maximale (l'échelle la plus fine). Ce paramètre détermine la "vitesse" de décroissance de l'échelle. Plus il est faible, plus la décroissance est rapide et inversement. Lorsque α tend vers l'infini, l'échelle est uniforme. Nous notons également que la loi de décroissance de l'échelle dans le modèle de Perry est obtenue à partir des données expérimentales sur la sensibilité au contraste en fonction de l'excentricité [Geisler and Perry, 1998].

Ainsi, pour une image donnée et un point de vue sur cette image, nous pouvons calculer l'échelle pour tous les pixels à la position (x, y) sur cette image en utilisant l'équation 5.1 ; cela crée une carte d'échelle $E(x, y)$. L'image est donc filtrée par un filtre passe-bas spatialement variant dont la fréquence de coupure, correspondant à l'échelle, diminue du point de vue à la périphérie pour obtenir en sortie une image de taille égale.

2.1.2 Implémentation du filtre passe-bas spatialement variant

Le filtrage passe-bas spatialement variant d'une image s'effectue à partir d'une décomposition pyramidale multirésolution de l'image. La pyramide dispose d'un certain nombre de niveaux correspondant à différentes résolutions allant de la résolution la plus nette à la plus floue. L'image d'entrée (P0) correspond au niveau 0 de la pyramide. Elle est ensuite filtrée par un filtre passe-bas et sous-échantillonnée par 2 pour

FIG. 5.2 – Exemple de la décompostion pyramidale d'une image selon quatre niveaux : l'image au niveau 0 est l'image originale (P0), l'image au niveau 1 est issue d'un filtrage passe-bas de l'image originale et puis sous-échantillonnée par 2 (P1). Le processus continue jusqu'au dernier niveau de la pyramide ([Perry and Geisler, 2002]).

donner l'image au niveau 1 (P1). Ce processus continue jusqu'au dernier niveau de la pyramide. La figure 5.2 illustre les images aux 4 premiers niveaux de la pyramide. L'image à chaque niveau de la pyramide est considérée comme la sortie du filtrage de l'image d'entrée par un filtre passe-bas dont la fréquence de coupure est égale à l'échelle à ce niveau de la pyramide. Les fonctions de transfert des filtres passe-bas utilisés par Perry sont illustrées à la figure 5.3. La fréquence de coupure du filtre passe-bas au niveau i est notée E_i et elle est déterminée par la fréquence spatiale à la mi-hauteur de la fonction de transfert (en valeur relative, $E_i = \frac{0.992}{2^i}$).

Cette décomposition pyramidale d'une image est utilisée pour effectuer le filtrage passe-bas spatialement variant. En réalité, le nombre de niveaux de la pyramide est limité tandis que la variation d'échelle selon l'équation 5.1 est continue. En conséquence, la détermination de la sortie d'un filtre passe-bas qui n'appartient pas à un niveau de la pyramide se fait grâce à l'interpolation des images à deux niveaux voisins. Par exemple, la sortie du filtre passe-bas correspondant à la fonction de transfert en pointillé dans la figure 5.3 est interpolée à partir des images I_2 au niveau 2 et I_3 au niveau 3. Les coefficients d'interpolation sont représentés par les fonctions B_i qui sont générées à tous les niveaux i de la pyramide. Pour chaque niveau et pour chaque pixel, ces coefficients B_i sont calculés :

$$B_i(x,y) = \begin{cases} 0 & \text{si } E(x,y) \leq E_i \\ \dfrac{0.5 - T_i\left(E(x,y)\right)}{T_{i-1}\left(E(x,y)\right) - T_i\left(E(x,y)\right)} & \text{si } E_i \leq E(x,y) \leq E_{i-1} \\ 1 & \text{si } E(x,y) \geq E_{i-1} \end{cases} \tag{5.2}$$

où (x,y) est la position d'un pixel
$E(x,y)$ l'échelle à la position (x,y)
E_i l'échelle au niveau i de la pyramide
T_i la fonction de transfert d'un filtre passe-bas au niveau i.

L'interpolation a lieu aux pixels où la valeur de B_i est strictement comprise entre 0 et 1 selon l'équation[1] :

[1]Pour combiner deux images de deux niveaux différents, il faut sur-échantillonner l'image de

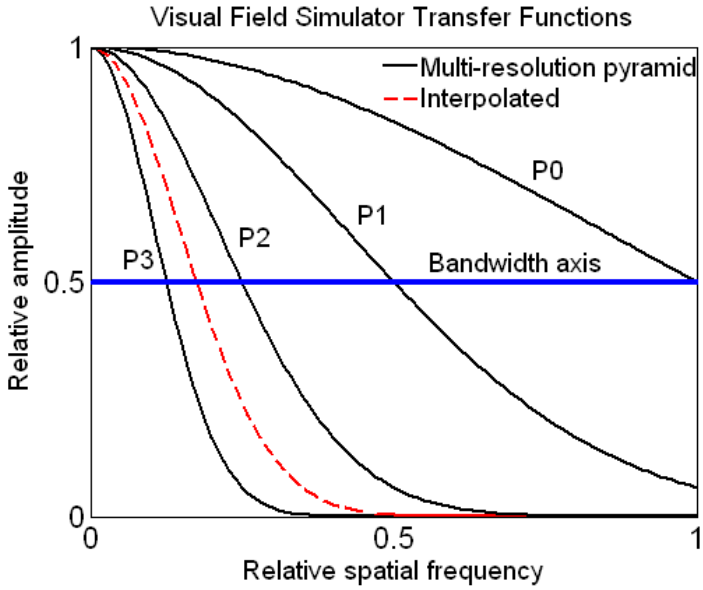


FIG. 5.3 – Les fonctions de transfert des filtres passe-bas (courbe pleine) utilisés pour créer les différentes images de la pyramide. La fonction de transfert d'un filtre passe-bas pour une fréquence de coupure quelconque (courbe pointillée) est interpolée à partir des fonctions de transfert de deux filtres voisins, ici P2 et P3 ([Perry and Geisler, 2002]).

$$O(x,y) = B_i(x,y)I_i(x,y) + (1 - B_i(x,y))\, I_{i-1}(x,y). \tag{5.3}$$

La combinaison des images à tous les niveaux de la pyramide commence par la résolution la plus floue et monte jusqu'à la plus nette. A chaque niveau, les pixels dont la valeur B_i est égale à 0, correspondent à la résolution plus floue et donc l'interpolation a déjà eu lieu. Les pixels dont B_i est égal à 1 correspondent à la résolution plus nette et pourront être interpolés. Finalement, les images à différents niveaux sont combinées pour créer l'image filtrée spatialement variante.

2.2 Le modèle proposé

2.2.1 Décroissance de l'échelle spatiale

Dans le modèle de Perry décrit ci-dessus, la variation de l'échelle est déterminée à partir de la fonction de sensibilité au contraste mesurée chez l'homme. Ici, nous préférons modéliser la décroissance de l'échelle en partant des propriétés anatomiques de la rétine, et plus particulièrement des courbes de densité des photorécepteurs et des courbes de densité des cellules ganglionnaires à la surface de la rétine.

Nous utilisons la courbe proposée par Osterberg (1935) et déjà présentée au chapitre 2 qui présente la densité des photorécepteurs à la surface de la rétine. Ici, nous nous intéressons uniquement aux cônes car ces photorécepteurs déterminent l'acuité visuelle. Nous utiliserons également une courbe qui représente l'évolution du rapport

résolution plus floue pour que les deux images à combiner aient la même taille.

cônes/ganglionnaires en fonction de l'excentricité par rapport à la fovéa (Fig. 5.4). Ainsi de la même manière que le nombre de cônes diminue avec l'excentricité par rapport à la fovéa, le nombre de cellules ganglionnaires diminuent également avec l'excentricité.

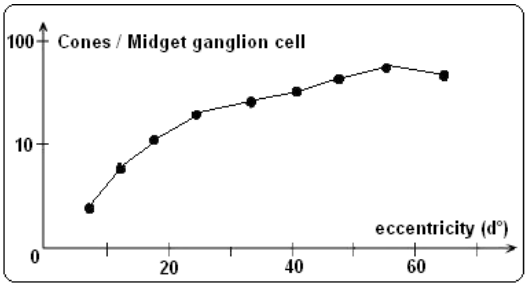


FIG. 5.4 – La courbe représentant le rapport cônes par cellule ganglionnaire en fonction de l'excentricité [Goodchild et al., 1996].

Nous allons maintenant modéliser la densité des cellules ganglionnaires, qui véhiculent les informations visuelles à la sortie de la rétine, pour représenter la décroissance de l'échelle spatiale des stimuli. Selon [Hérault, 2009], la densité des cellules ganglionnaires peut être approximée par cette formule :

$$d(e) = \sqrt{\frac{M_c(e)}{M_{cpg}(e)}} \tag{5.4}$$

où e est l'excentricité (en mm) sur la rétine par rapport à la fovéa,
M_c la densité des cônes en fonction de l'excentricité (en mm),
M_{cpg} le rapport cônes par cellule ganglionnaire en fonction de l'excentricité par rapport à la fovéa (en mm).
M_c est calculée à partir de la courbe de la figure 2.3 (cf. chapitre 2). Pour le rapport cônes par cellule ganglionnaire (Fig. 5.4), il est approximée par la formule suivante [Hérault, 2009] :

$$M_{cpg}(e) = 2 + 48\frac{e^3 + e^2 + e}{e^3 + e^2 + e + 700} \tag{5.5}$$

Finalement, nous obtenons la densité $d(e)$ des cellules ganglionnaires en fonction de l'excentricité par rapport à la fovéa (en mm). Nous convertissons ensuite l'excentricité en angle visuel selon le modèle de l'optique de l'œil [Drasdo and Fowler, 1974] pour obtenir la densité des cellules ganglionnaires en fonction de l'excentricité en degré. Cette courbe sera utilisée pour représenter la variation de l'échelle dans le filtrage passe-bas spatialement variant. La figure 5.5 représente la courbe de variation de l'échelle modélisée par la densité des cellules ganglionnaires et la courbe de variation de l'échelle du modèle de Perry. Ces deux courbes sont proches et pourtant elles sont issues de données différentes, la sensibilité au contraste pour le modèle de Perry et la courbe de densité des cellules ganglionnaires pour le modèle que nous proposons.

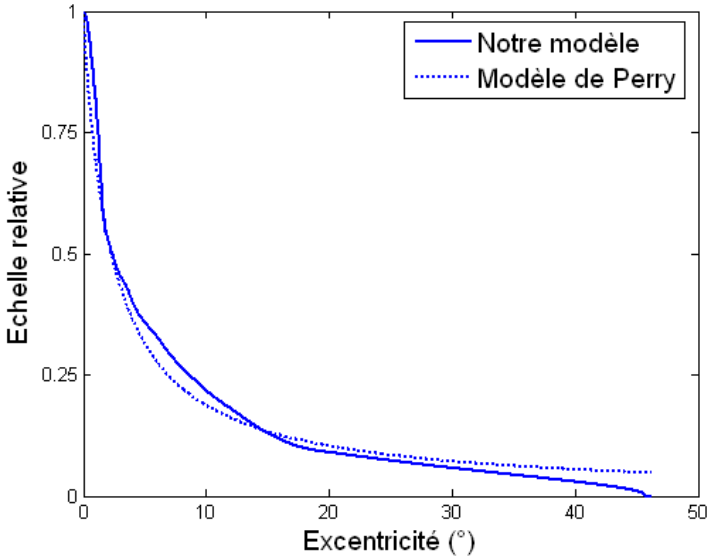


FIG. 5.5 – La courbe de variation de l'échelle selon le modèle de Perry (le paramètre contrôlant la vitesse de décroissance de l'échelle $\alpha = 2.3°$) et la courbe de variation de l'échelle selon la densité des cellules ganglionnaires à la surface de la rétine. L'excentricité 0° correspond au centre de la fovéa.

2.2.2 Implémentation du filtre passe-bas spatialement variant

La densité des cellules ganglionnaires est considérée comme la fréquence d'échantillonnage de la rétine. La fréquence de coupure du filtre passe-bas spatialement variant est choisie égale à la moitié de la fréquence d'échantillonnage (théorème de Shannon) :

$$\lambda_c(e) = \frac{1}{2} f_e(e) = \frac{1}{2} \frac{d(e)}{d_{max}} \tag{5.6}$$

avec $\lambda_c(e)$ la fréquence de coupure, $d(e)$ la densité des cellules ganglionnaires, d_{max} la valeur maximale de $d(e)$ et $f_e(e)$ la fréquence d'échantillonnage en fonction de l'excentricité (en degré).

Le filtre passe-bas spatialement variant est effectué par un filtre récursif de type passe-bas sous la forme suivante :

$$H(z) = (1-a)^2 \frac{1}{1-az} \frac{1}{1-az^{-1}} \tag{5.7}$$

où a est le paramètre du filtre passe-bas. En remplaçant $z = e^{j2\pi\lambda}$ avec λ la fréquence réduite, $H(z)$ devient :

$$H(\lambda) = \frac{(1-a)^2}{1 - 2a\cos(2\pi\lambda) + a^2} \tag{5.8}$$

La fréquence de coupure λ_c est déterminée par :

$$H(\lambda_c) = \frac{1}{\sqrt{2}} \tag{5.9}$$

ou :

$$(1-\sqrt{2})a^2 + 2\left(\sqrt{2} - \cos(2\pi\lambda_c)\right)a + (1-\sqrt{2}) = 0 \tag{5.10}$$

A partir de l'équation 5.10, on peut en déduire la valeur de a en choisissant $0 < |a| < 1$.

Pour une image donnée et un point de vue sur cette image, nous pouvons déterminer une carte de fréquences de coupure en fonction de l'excentricité par rapport à ce point, et puis une carte de paramètres a du filtre passe-bas $H(z)$. Finalement, pour effectuer le filtrage passe-bas spatialement variant d'une image, nous appliquons le filtre $H(z)$ suivant la direction horizontale et puis verticale. Cette méthode permet de filtrer une image en deux dimensions par un filtre à une dimension. L'implémentation de ce filtre dans le domaine spatial est présentée à l'annexe E.

2.3 Comparaison des deux modèles

Notre modèle de filtrage passe-bas spatialement variant diffère du modèle de Perry sur deux aspects. Premièrement, dans notre modèle, la courbe de variation de l'échelle est construite à partir de la densité des cellules ganglionnaires, tandis que dans le modèle de Perry, elle se base sur les données de la sensibilité au contraste. Deuxièmement, la différence entre les deux modèles réside dans la manière d'implémenter le filtrage passe-bas à chaque position. Alors que Perry a utilisé la décomposition pyramidale et l'interpolation des images à différents niveaux, nous avons utilisé un filtrage récursif.

Les résultats obtenus à partir de notre modèle et celui de Perry avec le paramètre contrôlant la vitesse de décroissance de l'échelle $\alpha = 2.3°$ sont très similaires (Fig. 5.6). Néanmoins, dans le modèle de Perry, on peut faire varier ce paramètre α, cela permet de tester l'influence de la vitesse de décroissance de l'échelle sur par exemple des paramètres de saccades générées par ce modèle. Le modèle de Perry sera utilisé pour l'étude sur la programmation de saccade dans ce chapitre. Notre modèle servira à l'échantillonnage à taux variable décrit au chapitre 6.

(a) Original (b) Perry (c) Notre modèle

FIG. 5.6 – Comparaison des résultats des filtres passe-bas spatialement variant du modèle de Perry avec le paramètre contrôlant la vitesse de décroissance de l'échelle $\alpha = 2.3°$ et de notre modèle. Le point de vue est au centre de l'image.

3 Modèles de programmation de saccade

Nous allons décrire en détails trois modèles de programmation de saccade. Nous supposons que les mouvements oculaires dépendent des caractéristiques visuelles de bas niveau des scènes naturelles et qu'ils peuvent être prédits par des modèles de saillance. Nous commençons donc par présenter la carte de saillance pour une image en niveau de gris. Nous utilisons ici le modèle pour prédire les 4 premières fixations après l'apparition de l'image.

3.1 Carte de saillance

Nous présentons la carte de saillance pour des scènes en niveau de gris. Elle correspond principalement à la voie de luminance du modèle d'attention visuelle décrit au chapitre 2 avec certaines modifications.

Dans un premier temps, nous intégrons le filtrage passe-bas spatialement variant comme la première étape du traitement rétinien. Ainsi, le filtre passe-bas spatialement variant de Perry est appliqué pour une image d'entrée afin de modéliser la décroissance de l'échelle des stimuli en fonction de l'excentricité. Le filtrage est suivi par le traitement rétinien pour la voie de luminance comme décrit au chapitre 2. La sortie de la rétine est ici la combinaison linéaire des cellules parvocellulaires et des cellules magnocellulaires. La présence des basses fréquences (cellules magnocellulaires) à l'issue du traitement rétinien est également appropriée à la propriété de la perception visuelle dans laquelle les basses fréquences précèdent les hautes fréquences [Navon, 1977].

La sortie de la rétine est ensuite décomposée par le banc de filtres corticaux décrits au chapitre 2. Il comporte un banc de filtres LogNormaux suivant 8 orientations et 4 bandes de fréquence, les normalisations des sorties des filtres, les interactions entre les sorties des filtres et la fusion des sorties des filtres afin de créer la carte de saillance de l'image.

3.2 Description des modèles

En s'appuyant sur la carte de saillance, nous testons trois modèles de programmation de saccade qui sont résumés à la figure 5.7. Pour le premier modèle, à partir d'un point de vue (d'une carte de saillance), 4 fixations sont prédites. Ce modèle est appelé "1M". Le modèle "1M" servira d'un modèle de base pour comparer les deux autres modèles. Le deuxième modèle appelé "2M" prédit à partir d'un point de vue les 2 fixations suivantes, et puis, à partir d'un deuxième point de vue (la fixation X_2) les 2 fixations suivantes. Ce modèle nécessite donc 2 cartes de saillance. Le troisième modèle ("4M") prédit à partir d'un point de vue uniquement la fixation suivante. Ainsi, il nécessite 4 cartes de saillance ; chacune permettant de prédire une fixation.

La notion de point de vue est importante et est liée au champ visuel d'un sujet. Nous supposons que l'aire du champ visuel est fixée et égale à la taille de l'écran (ou d'une image). Quand le point de vue se déplace, la portion de l'image originale tombant dans le champ visuel change aussi. Le champ visuel est alors comblé en

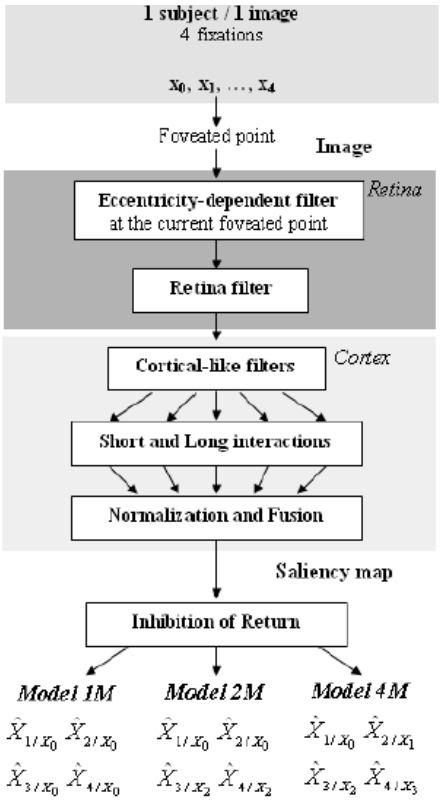
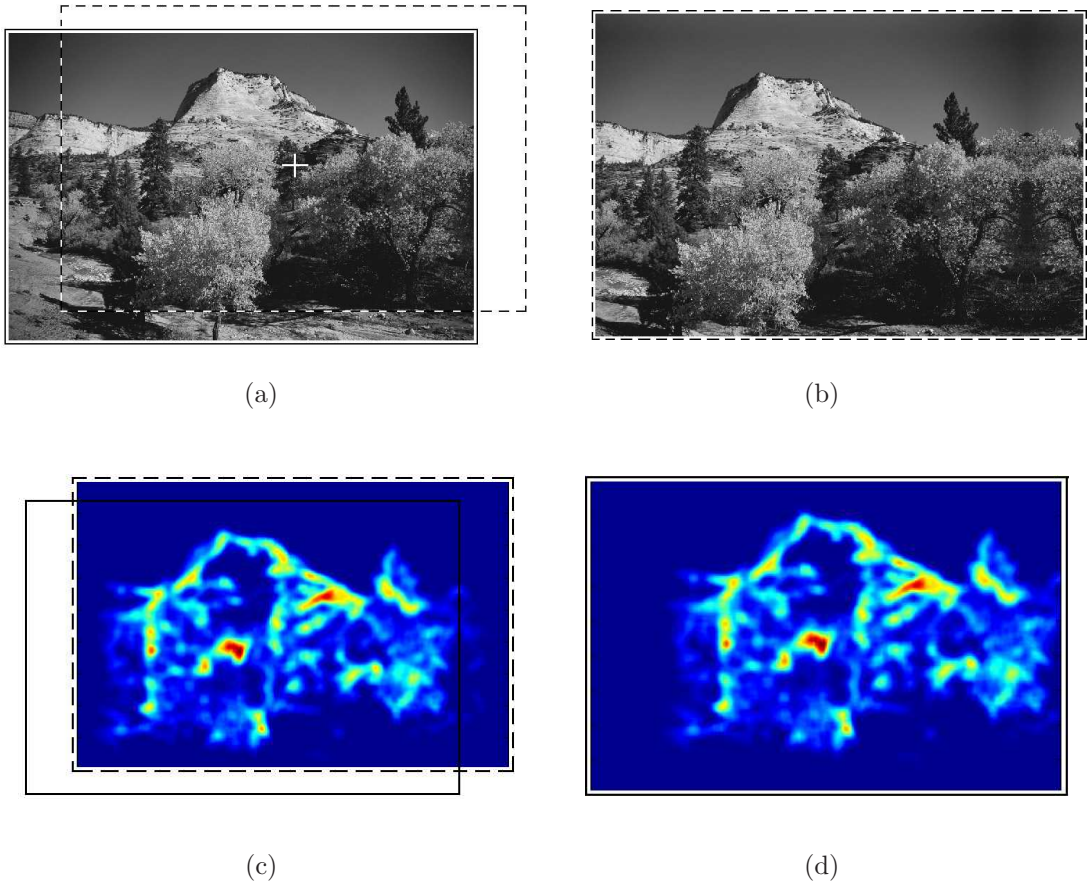


FIG. 5.7 – Les trois modèles de programmation de saccade se basent sur la carte de saillance et l'inhibition de retour pour prédire les 4 premières fixations ou saccades [Ho-Phuoc et al., 2010].

prenant la symétrisation au bord de l'image originale. Alors que ce champ visuel créé de cette façon est artificiel, il permet d'éviter des effets de bord lors du calcul d'une carte de saillance. De plus, la partie étendue par la symétrisation n'est pas gardée pour l'analyse. La figure 5.8 illustre un exemple du calcul d'une carte de saillance pour un point de vue. Dans cette figure, le cadre de l'écran est représenté par le rectangle en trait plein. Le champ visuel représenté par le rectangle pointillé peut se déplacer mais est toujours égal à la taille de l'écran. La carte de saillance est calculée dans le champ visuel actuel et puis ramenée au champ visuel correspondant à l'écran (Fig. 5.8d).

Dans le modèle "1M", à partir du point de vue initial, ici au centre de l'image, le filtre passe-bas spatialement variant est appliqué. Ensuite, les traitements rétinien et cortical interviennent pour créer la carte de saillance. Cette dernière permet de prédire les 4 premières saccades correspondant aux 4 premières fixations. Pour cela nous utilisons l'inhibition de retour (IOR) (cf. chapitre 1). En s'appuyant sur la carte de saillance, la première fixation est choisie comme le pixel dont la saillance est maximale. Puis, le mécanisme d'IOR est appliqué sur ce pixel pour empêcher la fixation suivante d'y revenir. IOR est un masque de 1° de rayon permettant de mettre à zéro tous les pixels à l'intérieur de ce masque. Ensuite, le maximum de la



(a) (b)

(c) (d)

FIG. 5.8 – Exemple de calcul d'une carte de saillance lors du déplacement du point de vue : (a) L'image originale (rectangle en trait plein) et le champ visuel (rectangle en pointillé) avec le point de vue actuel (croix) ; (b) Le champ visuel actuel est rempli par la symétrisation au bord de l'image originale ; (c) La carte de saillance dans le champ visuel actuel ; (d) La carte de saillance ramenée dans le cadre de l'image originale.

carte de saillance masquée par IOR est cherchée et correspond à la deuxième fixation. Puis, le masque IOR intervient à nouveau. Cette démarche se poursuit jusqu'à la quatrième fixation.

Le modèle "2M" utilise 2 points de vue différents, et à partir de ces 2 points de vue, le modèle d'attention visuelle est appliqué. Dans un premier temps, comme dans le modèle "1M", à partir du point de vue initial au centre de l'image, la première carte de saillance est calculée. Ensuite, nous appliquons le mécanisme d'IOR pour choisir les 2 premières fixations. Maintenant, contrairement au modèle "1M", le point de vue est réinitialisé et à partir de ce point de vue la nouvelle carte de saillance est calculée. Elle permet de prédire les 2 fixations suivantes. Ce qui est important dans le modèle "2M" est la réinitialisation du point de vue. Le deuxième point de vue est initialisé par la deuxième fixation d'un sujet sur une image. Car chaque sujet a une trajectoire différente de mouvements oculaires, le deuxième point de vue dépend du sujet et de l'image. Par conséquent, le modèle "2M" va être appliqué pour chaque

couple “image × sujet”.

De façon similaire, le modèle “4M” nécessite 4 points de vue différents. Tandis que le premier point de vue est commun pour tous les sujets et correspond au centre de l'image, les trois points de vue suivants dépendent de chaque sujet et de chaque image. Pour chaque point de vue, une carte de saillance est calculée afin de prédire la fixation suivante. Puis, le point de vue est réinitialisé selon chaque couple “image × sujet”.

Au niveau computationnel, le modèle “1M” est le plus simple, le modèle “4M” est le plus compliqué car il nécessite le calcul de 4 cartes de saillance par image et par sujet et donc prend un temps de calcul important.

4 Expérience

Nous avons mené ici une expérience similaire à celle décrite au chapitre 4.

4.1 Stimuli

Les stimuli correspondent à 37 images de scènes naturelles en niveau de gris appartenant à plusieurs catégories : “paysage”, “objet”, “personne” et “habitacle de voiture”. Les images sont de taille de 768 × 1024 pixels (Fig. 5.9).

Un stimulus est présenté sur l'écran d'ordinateur 21” avec un taux de rafraîchissement de 75 Hz. La résolution de l'écran est de 768 × 1024 pixels et les images sont présentées au centre de l'écran.

4.2 Sujets

11 sujets ont participé à l'expérience. La plupart des sujets sont des étudiants en master ou en doctorat du laboratoire. Tous les sujets ont une acuité visuelle normale ou corrigée à la normale.

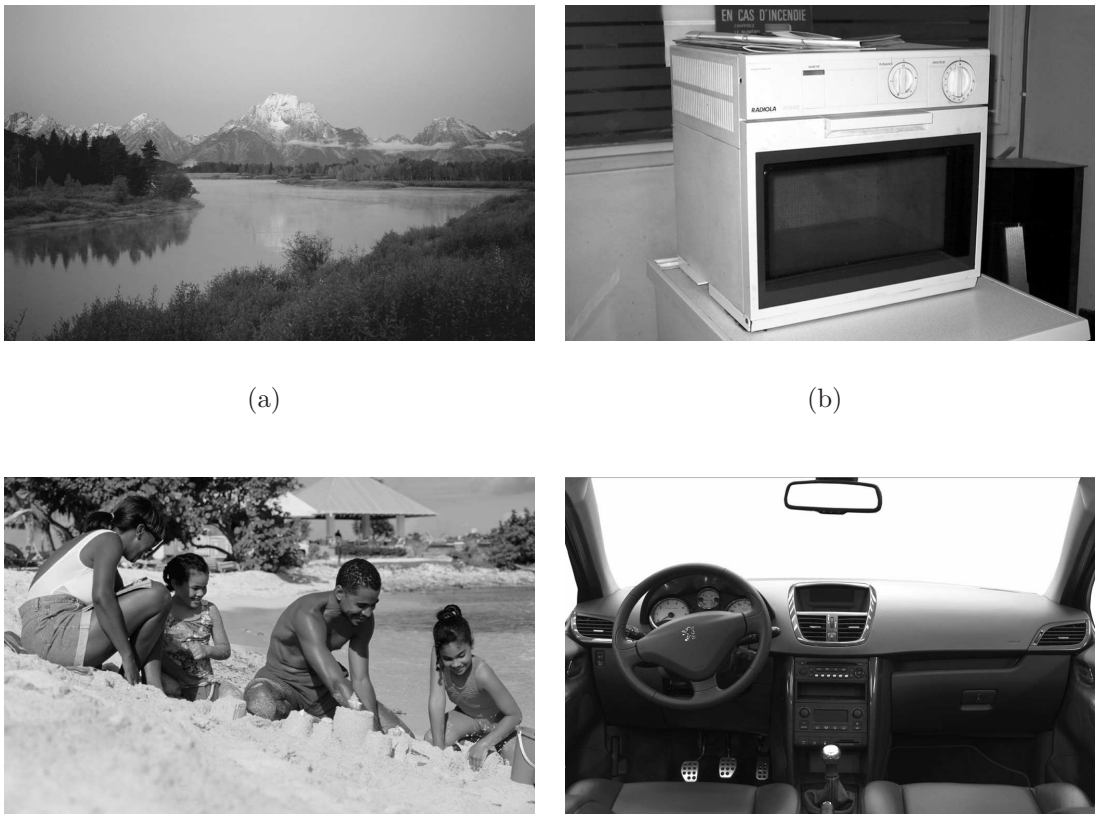
4.3 Dispositifs d'enregistrement - Eyelink

Les dispositifs sont les mêmes que ceux utilisés dans l'expérience au chapitre précédent.

4.4 Démarche

Les sujets sont assis à une distance de 57 cm devant l'écran sur lequel sont présentés les stimuli. L'angle visuel correspond à 30° × 40° degrés. Le menton du sujet est soutenu par une barre horizontale fixée et le casque de l'oculomètre est placé sur sa tête.

Chaque essai se déroule selon les étapes suivantes. D'abord, une cible de fixation apparaît au centre de l'écran où le sujet doit stabiliser son regard. Ensuite, une image



(a) (b)

(c) (d)

FIG. 5.9 – Exemple de 4 images en niveau de gris utilisées dans l'expérience oculométrique. (a) Paysage ; (b) Objet ; (c) Personne ; (d) Habitacle de voiture.

apparaît pendant 1.5 s. Elle est suivie par un écran en niveau de gris moyen pendant 1 s. Il faut noter que la cible de fixation est toujours au centre de l'écran. Par contre, l'ordre des images est aléatoire. Chaque sujet regarde les 37 images librement sans consigne. L'expérience dure environ 15 minutes pour un sujet.

4.5 Données expérimentales

Pour chaque couple "sujet × image", nous avons gardé les positions oculaires si le sujet avait bien stabilisé son regard sur la cible de fixation au centre de l'écran avant l'apparition de l'image (nous avons éliminé les données de 3 couples "sujet × image" sur 407 couples au total). Nous analysons les positions des fixations et les amplitudes des saccades pendant toute la durée où les images sont présentées.

5 Evaluation des trois modèles de programmation de saccade

Pour évaluer la qualité des trois modèles de programmation de saccade, il est nécessaire de les comparer aux données expérimentales. Deux critères de comparai-

son sont ici utilisés : la prédiction de fixation et la distribution des amplitudes de saccades.

5.1 Prédiction de fixation

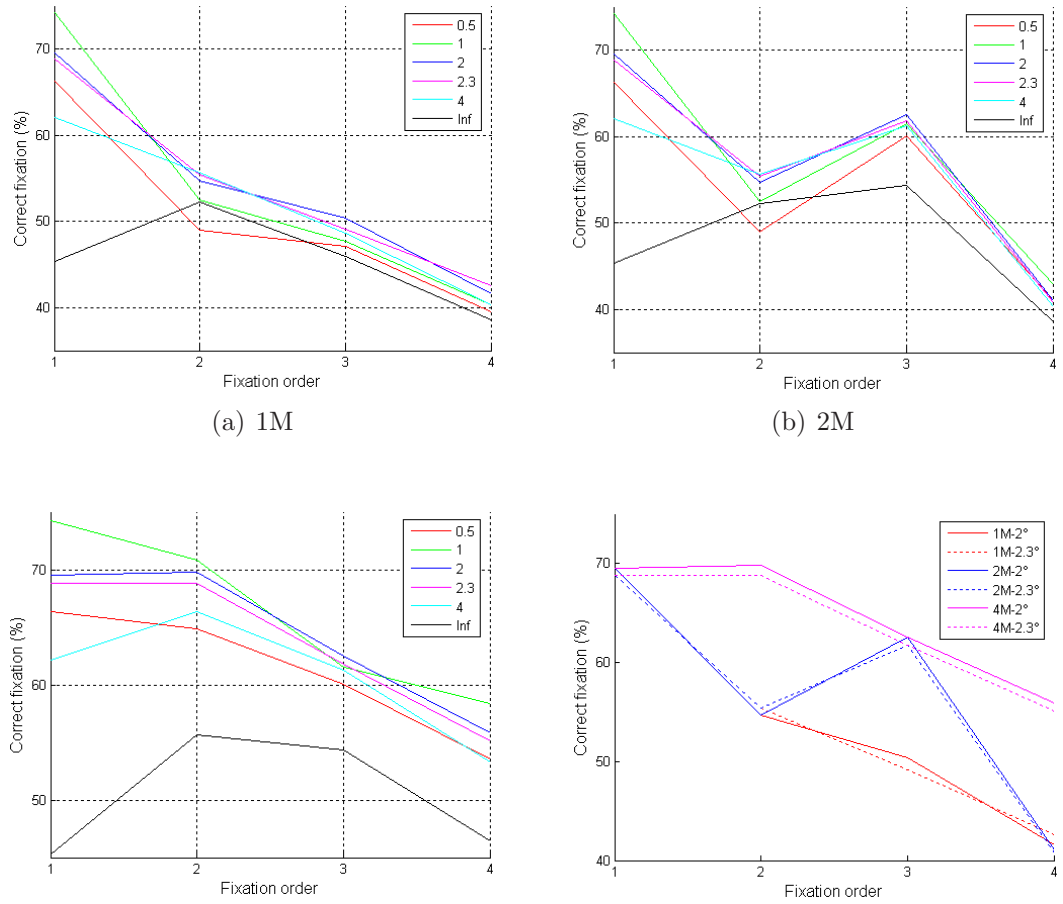
La qualité de prédiction de fixation est mesurée par le taux de fixations correctes [Torralba et al., 2006] présenté au chapitre 3. Ce critère permet d'évaluer le pourcentage de fixations localisées dans les zones saillantes extraites à partir de la carte de saillance. Comme nous utilisons trois modèles de programmation de saccade différents, nous obtenons trois jeux de fixations prédites différents. Pour chaque couple "image × sujet" dans le modèle "1M", une carte de saillance est utilisée pour mesurer le taux de fixations correctes pour 4 fixations. Dans le modèle "2M", la première carte de saillance permet de calculer le taux pour les 2 premières fixations, puis la deuxième carte pour les 2 fixations suivantes. De la même façon, chaque carte de saillance sert au calcul d'une fixation dans le modèle "4M". Nous remarquons que pour le modèle "1M", la carte de saillance d'une image est la même pour tous les sujets. En revanche, ce n'est plus le cas pour les deux autres modèles.

Pour ce critère, nous mesurons le taux de fixations correctes en fonction de l'ordre des fixations pour étudier la variation temporelle de la qualité de la carte de saillance. Pour chaque modèle, le taux de fixations correctes est calculé pour plusieurs valeurs du paramètre α, contrôlant la vitesse de décroissance de l'échelle du filtre passe-bas spatialement variant. Plus α est petit, plus l'image se floute rapidement avec l'excentricité. Afin d'étudier l'influence de ce paramètre sur les mouvements oculaires, nous le faisons varier dans un large éventail de 0.5° à l'infini. Cette dernière valeur correspond à une échelle uniforme, qui peut servir du modèle de base pour comparer à d'autres valeurs α. En total, nous avons testé 6 valeurs du paramètre α : 0.5°, 1°, 2°, 2.3°, 4° et Inf. La valeur 2.3° est celle optimale d'après la littérature [Geisler et al., 2006; Perry, 2002].

5.1.1 Comparaison entre les trois modèles

La figure 5.10a,b,c représente les taux de fixations correctes pour les trois modèles calculés sur les 4 premières fixations. Pour chaque modèle, les résultats sont montrés pour différentes valeurs du paramètres α. Le modèle "1M", qui utilise une seule carte de saillance pour prédire les 4 fixations, joue le rôle de modèle de base. Cette stratégie semble la moins efficace. En partageant avec "1M" le même résultat pour les deux premières fixations, le modèle "2M" rend un meilleur taux de fixations correctes pour la troisième fixation.

Le modèle "4M" améliore encore les résultats. La qualité de prédiction du modèle a nettement été améliorée à la deuxième, troisième et quatrième fixation par rapport au modèle "1M" et "2M". Ce modèle vise à calculer une nouvelle carte de saillance pour prédire chaque fixation. Ainsi, à partir d'un point de vue, seule la fixation suivante est programmée. Cette stratégie semble la plus efficace.



(a) 1M (b) 2M

(c) 4M (d) Trois modèles

FIG. 5.10 – Le taux de fixations correctes de différents modèles en fonction du rang de la fixation et pour différentes valeurs du paramètre α contrôlant la vitesse de décroissance de l'échelle : (a) modèle "1M" ; (b) modèle "2M" ; (c) modèle "4M" ; (d) Les trois modèles pour $\alpha = 2$ et $\alpha = 2.3$.

La différence de qualité entre les modèles est testée par le t-test (cf. chapitre 3). A titre d'exemple, pour $\alpha = 2$, nous observons une différence significative entre "4M" et "1M" (t-test, $p = 8.6985e-6$, $p = 4.8524e-4$ et $p = 5.9309e-5$ respectivement pour la 2ème, 3ème et 4ème fixation) ("2M" partage le résultat à la fois de "1M" et "4M" sauf la 4ème fixation ; pourtant à cette dernière, le résultat de "1M" et "2M" est presque le même). Cela montre que la réinitialisation de point de vue a significativement amélioré le résultat. Parmi ces modèles, "4M" a donné le meilleur résultat. La performance du modèle "2M" est intermédiaire de celle de "1M" et celle de "4M".

Le modèle "2M" correspond à une programmation de 2 saccades en parallèle comme proposé dans d'autres études sur les stimuli artificiels [Becker and Jürgens, 1979; Hooge and Erkelens, 1998; McPeek et al., 2000]. Néanmoins, la programmation en parallèle ne semble pas correspondre aux données expérimentales lors de l'exploration libre de scènes naturelles. Si la programmation de saccades en parallèle avait existé, le taux de fixations correctes aurait été meilleur pour le modèle "2M" par rapport au modèle "4M" à la deuxième et à la quatrième fixations. Pourtant, dans

notre expérience, la réinitialisation du point de vue après chaque saccade ou la programmation d'une seule saccade constitue la stratégie la plus efficace. Les meilleurs taux de fixations correctes sont obtenus pour le modèle "4M". En résumé, nous pouvons dire que lors de l'exploration d'une scène naturelle, les sujets planifient leurs saccades les unes après les autres, de manière séquentielle.

Une autre conclusion que nous pouvons tirer à partir des résultats est que la capacité de prédiction de fixation des cartes de saillance diminue au cours du temps. La diminution la plus forte est observée pour le modèle "1M". Cette diminution est plus lente pour le modèle "2M" et plus encore pour le modèle "4M" due à la réinitialisation du point de vue. Ce phénomène s'explique par l'effet des facteurs de haut niveau au cours du temps. Comme nous avons dit au chapitre précédent, bien que l'influence des facteurs de bas niveau ne change pas, celle des facteurs de haut niveau augmente fortement. Cela pénalise la qualité de prédiction de notre modèle qui se base principalement sur les facteurs de bas niveau. Néanmoins, notre modèle de saillance prédit bien des fixations, du moins, pour les premières fixations lors d'une exploration libre. Le taux de fixations correctes est toujours largement supérieur à la valeur du hasard qui est de 20%.

5.1.2 Influence du paramètre contrôlant la vitesse de décroissance de l'échelle

Les résultats de la figure 5.10 montrent l'effet du paramètre α contrôlant la vitesse de décroissance de l'échelle sur le taux de fixations correctes pour les trois modèles. La différence de taux de fixations correctes entre plusieurs valeurs de α dans le modèle "4M" semble la plus claire. D'abord, il existe une différence entre le modèle avec filtre passe-bas spatialement variant et celui qui ne l'a pas (α égal à l'infini). En utilisant le t-test, nous avons trouvé une différence significative entre le taux de fixations correctes pour $\alpha = Inf$ et celui pour chacune des autres valeurs de ce paramètre (sauf $\alpha = 0.5$ pour la troisième fixation et $\alpha = 4$ pour la quatrième). En tous les cas, en regardant la figure 5.10, nous pouvons voir un écart net entre le modèle sans filtre passe-bas spatialement variant et celui avec. Alors, l'implémentation de la variation de l'échelle des stimuli permet d'augmenter la qualité du modèle de programmation de saccade. Pourtant, entre les valeurs α de 0.5 à 4, il n'y a pas de différence significative sauf à la première fixation pour certains cas (t-test, $p = 0.0137$ entre $\alpha = 0.5$ et $\alpha = 1$; $p = 2.0610e - 004$ entre $\alpha = 1$ et $\alpha = 4$; $p = 0.0261$ entre $\alpha = 2$ et $\alpha = 4$; $p = 0.0458$ entre $\alpha = 2.3$ et $\alpha = 4$).

De plus, la figure 5.10 montre que le modèle pourrait obtenir le meilleur résultat pour les valeurs de α près de 1° ou 2°. La valeur biologique 2.3°, qui a été révélée dans la littérature [Geisler et al., 2006; Perry, 2002], donne elle aussi un bon taux de fixations correctes et très proche de celui de $\alpha = 2$ (Fig. 5.10d). Le modèle qui n'intègre pas le filtre passe-bas spatialement variant ($\alpha = Inf$) ou dans lequel la décroissance de l'échelle est trop rapide ($\alpha = 0.5$) ne prédit pas bien les saccades ou les fixations des sujets.

5.2 Distribution des amplitudes de saccades

La plupart des saccades ont de petites amplitudes [Bahill et al., 1975]. Dans [Parkhurst et al., 2002], à partir du modèle d'attention visuelle, il a trouvé une distribution des amplitudes de saccades similaire à celles expérimentales en appliquant simplement un masque gaussien à la carte de saillance. Nous voulons ici tester quel modèle de programmation de saccade permet d'obtenir une distribution des amplitudes de saccades similaire à celle obtenue sur des données comportementales.

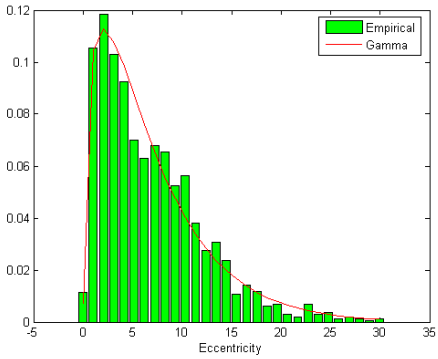
Nous traçons la distribution des amplitudes de saccades obtenues à partir des données expérimentales et nous la considérons comme la distribution de référence. Ensuite, la distribution des amplitudes de saccades prédites par nos modèles est examinée pour plusieurs valeurs du paramètre α contrôlant la vitesse de décroissance de l'échelle du modèle de filtrage passe-bas spatialement variant de Perry. Dans un premier temps, la comparaison de la distribution des amplitudes de saccades prédites avec la distribution de référence est effectuée qualitativement. Dans un deuxième temps, nous présentons une comparaison quantitative entre des distributions basée sur une modélisation des distributions par la loi Gamma (cf. annexe F2).

Nous choisissons d'étudier la distribution des amplitudes de saccades prédites par le modèle "4M" qui est le plus efficace pour prédire les fixations et donc les saccades des sujets lors d'une exploration libre de scènes naturelles. En faisant varier le paramètre α contrôlant la vitesse de décroissance de l'échelle, nous voulons observer une influence sur la distribution des amplitudes de saccades. De plus, alors que le critère de taux de fixations correctes n'a pas précisé exactement la meilleure valeur de α, nous espérons que ce deuxième critère permette de trouver une valeur proche de 2.3°.

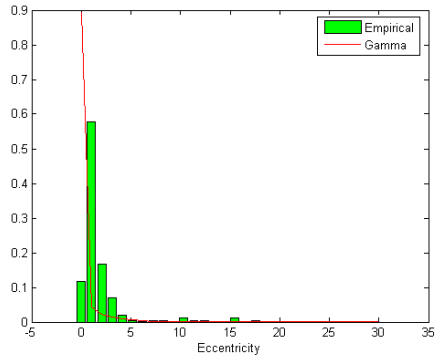
5.2.1 Comparaison qualitative

A partir de la figure 5.11a, nous pouvons voir qu'une grande partie des saccades expérimentales ont de petites amplitudes inférieures à 15° comme révélé par Bahill [Bahill et al., 1975]. Sur cette distribution, il existe deux propriétés importantes : la position du mode et l'étendue.

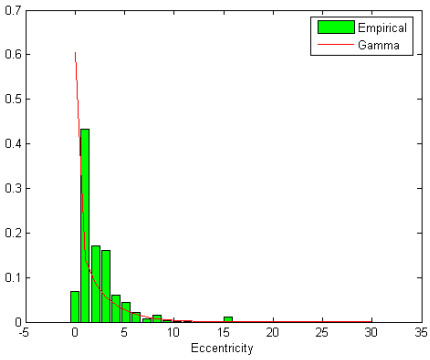
En même temps, en regardant les distributions des amplitudes de saccades prédites pour différentes valeurs de α allant de 0.5 à l'infini, nous avons bien observé la variation de la distribution des amplitudes de saccades, à la fois au niveau de la position du mode et au niveau de l'étendue. Pour α petit, la distribution des amplitudes de saccades se concentre sur les petites saccades et son étendue est petite. Lorsque α augmente, l'étendue de la distribution augmente et le mode se déplace à droite. Quand α tend vers l'infini, c'est-à-dire qu'il n'y a plus de filtrage passe-bas spatialement variant, la distribution semble devenir uniforme, ce qui a été observé par Parkhurst [Parkhurst et al., 2002]. En comparant avec la forme de la distribution des amplitudes de saccades expérimentales, les distributions modélisées pour $\alpha = 2$ ou $\alpha = 2.3$ donnent les meilleurs résultats en prenant le compromis entre la position du mode et l'étendue.



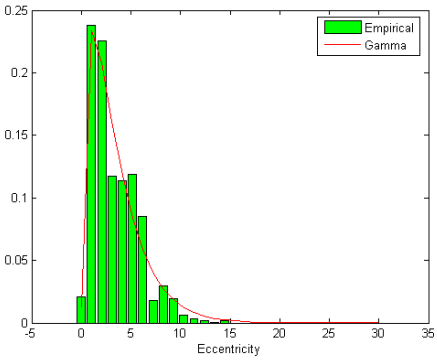
(a) Données expérimentales



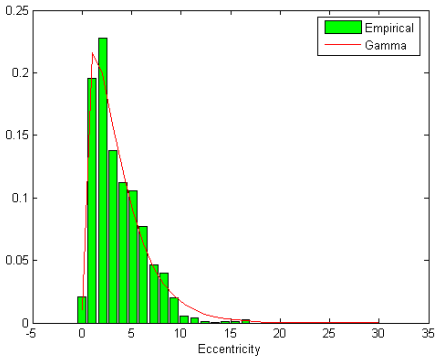
(b) $\alpha = 0.5°$



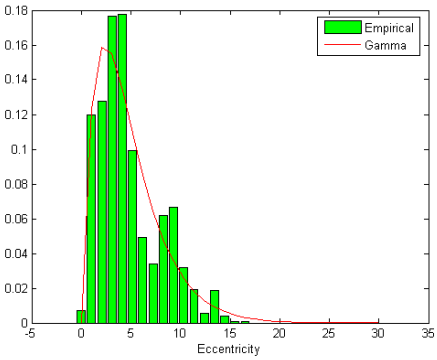
(c) $\alpha = 1°$



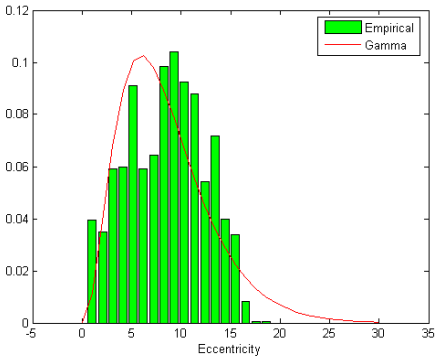
(d) $\alpha = 2°$



(e) $\alpha = 2.3°$



(f) $\alpha = 4°$



(g) $\alpha = Inf$

FIG. 5.11 – Distributions empiriques des amplitudes de saccades et leur modélisation par la distribution Gamma (voir texte).

5.2.2 Comparaison quantitative

Nous pouvons comparer quantitativement les distributions des amplitudes de saccades prédites et la distribution des amplitudes de saccades expérimentales en les modélisant par la distribution Gamma. Chaque distribution Gamma est caractérisée par deux paramètres : l'allure k et l'échelle θ. Ces valeurs sont données dans le tableau 5.1.

En s'appuyant sur le paramètre d'allure k, nous pouvons dire que la distribution des amplitudes de saccades prédites pour α égal à 2° ou 2.3° a une allure très proche de celle de la distribution expérimentale. Néanmoins, pour les paramètres θ des distributions obtenues à partir des modèles, ils sont proches entre eux mais éloignés de celui de la distribution expérimentale. Cela s'explique par une grande étendue de la distribution expérimentale par rapport aux étendues des distributions modélisées. En effet, la distribution expérimentale a plus de longues saccades que celle obtenue à partir des modèles de programmation de saccade. Ce phénomène est lié à la manière dont les fixations sont prédites. Pour être moins coûteux, le champ visuel est fixé égal à la taille de l'image (768 × 1024 pixels). Ainsi, quand le point de vue se déplace, le champ visuel change mais garde toujours la même taille. D'ailleurs, le point de vue actuel est au centre de champ visuel. Par conséquent, l'amplitude de saccade prédite, qui est la distance entre le point de vue et la fixation prédite, ne peut pas dépasser la moitié de la diagonale de l'image. Quant aux données expérimentales, la saccade maximale peut théoriquement atteindre à la taille de la diagonale de l'image (50°).

Tab. 5.1 – Paramètres estimés de la distribution Gamma

α	k	θ
Expérience	**1.4146**	**4.8707**
0.5	06150	2.9614
1	0.8422	2.7956
2	1.4822	2.2969
2.3	1.4917	2.4445
4	1.9904	2.4136
Inf	3.3740	2.5108

6 Conclusion

Dans ce chapitre, nous avons étudié trois modèles de programmation de saccade. Ces modèles diffèrent sur le nombre de fixations prédites à partir d'un point de vue. Le modèle "1M" est le plus simple ; il permet de prédire 4 fixations depuis un point de vue. Le modèle "2M" (respectivement "4M") prédit 2 (respectivement 1) fixations. Tous ces modèles prédisent les 4 premières fixations (ou saccades) lors de l'exploration libre d'une scène naturelle. Ces fixations correspondent à une courte

durée au début de la perception visuelle dans laquelle les facteurs de bas niveau dominent le traitement visuel. Ce choix est ainsi compatible avec le modèle de programmation de saccade qui se base sur la carte de saillance calculée à partir de facteurs de bas niveau.

Les résultats ont montré que le modèle "4M" est le plus efficace. Le modèle "1M" a présenté le résultat le moins efficace et "2M" se situe entre ces deux modèles. A partir d'un point de vue, il semble, dans l'exploration libre de scènes naturelles, que les sujets programment une seule saccade. Ensuite, le point de vue se déplace et la saccade suivante sera programmée.

Dans notre étude, la programmation en parallèle, qui a été proposée dans la littérature [Becker and Jürgens, 1979; Hooge and Erkelens, 1998; McPeek et al., 2000], semble moins efficace que la programmation d'une seule saccade à partir d'un point de vue. Ce phénomène peut être expliqué par deux raisons. La première concerne la tâche. La programmation de saccades en parallèle a auparavant été étudiée dans le contexte de la recherche visuelle. En revanche, dans notre expérience, il n'y avait aucune tâche spécifique. La deuxième raison peut être liée aux stimuli utilisés. Dans les études précédentes, les stimuli sont artificiels et très simples (une cible et des distracteurs). Ainsi, un petit nombre de zones saillantes dans l'image peut faciliter la programmation de saccades en parallèle. Quant aux scènes naturelles utilisées dans notre expérience, elles sont bien plus complexes et contiennent plus de zones saillantes. Il est alors plus difficile de programmer plus qu'une saccade en même temps.

Nous avons également montré le rôle de la variation de l'échelle effectuée par le filtrage passe-bas spatialement variant dans la programmation de saccade. L'implémentation du filtre passe-bas spatialement variant a donné une meilleure prédiction de fixation par rapport au modèle avec une échelle constante. De plus, la variation de l'échelle a influencé la distribution des amplitudes de saccades. En faisant varier le paramètre contrôlant la vitesse de décroissance de l'échelle du filtre passe-bas spatialement variant, nous avons bien observé une variation de la forme de la distribution des amplitudes de saccades prédites. Les résultats obtenus du filtre passe-bas spatialement variant dans ce chapitre supportent la conclusion que ce filtre devrait être implémenté dans les modèles de perception visuelle et plus particulièrement dans les modèles d'attention visuelle.

Chapitre 6

Traitement spatio-temporel de la rétine et filtrage spatialement variant

1 Introduction

L'objectif de ce chapitre est de présenter deux évolutions qui nous semblent importantes à donner à notre modèle d'attention visuelle. Ces deux évolutions concernent d'une part la prise en compte des caractéristiques temporelles de la rétine et d'autre part l'impact au niveau du filtrage cortical, de la densité non homogène des photorécepteurs et des cellules ganglionnaires.

Ainsi jusqu'à présent, nous avons construit le modèle de saillance en utilisant les caractéristiques statiques du modèle fonctionnel de la rétine. Or le processus de scrutation des scènes naturelles induit, par les mouvements oculaires, un flux visuel dynamique sur les photorécepteurs de la rétine. Ce flux est caractérisé par des séquences alternant la stabilisation du flux visuel durant les fixations et le déplacement rapide durant les saccades. Nous allons donc décrire le modèle spatio-temporel et le simuler dans le contexte des mouvements oculaires. Cette dynamique spécifique va nous permettre de tester le modèle dans des conditions d'application différentes de celles plus usuelles des scènes en mouvement [Torralba and Hérault, 1997; Torralba, 1999]. De plus en analysant les sorties des voies rétiniennes selon leur dynamique spatio-temporelle durant les fixations et les saccades, nous allons pouvoir confronter le modèle relativement aux résultats connus sur l'inhibition de l'activité visuelle durant les saccades.

Le deuxième point abordé durant ce chapitre concerne la densité non uniforme des cellules ganglionnaires à la surface de la rétine. Comme nous l'avons vu au chapitre précédent, cette densité est maximale à la fovéa et diminue rapidement lorsque l'on va à la périphérie de la rétine. Ainsi, sur une image, la zone autour du point de vue est perçue avec la résolution la plus importante tandis que la périphérie de ce point est perçue avec une résolution plus faible. Ce phénomène avait été modélisé par un filtrage passe-bas spatialement variant dont la fréquence de coupure diminue du point de vue à la périphérie pour obtenir en sortie une image de taille égale, mais

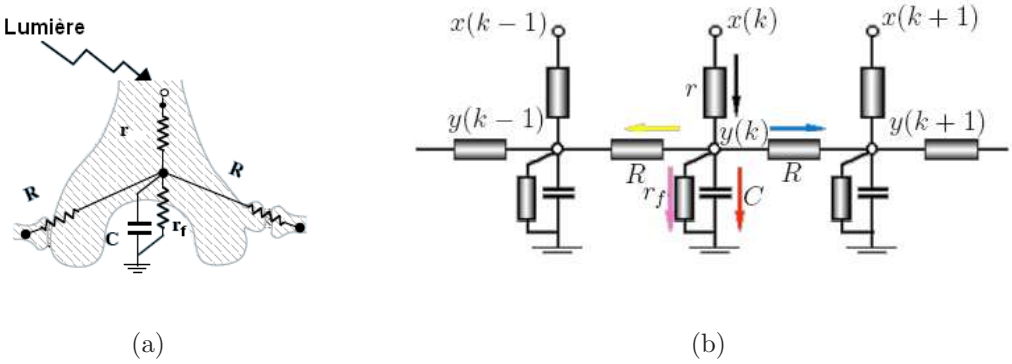
dont l'échelle diminue avec l'excentricité. Nous allons donc maintenant implémenter la résolution spatialement variante par un taux de sous-échantillonnage également spatialement variant, et surtout, en poursuivant cette modélisation jusqu'au niveau cortical. Il est intéressant d'étudier l'impact de ce sous-échantillonnage sur le banc de filtres LogNormaux modélisant les cellules corticales.

2 Traitement spatio-temporel de la rétine

2.1 Filtrage passe-bas spatio-temporel

Pour la rétine statique, tous les filtrages sont modélisés à partir d'un seul type, le filtrage passe-bas, comme module de lissage mais aussi de réhaussement du contraste grâce à un filtrage passe-haut réalisé par différence de deux filtres passe-bas. Dans le cas du modèle dynamique, le filtrage passe-bas est spatio-temporel. Nous allons donc d'abord décrire ce module "générique" puis ensuite le positionner aux différents niveaux de la rétine.

Ce filtrage se base sur le modèle électrique du réseau de photorécepteurs représenté à la figure 6.1b [Beaudot, 1994; Hérault, 2001]. Selon ce modèle, un photorécepteur est modélisé par des composantes passives : deux résistances R pour les jonctions électriques avec les photorécepteurs voisins, une capacité membranaire C et une résistance de fuite r_f pour les caractéristiques membranaires du photorécepteur, et enfin une résistance cytoplasmique r (Fig. 6.1a).



(a) (b)

FIG. 6.1 – Modèle électrique des photorécepteurs : (a) Modèle d'un photorécepteur ; (b) Schéma électrique représentant plusieurs photorécepteurs connectés [Hérault, 2001].

A partir du schéma de la figure 6.1b, la relation entre la sortie $y(k)$ et l'entrée $x(k)$ d'un photorécepteur à un instant t est donnée par l'équation suivante :

$$\frac{x(k,t)-y(k,t)}{r} = \frac{y(k,t)-y(k-1,t)}{R} + \frac{y(k,t)-y(k+1,t)}{R} + \frac{y(k,t)}{r_f} + C\frac{dy(k,t)}{dt} \tag{6.1}$$

avec k, la position dans le domaine spatial et t, le temps.

En posant $\alpha = \frac{r}{R}$, $\beta = \frac{r}{r_f}$ et $\tau = rC$, on a :

$$x(k,t) = (1 + 2\alpha + \beta)y(k,t) - \alpha y(k-1,t) - \alpha y(k+1,t) + \tau \frac{dy(k,t)}{dt}$$

En prenant la transformée Z pour la variable spatiale discrète et la transformée de Fourier pour la variable temporelle continue, on obtient :

$$X(z, f_t) = (1 + 2\alpha + \beta)Y(z, f_t) - \alpha(z + z^{-1})Y(z, f_t) + \tau j 2\pi f_t Y(z, f_t)$$

On obtient ainsi la fonction de transfert $H(z, f_t)$:

$$H(z, f_t) = \frac{Y(z, f_t)}{X(z, f_t)} = \frac{1}{1 + \beta + \alpha(2 - z - z^{-1}) + \tau j 2\pi f_t} \tag{6.2}$$

Dans le domaine des fréquences spatiales, en remplaçant z par $e^{j2\pi f_s}$ la fonction de transfert devient :

$$H(f_s, f_t) = \frac{Y(f_s, f_t)}{X(f_s, f_t)} = \frac{1}{1 + \beta + 4\alpha \sin(\pi f_s)^2 + \tau j 2\pi f_t} \tag{6.3}$$

avec f_s la fréquence spatiale et f_t la fréquence temporelle.

Cette fonction de transfert correspond à un filtre passe-bas spatio-temporel que nous désignons H_{BF} dans la suite. Le paramètre β détermine le gain du filtre aux fréquences nulles en spatial et en temporel. Le paramètre α détermine l'influence de la fréquence spatiale sur le filtre et le paramètre τ est la constante de temps du filtre temporel. La figure 6.2a illustre un exemple de la fonction de transfert du filtre passe-bas spatio-temporel.

Cette fonction de transfert $H_{BF}(f_s, f_t)$ peut être réécrite sous la forme :

$$\begin{aligned} H_{BF}(f_s, f_t) &= \frac{\frac{1}{\tau}}{\frac{1+\beta+4\alpha \sin(\pi f_s)^2}{\tau} + j 2\pi f_t} \\ &= \frac{\frac{1}{\tau}}{\frac{1}{T} + j 2\pi f_t} \end{aligned} \tag{6.4}$$

où $T = \frac{\tau}{1 + \beta + 4\alpha \sin(\pi f_s)^2}$ est la nouvelle constante de temps. Il est important de noter qu'elle varie en fonction de la fréquence spatiale. Plus la fréquence spatiale est importante, plus la constante de temps T est faible, et *vice versa*. Cela implique une inséparabilité des variables spatiale et temporelle du filtre passe-bas spatio-temporel.

La réponse impulsionnelle calculée par la transformée de Fourier inverse pour une fréquence spatiale f_s est :

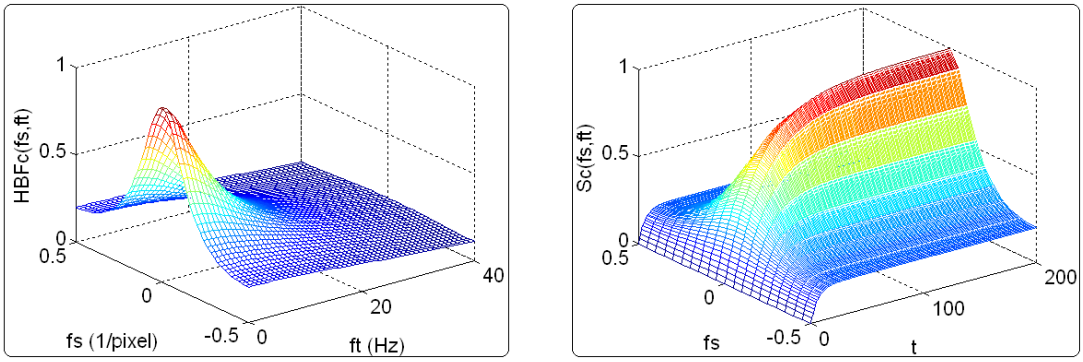
$$h_{BF}(f_s, t) = \frac{1}{\tau} e^{-\frac{t}{T}} \cdot u(t) \tag{6.5}$$

avec $u(t)$ une fonction échelon.

Avec une entrée "échelon" $e(k,t) = \delta(k) \cdot u(t)$, alors $e(f_s,t) = u(t)$ représentant un Dirac dans le domaine spatial et une fonction échelon dans le temps, la sortie du filtre passe-bas spatio-temporel $s(f_s,t)$ est[1] :

$$\begin{aligned} s(f_s,t) &= h_{BF}(f_s,t) * e(f_s,t) \\ s(f_s,t) &= \frac{T}{\tau}\left(1 - \exp(-\frac{t}{T})\right) u(t) \end{aligned} \tag{6.6}$$

La figure 6.2b illustre la réponse du filtre pour une entrée "échelon" $e(f_s,t) = u(t)$. Nous remarquons qu'il faut un certain temps dépendant de la constante de temps T pour que la réponse soit stable.



(a) (b)

FIG. 6.2 – Modélisation du filtre passe-bas spatio-temporel : (a) Fonction de transfert ; (b) Sa réponse au stimulus "échelon". Les paramètres sont $\alpha = 1$, $\beta = 0$ et $\tau = 30$.

2.2 Modèle spatio-temporel de la rétine

2.2.1 Modélisation

Notre modélisation du traitement rétinien s'effectue selon le schéma de la figure 6.3. En effet, ce schéma est une extension du modèle fonctionnel de la rétine statique (Fig. 2.5) décrit au chapitre 2 en ajoutant la voie Magno qui fait partie de la rétine dynamique. Les traitements dans ce schéma s'appuient majoritairement sur le filtrage linéaire des cellules rétiniennes. En effet, l'étape non linéaire d'adaptation à la luminance (cf. chapitre 2) qui intervient dans le fonctionnement des

[1] $\mathcal{F}_x\left\{f(x,y) \underbrace{*}_{(x,y)} g(x,y)\right\} = \mathcal{F}_x\{f(x,y)\} \underbrace{*}_{(y)} \mathcal{F}_x\{g(x,y)\}$ avec $\mathcal{F}_x$ la transformée de Fourier pour la variable x et $\underbrace{*}_{(x)}$ convolution pour la variable x. Dans la suite, sans précision supplémentaire, la convolution est effectuée sur la variable t.

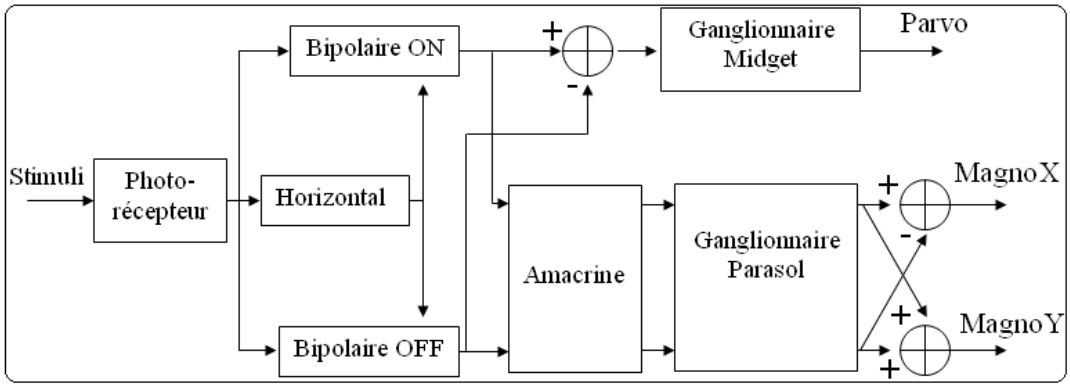


FIG. 6.3 – Modèle spatio-temporel de la rétine.

photorécepteurs et des cellules bipolaires ne modifie pas le principe du traitement spatio-temporel de la rétine. Ainsi, nous passons cette étape non-linéaire sous silence dans les modélisations suivantes. Néanmoins, l'adaptation à la luminance sera intégrée dans le simulateur de la rétine (cf. section §2.2.2).

Dans la suite, nous allons modéliser les traitements des différentes étapes du schéma de la figure 6.3.

Les photorécepteurs

Le fonctionnement des photorécepteurs est modélisé par le filtre passe-bas spatio-temporel présenté ci-dessus et appelé H_{BFc} avec les paramètres τ_c et T_c incluant les deux paramètres α_c et β_c. En reprenant l'équation 6.4, la fonction de transfert des photorécepteurs est alors :

$$H_{BFc}(f_s, f_t) = \frac{\frac{1}{\tau_c}}{\frac{1}{T_c} + j2\pi f_t}, \tag{6.7}$$

et la réponse impulsionnelle pour une fréquence spatiale f_s donnée est :

$$h_{BFc}(f_s, t) = \frac{1}{\tau_c} e^{-\frac{t}{T_c}} \cdot u(t) \tag{6.8}$$

Les cellules horizontales

Le fonctionnement des cellules horizontales est modélisé par le même filtrage passe-bas à partir de la sortie des photorécepteurs. Comme les photorécepteurs, la fonction de transfert H_{BFh} des cellules horizontales est caractérisée par les paramètres τ_h et T_h (contenant les deux autres paramètres α_h et β_h) :

$$H_{BFh}(f_s, f_t) = \frac{\frac{1}{\tau_h}}{\frac{1}{T_h} + j2\pi f_t} \tag{6.9}$$

et la réponse impulsionnelle pour une fréquence spatiale f_s donnée est :

$$h_{BFh}(f_s, t) = \frac{1}{\tau_c} e^{-\frac{t}{T_c}} \cdot u(t) \tag{6.10}$$

Nous notons que puisque les cellules horizontales fonctionnent comme un filtrage passe-bas de la sortie des photorécepteurs, qui sont aussi un filtre passe-bas, la sortie des cellules horizontales est encore plus basse fréquence.

Les cellules bipolaires et la voie Parvo

En absence de l'étape d'adaptation à la luminance, les cellules bipolaires sont modélisées par la différence entre les photorécepteurs et les cellules horizontales. Les cellules bipolaires sont en entrée des cellules ganglionnaires midgets qui donnent en sortie la voie Parvo. Dans le modèle, on considère les cellules ganglionnaires midgets comme un relais avec une fonction de transfert constante unitaire. Alors que les réponses des cellules bipolaires sont identiques à celles de la voie Parvo, nous allons présenter par la suite uniquement la voie Parvo. Ainsi, la fonction de transfert $H_P(f_s, f_t)$ de la voie Parvo est représentée par l'équation suivante :

$$\begin{aligned} H_B(f_s, f_t) &= H_{BFc}(f_s, f_t) - H_{BFc}(f_s, f_t) \cdot H_{BFh}(f_s, f_t) \\ &= H_{BFc}(f_s, f_t) \cdot (1 - H_{BFh}(f_s, f_t)) \\ &= \frac{\frac{1}{\tau_c T_h} - \frac{1}{\tau_c \tau_h} + j\frac{2\pi}{\tau_c} f_t}{\left(\frac{1}{T_c} + j2\pi f_t\right)\left(\frac{1}{T_h} + j2\pi f_t\right)} \end{aligned} \tag{6.11}$$

La figure 6.4a représente la fonction de transfert de la voie Parvo dans le domaine des fréquences spatiales et temporelles ; elle correspond à un filtre passe-bande. Comme pour la rétine statique, le filtre passe-bande est construit en se basant uniquement sur le filtre passe-bas. De plus, la voie Parvo véhicule des informations hautes fréquences spatiales à la fréquence temporelle nulle. Pour les stimuli haute fréquence temporelle (i.e., variation rapide dans le temps), la voie Parvo transmet des informations basses fréquences spatiales.

Si l'on veut étudier la fonction de transfert de la voie Parvo selon la variable temporelle, cette fonction est réécrite en la mettant sous la forme suivante :

$$H_P(f_s, f_t) = \frac{A_1}{\frac{1}{T_c} + j2\pi f_t} + \frac{A_2}{\frac{1}{T_h} + j2\pi f_t} \tag{6.12}$$

avec

$$A_1 = \frac{1}{\tau_c} - \frac{T_c T_h}{\tau_c \tau_h (T_c - T_h)} \text{ et } A_2 = \frac{T_c T_h}{\tau_c \tau_h (T_c - T_h)}.$$

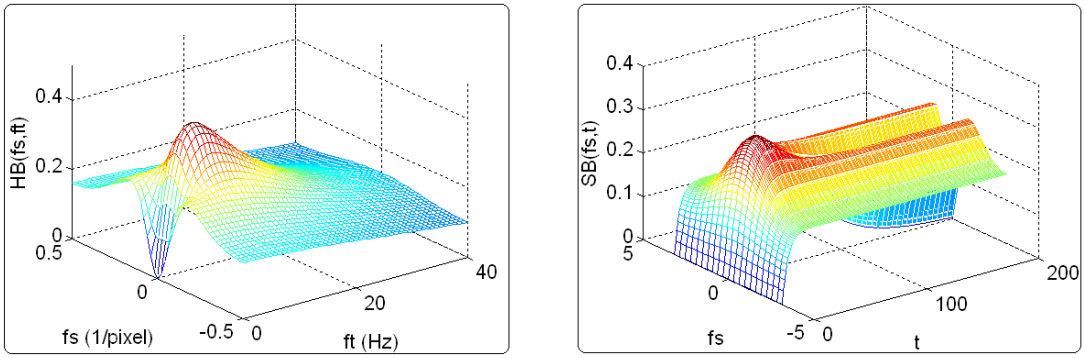
Les valeur de A_1, A_2, T_c, T_h dépendent de la fréquence spatiale. Ainsi, la réponse impulsionnelle de la voie Parvo est :

$$h_P(f_s, t) = A_1 \exp\left(-\frac{t}{T_c}\right) u(t) + A_2 \exp\left(-\frac{t}{T_h}\right) u(t) \tag{6.13}$$

Pour une entrée "échelon" $e(f_s, t) = u(t)$ représentant un Dirac dans le domaine spatial et une fonction échelon dans le temps, la réponse $s_P(f_s, t)$ de la voie Parvo est la suivante :

$$\begin{aligned} s_B(f_s, t) &= h_B(f_s, t) * e(f_s, t) \\ s_B(f_s, t) &= A_1 T_c \left(1 - \exp(-\frac{t}{T_c})\right) u(t) + A_2 T_h \left(1 - \exp(-\frac{t}{T_h})\right) u(t) \end{aligned}$$

Un exemple de la réponse spatio-temporelle de la voie Parvo pour une entrée "échelon" est illustré à la figure 6.4b. Au début, elles fonctionnent comme un filtre passe-bas spatial. Puis au cours du temps, elles fonctionnent comme un filtre passe-bande en transmettant les détails des stimuli.



(a) (b)

FIG. 6.4 – Modélisation de la voie Parvo (ou des cellules bipolaires) : (a) Fonction de transfert ; (b) Sa réponse au stimulus "échelon". Les paramètres pour les photorécepteurs sont $\alpha_c = 1$, $\beta_c = 0$ et $\tau_c = 30$. Pour les cellules horizontales, ils sont $\alpha_h = 1$, $\beta_h = 0$ et $\tau_h = 22$.

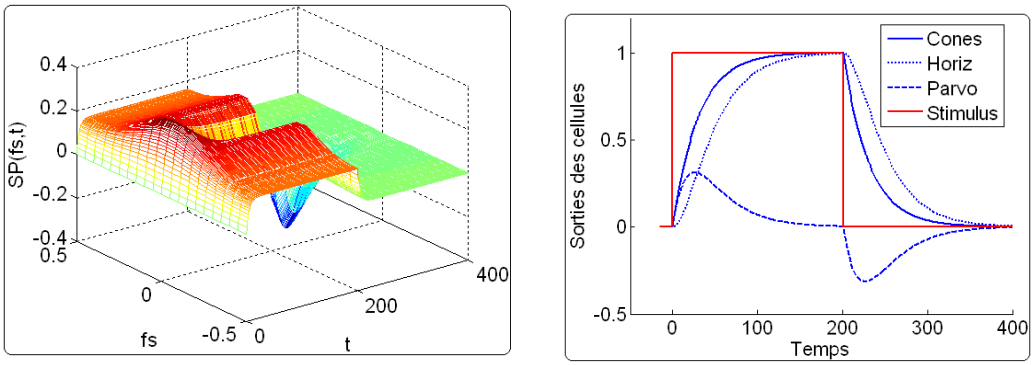
Pour une entrée "porte" représentant un Dirac dans le domaine spatial et une fonction "porte" dans le temps, lorsque le stimulus disparaît (i.e., le stimulus tombe de 1 à zéro), la sortie $s_P(f_s, t)$ est la suivante :

$$s_P(f_s, t) = aT_c \exp(-\frac{t}{T_c})u(t) + bT_h \exp(-\frac{t}{T_h})u(t)$$

La figure 6.5a illustre la voie Parvo pour un stimulus suivant une fonction "porte" temporelle simulant la présentation d'une image statique durant un intervalle de temps, puis sa disparition. Lorsque le stimulus apparaît, la voie Parvo prend l'information basse fréquence spatiale. Quand la voie Parvo est stable, elle représente l'information passe-bande du stimulus. Autrement dit, lors d'une exploration de scène, nous percevons d'abord l'information globale de la scène, et puis au cours du temps, les détails. Ce processus nous permet d'obtenir d'une même scène l'information visuelle à différentes échelles spatiales selon le temps. Cela correspond au phénomène *"coarse-to-fine"* qui a été observé dans la littérature. Lorsque le stimulus disparaît, la voie Parvo change son signe avant de revenir à zéro. La figure 6.5b montre l'évolution de la voie Parvo à la fréquence spatiale nulle. Alors que la voie Parvo extrait l'information passe-bande, sa réponse à la fréquence spatiale nulle tombe à zéro quand elle est stable. En revanche, la voie Parvo à la fréquence spatiale non nulle est constante et non nulle au régime permanent (Fig. 6.6d).

La voie Magno

La voie Magno, correspondant aux cellules ganglionnaires parasols, peut en fait



(a) (b)

FIG. 6.5 – Réponse de la voie Parvo au stimulus "porte" : (a) Réponse de la voie Parvo en fonction de la fréquence spatiale et du temps ; (b) Réponses des photorécepteurs, des cellules horizontales et de la voie parvocellulaire à la fréquence spatiale nulle. Les paramètres sont $\alpha_c = 1$, $\beta_c = 0$, $\tau_c = 30$ pour les photorécepteurs et $\alpha_h = 1$, $\beta_h = 0$, $\tau_h = 22$ pour les cellules horizontales.

se séparer en deux "sous voies" : la voie Magno X et la voie Magno Y. Ces deux voies magnocellulaires se distinguent par la manière de combiner les signaux dans les champs récepteurs des cellules : une combinaison linéaire pour la voie Magno X et une combinaison non-linéaire pour la voie Magno Y [Shapley et al., 1981]. Ainsi, c'est la voie Magno X qui va être modélisée par le filtrage linéaire. La voie Magno Y peut pourtant être expliquée à partir de la voie Magno X car ces deux voies partagent presque les mêmes étapes dans leur traitement.

La voie Magno X est obtenue à partir de la voie Parvo (ou la sortie des cellules bipolaires) après être passée par les cellules amacrines et les cellules ganglionnaires parasols. Le fonctionnement des cellules ganglionnaires parasols est modélisé par un filtrage passe-bas spatio-temporel comme les photorécepteurs ou les cellules horizontales. Quant aux cellules amacrines, elles fonctionnent comme un filtre passe-haut temporel. Ainsi, la voie Magno X est modélisée par le filtrage passe-bas spatio-temporel de la sortie de la voie Parvo, suivi par un filtrage passe-haut temporel. La fonction de transfert de la voie Magno X est alors :

$$H_{Mx}(f_s, f_t) = H_P(f_s, t) \cdot H_{BFg}(f_s, f_t) \cdot H_{HF}(f_t) \tag{6.14}$$

où $H_{BFg}(f_s, f_t) = \frac{\frac{1}{\tau_g}}{\frac{1}{T_g} + j2\pi f_t}$ est la fonction de transfert des cellules ganglionnaires midgets représentant un filtre passe-bas spatio-temporel avec les paramètres τ_g et T_g (comportant les deux autres paramètres α_g et β_g),
et $H_{HF}(f_t) = \frac{j2\pi\tau_a f_t}{1+j2\pi\tau_a f_t}$ est la fonction de transfert des cellules amacrines représentant un filtre passe-haut temporel, caractérisé par la constante de temps τ_a.

La figure 6.6a illustre la fonction de transfert de la voie Magno X. Cette dernière ne répond qu'aux stimuli variants dans le temps en véhiculant leurs informations en basses fréquences spatiales.

La réponse impulsionnelle de la voie Magno X pour une fréquence spatiale f_s donnée est alors :

$$h_{Mx}(f_s,t) = C_1 \exp(-\frac{t}{T_g})u(t) + C_2 \exp(-\frac{t}{T_c})u(t) + C_3 \exp(-\frac{t}{T_h})u(t) + C_4 \exp(-\frac{t}{\tau_a})u(t) \tag{6.15}$$

avec

$$\begin{aligned} C_1 &= -\frac{\tau_a T_g}{\tau_g(T_g-\tau_a)}\left(\frac{A_1 T_c}{T_g-T_c} + \frac{A_2 T_h}{T_g-T_h}\right) \\ C_2 &= \frac{\tau_a A_1 T_c T_g}{\tau_g(T_g-\tau_a)}\left(\frac{1}{T_g-T_c} - \frac{1}{\tau_a-T_c}\right) \\ C_3 &= \frac{\tau_a A_2 T_h T_g}{\tau_g(T_g-\tau_a)}\left(\frac{1}{T_g-T_h} - \frac{1}{\tau_a-T_h}\right) \\ C_4 &= \frac{\tau_a T_g}{\tau_g(T_g-\tau_a)}\left(\frac{A_1 T_c}{\tau_a-T_c} + \frac{A_2 T_h}{\tau_a-T_h}\right) \end{aligned}$$

Les valeurs de C_1, C_2, C_3, C_4 dépendent de la fréquence spatiale.

La figure 6.6b illustre la voie Magno X pour un même stimulus de type fonction "porte" temporelle. D'une part, la voie Magno X représente la partie basse fréquence spatiale de la voie Parvo. D'autre part, elle reflète la variation temporelle de cette dernière. Il est à noter que l'amplitude de la voie Magno X est plus faible que celle de la voie Parvo. De plus, nous observons un passage par zéro de la voie Magno X, notamment à la fréquence spatiale nulle (Fig. 6.6c). Pour les hautes fréquences spatiales, la voie Magno X tend plus rapidement vers le régime permanent (Fig. 6.6d). Comme la voie Magno X répond à la variation temporelle, leur réponse devient nulle lorsque la voie Parvo est stable. Lorsque le stimulus disparaît, la voie Magno X change son signe comme la voie Parvo avant de revenir à zéro.

La voie Magno Y est soumise aux mêmes traitements que la voie Magno X sauf la dernière étape. La somme des voies ON et OFF pour construire la voie Magno Y est une étape non-linéaire. Ainsi, le fonctionnement de la voie Magno Y est représenté par une fonction "valeur absolue" des cellules bipolaires ou de la voie Parvo [Hérault, 2001]. La voie Magno Y simule donc la variance locale des stimuli.

Paramétrage des constantes de temps

Après avoir décrit le comportement spatio-temporel des voies Parvo et Magno, nous allons maintenant choisir les constantes de temps afin qu'elles répondent aux caractéristiques suivantes durant une fixation oculaire. La durée de la fixation oculaire est choisie égale à 200 ms, c'est la valeur moyenne des durées expérimentales que nous avons observées. Nous voulons que la voie Parvo évolue et devient stable en fin de fixation. Pour la voie Magno X, elle évolue plus rapidement que la voie Parvo et s'arrête plus tôt. De plus, la voie Magno est censée véhiculer l'information globale de la scène au début d'une exploration. Cette information globale peut être obtenue après environ 40 ms [Castelhano and Henderson, 2008]. Ici, les paramètres sont choisis pour que la voie Magno X passe par la valeur nulle après 50 ms environ. Ainsi, nous avons choisi les constantes de temps suivantes : $\tau_c = 30$ ms pour

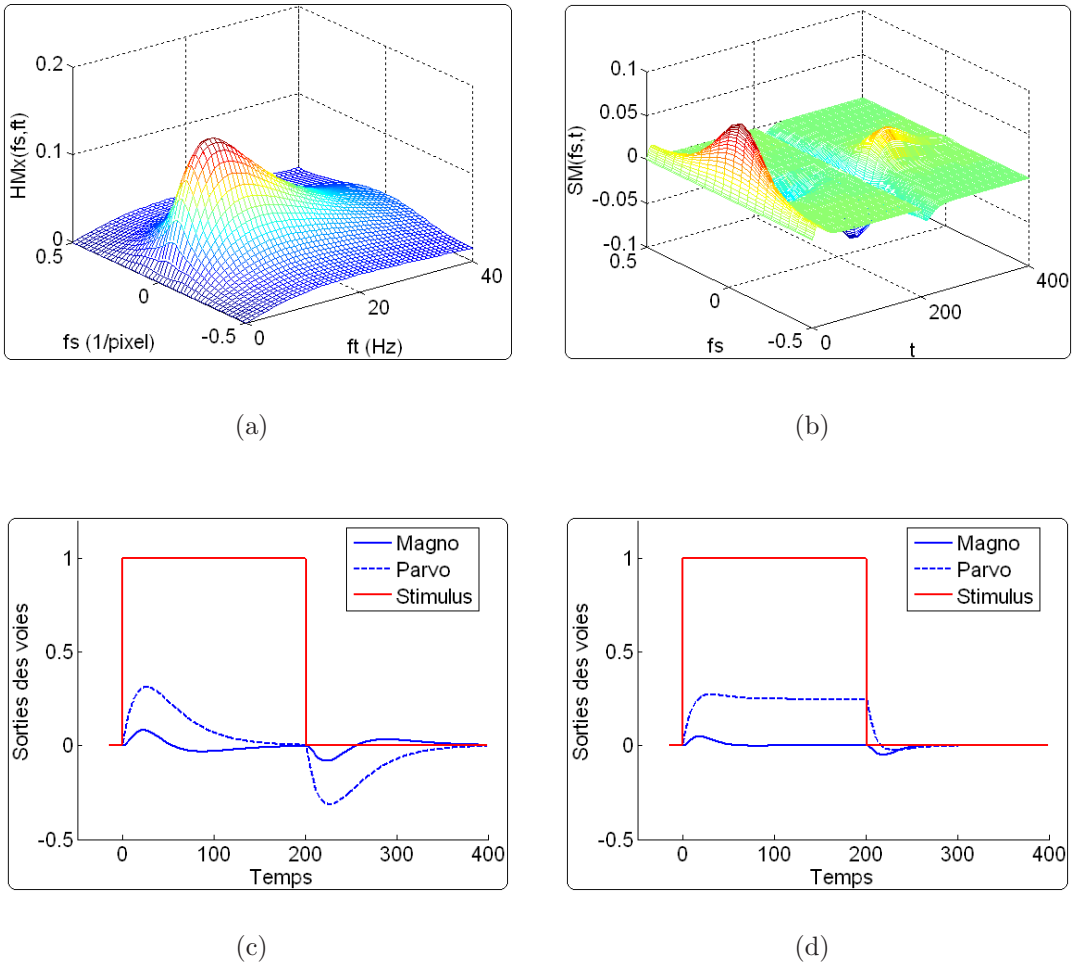


FIG. 6.6 – Réponses de la voie Magno X au stimulus "porte" : (a) Fonction de transfert ; (b) Réponse de la voie Magno X en fonction de la fréquence spatiale et du temps ; (c) Réponses de la voie Parvo et de la voie Magno X à la fréquence spatiale nulle ; (d) Réponses de la voie Parvo et de la voie Magno X à la fréquence spatiale non-nulle ($f_s = 0.15$). Les paramètres pour les photorécepteurs et les cellules horizontales sont ceux utilisés à la figure 6.5. De plus, les paramètres pour les cellules ganglionnaires parasols sont $\alpha_g = 1$, $\beta_g = 0$, $\tau_g = 15$. Pour les cellules amacrines, $\tau_a = 10$.

les photorécepteurs, $\tau_h = 22$ ms pour les cellules horizontales, $\tau_g = 15$ ms pour les cellules ganglionnaires parasols, et $\tau_a = 10$ ms pour les cellules amacrines. En ayant ainsi fixé les constantes de temps pour un fonctionnement typique durant la fixation, nous allons ensuite observer le fonctionnement avec les mêmes paramètres, durant la saccade (cf. section §2.3)

2.2.2 Simulation d'une exploration de scène naturelle

La modélisation spatio-temporelle des voies Parvo et Magno ci-dessus permet de simuler le traitement dynamique de la rétine lors de l'exploration d'une scène naturelle. Les traitements rétiniens sont simulés suivant le processus "fixation-saccade" des mouvements oculaires présenté dans le schéma de la figure 6.7. Pour cette première simulation, nous nous concentrons sur les fixations, en considérant la disparition du signal durant la saccade. La réponse à une translation du flux visuel durant la saccade sera présentée au paragraphe §2.3. Pendant la fixation, nous considérons les stimuli immobiles (nous ne modélisons pas des mouvements de micro-saccades durant la fixation), puis ils disparaissent à la fin de la fixation, suivant un comportement temporel du type "porte".

Le simulateur est conçu pour recevoir en entrée les fichiers issus de l'*eye tracker* décrivant les mouvements oculaires (durée, position) d'un sujet visionnant une scène (Fig. 6.8). Ainsi, pour une fixation donnée, l'image est captée autour de cette position et filtrée par un filtre passe-bas spatialement variant pour modéliser la décroissance de l'échelle spatiale de l'image suivant l'excentricité par rapport à cette position. Le traitement spatio-temporel de la rétine est simulé avec les trois voies : Parvo, Magno X et Magno Y, qui représentent l'évolution des informations visuelles à l'issue de la rétine lors d'une fixation. La simulation s'effectue suivant un temps proportionnel au temps réel.

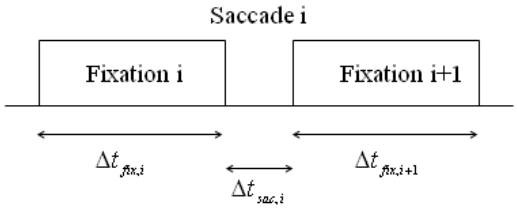


FIG. 6.7 – Diagramme de temps des fixations et des saccades utilisées dans la simulation.

La figure 6.9 illustre l'évolution des voies Parvo, Magno X et Magno Y lors d'une fixation sur une scène naturelle. Dans cette figure, le point de vue utilisé pour le filtrage passe-bas spatialement variant, qui est aussi la position de fixation, est au centre de l'image. En raison de ce filtrage, les contrastes spatiaux sont moins importants à la périphérie qu'au centre de l'image, notamment pour la voie Parvo. Au cours du temps, la voie Parvo évolue de la manière *"coarse-to-fine"*, qui représente les basses fréquences spatiales de la scène au début d'une fixation, et puis, les hautes fréquences spatiales avec des contrastes de plus en plus importants. La voie Magno X fonctionne comme le filtrage passe-bas de la voie Parvo et représente donc les

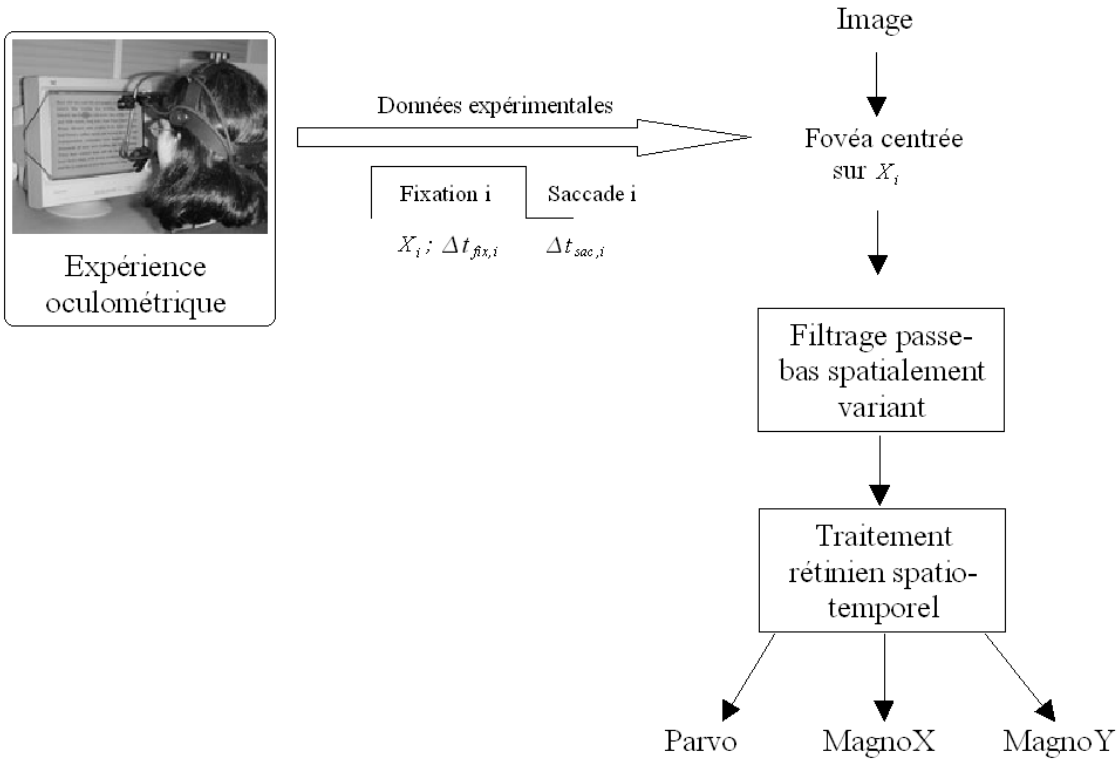


FIG. 6.8 – Déroulement de la simulation du traitement rétinien lors de l'exploration d'une image.

basses fréquences spatiales de la scène. La voie Magno Y répond à la variance locale en basses fréquences spatiales. Contrairement à la voie Parvo qui persiste jusqu'en fin de fixation, les voies Magno X et Y s'atténuent au fur et à mesure que la voie Parvo devient stable.

2.3 Traitement pendant une saccade

Nous allons maintenant étudier les traitements rétiniens pendant une saccade. En effet, pendant une saccade les stimuli projetés sur la rétine varient rapidement au niveau temporel. En considérant les fonctions de transfert de la voie Parvo et de la voie Magno X (Fig. 6.4a et 6.6a), nous voyons que les réponses de ces deux voies se ressemblent pour les hautes fréquences temporelles en véhiculant les informations basses fréquences spatiales.

Pour analyser les réponses des voies rétiniennes durant la saccade, nous allons nous placer dans un cadre réaliste issu des valeurs typiques de vitesse et d'amplitude de saccade. Nous avons considéré une vitesse moyenne de $300°/s$, pour une durée de 40 ms, soit un déplacement de 12° par saccade. La distance du sujet à l'écran est celle utilisée pour l'expérience, l'ouverture angulaire est de $1°/25$pixels. Ainsi durant une saccade, le flux visuel est translaté de 300 pixels en 40 ms.

Par manque de temps, les simulations que nous présentons ici sont des simula-

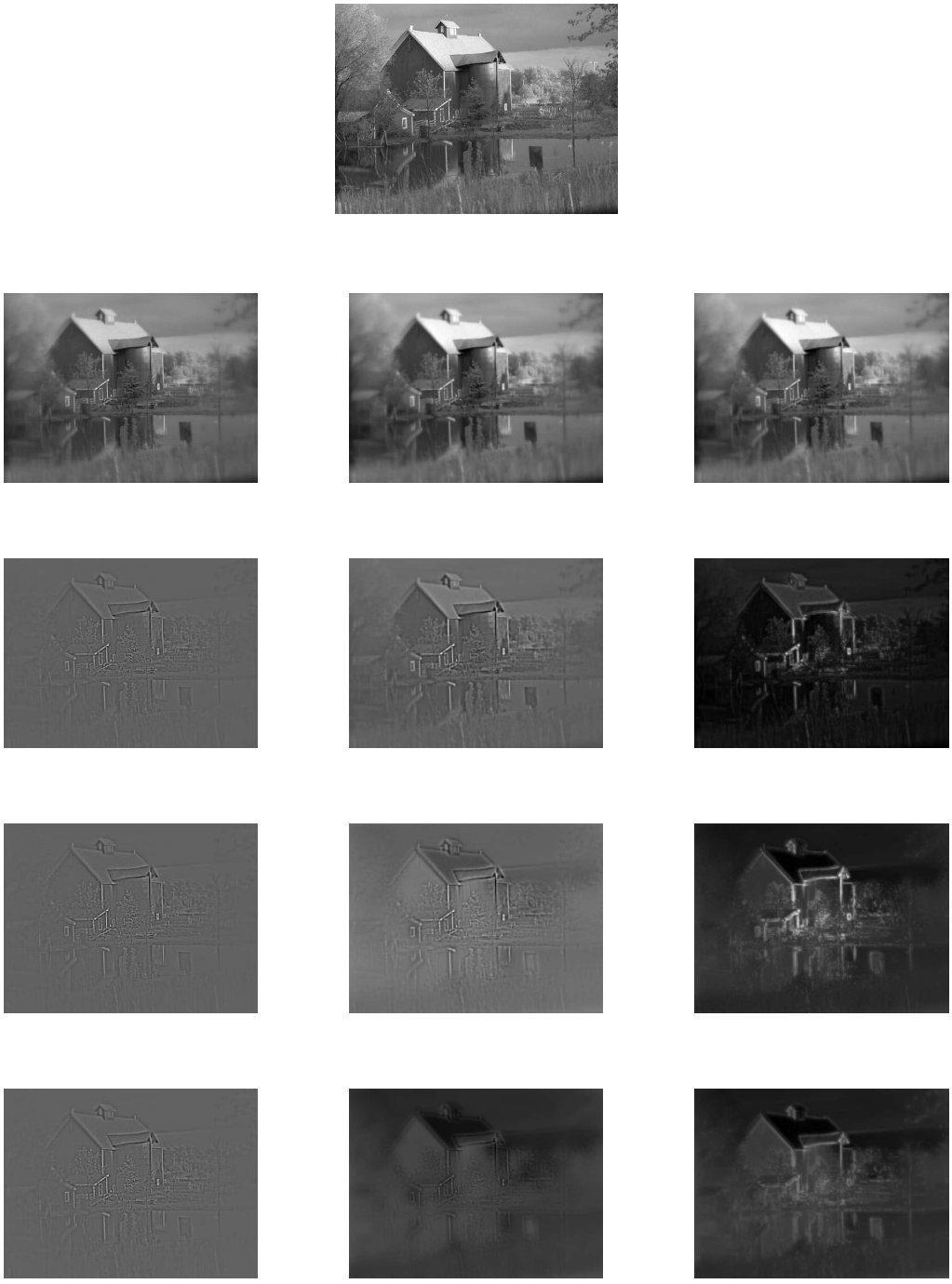
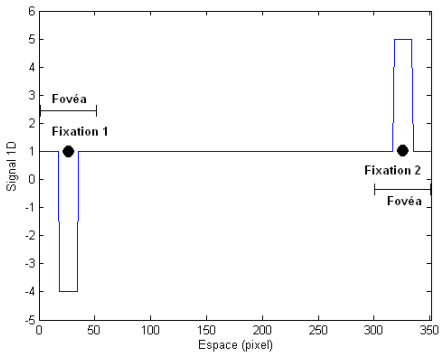
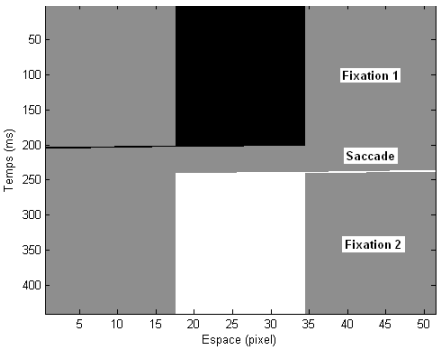


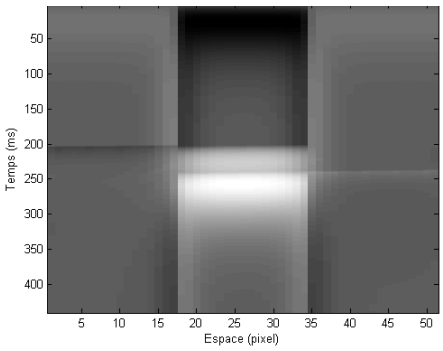
FIG. 6.9 – Exemple des trois voies : Parvo, Magno X et Magno Y (de gauche à droite) pendant une fixation sur le centre une image naturelle (la première ligne). La durée de fixation $\Delta t = 200$ ms. La 2ème ligne représente les voies à $t = 20$ ms, la 3ème ligne $t = 40$ ms, la 4ème ligne $t = 80$ ms, la 5ème ligne $t = 160$ ms (voir texte).



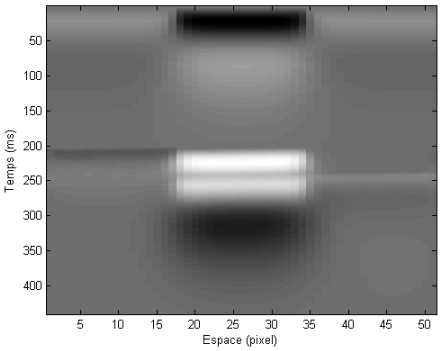
(a)



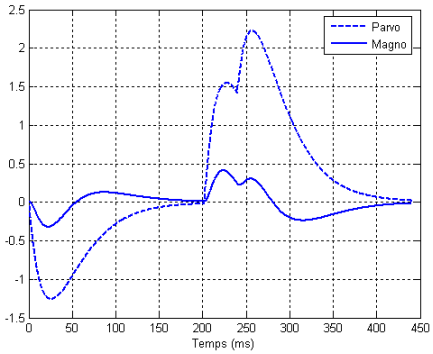
(b)



(c)



(d)



(e)

FIG. 6.10 – Modélisation du traitement rétinien pendant une saccade. (a) Signal en une dimension sur lequel une saccade s'effectue ; (b) Signal présenté sur la rétine selon le temps. (c) La voie Parvo ; (d) La voie Magno X ; (e) Les réponses de la voie Parvo et de la voie Magno X au centre de la rétine.

tions sur stimuli artificiels en une dimension. Mais cela nous permet déjà d'observer des comportements intéressants et d'offrir des perspectives à ce travail. Le stimulus artificiel est ainsi construit : durant une fixation F_1, un objet sombre sur un fond gris est complètement inclu dans la fovéa. A une distance de 300 pixels (F_2), il y a un objet de même taille de contraste opposé. Après une fixation de 200 ms sur F_1, l'œil va aller sur F_2 avec une saccade de 40 ms et rester 200 ms. La simulation s'effectue par pas de temps de 0.5 ms. La figure 6.10a illustre ce stimulus spatial. La figure 6.10b représente l'évolution spatio-temporelle de la scène visuelle au centre de la rétine à chaque pas de temps suivant le décours temporel suivant : stabilisation de 200 ms en F_1, translation vers F_2 durant 40 ms, puis de nouveau stabilisation de 200 ms en F_2. Ainsi, la figure 6.10c,d représente la voie Parvo et la voie Magno X pendant cette durée. Au début de la fixation F_1, la voie Parvo répond aux basses fréquences spatiales, et puis aux hautes fréquences spatiales en fin de fixation. La voie Magno X véhicule les informations basses fréquences spatiales au début d'une fixation et puis diminue à zéro pendant la fixation F_1 (Fig. 6.10d). En fin de fixation, il ne reste que la réponse de la voie Parvo.

Durant la saccade, les deux voies réagissent à la variation des stimuli. Comme la durée d'une saccade est très courte (40 ms), les réponses de la voie Parvo et de la voie Magno ne sont pas stables. Ainsi, les deux voies véhiculent les informations basses fréquences spatiales. En outre, la variation temporelle de la voie Magno X est plus rapide que la voie Parvo (Fig. 6.10e) car la voie Magno X ne contient que des hautes fréquences temporelles tandis que la voie Parvo contient aussi des basses fréquences temporelles. Pendant la fixation 2, les deux voies réagissent de la même manière que pendant la fixation 1.

2.4 Synthèse et Perspectives

Les simulations ci-dessus illustrent le fonctionnement des voies Parvo et Magno X tel que fourni par le modèle, durant une fixation puis en enchaînement avec une saccade. Nous observons bien des réponses différenciées durant les fixations et les saccades. Lors d'une fixation, les réponses de ces deux voies sont "phasiques" au début de fixation et "toniques" en fin de fixation. Lors d'une saccade, les réponses sont "phasiques" car la durée d'une saccade est courte et les stimuli varient rapidement. De plus au niveau des fréquences spatiales, alors qu'en fin de fixation la voie Parvo véhiculent les hautes fréquences, pendant une saccade, les voies Parvo et Magno véhiculent les informations en basses fréquences spatiales, avec un gain d'autant plus faible que la fréquence temporelle est grande. Cela montre une diminution de la sensibilité au contraste pendant une saccade, comme cela a été montré dans les expérimentations manipulant les stimuli visuels durant la saccade [Diamond et al., 2000], pour expliquer le phénomène d'"inhibition saccadique" où l'on ne perçoit pas le déplacement du flux visuel durant la saccade. Dans ces expérimentations, la décroissance de la sensibilité au contraste spatial a été observée lors des saccades et cela est de façon robuste pour différents contrastes. La sensibilité au contraste spatial est d'autant plus faible que la fréquence temporelle augmente. Les données physiologiques montrent que c'est principalement un contrôle du gain sur les voies magnocellulaires qui expliquerait l'inhibition saccadique, alors que les voies parvo-

cellulaires seraient préservées durant la saccade. Notre modélisation du traitement spatio-temporel de la rétine durant la saccade montre une activité sur les voies Parvo et Magno X avec un fonctionnement assez similaire de filtrage spatial passe-bas d'autant plus que la fréquence temporelle augmente, avec une réponse temporelle plus rapide pour la voie Magno X par rapport à la voie Parvo. On aurait pu s'attendre à une réponse plus lente pour la voie Parvo. Cet écart est peut-être dû au choix de constantes de temps qui a été fait suivant des critères de décours temporel durant les fixations, car nous voulions en premier lieu tester ces configurations en mode "fixation" et voir leur validité pour les saccades. Il nous semble donc nécessaire de préciser ces choix de constantes de temps en tenant compte également des données issues d'expérimentations durant les saccades. Ces simulations réalisées en une dimension seront facilement étendues à deux dimensions sur des scènes visuelles naturelles pour obtenir une modélisation plus complète du traitement spatio-temporel de la rétine lors des mouvements oculaires.

Cette première ébauche d'utilisation du modèle fonctionnel spatio-temporel de la rétine pour l'étude des mouvements oculaires pourrait se poursuivre pour également étudier les microsaccades. En réalité, les yeux sont toujours en mouvement. Même lors d'une fixation, les yeux bougent autour de cette position. Les microsaccades entraînent une fluctuation de luminance de l'image projetée sur la rétine. Selon Rucci [Rucci, 2008], cette fluctuation de luminance joue un rôle important dans l'égalisation du spectre des scènes naturelles sur différentes fréquences spatiales. En utilisant la fluctuation de luminance causée par des microsaccades, combinée avec les modèles des cellules ganglionnaires de la rétine, Rucci a réussi à expliquer certaines propriétés des réponses rétiniennes comme la réponse permanente de la voie Parvo et l'insensibilité de la voie Magno pour les stimuli statiques. Cependant, les modèles des cellules ganglionnaires dans [Rucci, 2008] ont été modélisés par le produit de deux fonctions de transfert séparées en spatial et en temporel en utilisant les données physiologiques obtenues sur les cellules ganglionnaires. Dans le modèle de rétine développé par J. Hérault et utilisé dans cette thèse, l'inséparabilité du comportement spatio-temporel est mise en avant. Il nous semble alors intéressant d'analyser, à partir des travaux de Rucci, l'impact du traitement inséparable en temps et en espace en utilisant notre modèle de rétine dans un contexte de fluctuations visuelles par les microsaccades. Le rôle fonctionnel des microsaccades est actuellement une question scientifique très débattue, il nous semble que ce modèle spatio-temporel de la rétine serait un bon outil d'analyse.

De plus, en se focalisant plus spécialement sur notre objet d'étude, la carte de saillance, il nous semble important dans les travaux futurs de mieux prendre en compte ce fonctionnement spatio-temporel jusqu'au niveau cortical pour l'établissement de la saillance visuelle. Dans les chapitres précédents, nous avons vu que durant une fixation, les informations visuelles autour de ce point fixé sont analysées et puis fusionnées pour créer une carte de saillance prédisant la fixation suivante. Cette carte a été construite pour chaque fixation en tenant compte des informations visuelles disponibles en fin de fixation. Cependant, la perception des stimuli visuels par la rétine et puis par le cortex visuel primaire est un processus dynamique. La création de la saillance doit être aussi un processus dynamique suivant le flux visuel rétinien.

Ainsi, la carte de saillance doit être mise à jour au cours du temps en se basant sur les informations véhiculées par la voies Parvo et Magno, selon un processus temporel *"coarse to fine"*. Au début d'une fixation, les informations visuelles transmises sont en basses fréquences spatiales. La voie Magno peut donc contribuer à la saillance, majoritairement et au plus tôt en début de fixation. En fin de fixation, la voie Parvo représente les informations en hautes fréquences spatiales des stimuli tandis que la voie Magno disparaît. Ainsi c'est la voie Parvo qui joue un rôle plus important dans la saillance en fin de fixation.

3 Filtrage spatialement variant

Abordons maintenant, une deuxième évolution à apporter à notre modèle. Cela concerne l'échantillonnage spatialement variant (SV) dans la rétine. Jusqu'à présent, nous avons modélisé le traitement par échantillonnage régulier mais avec une fonction d'échelle spatialement variante à travers un filtre passe-bas dont la fréquence de coupure diminue quand l'excentricité augmente. Cependant la densité des photorécepteurs et des cellules ganglionnaires n'est pas uniforme et conduit à un échantillonnage à taux variable augmentant avec l'excentricité et donc un traitement en multi-résolution depuis la fovéa jusqu'à la périphérie. Cet échantillonnage à taux variable va avoir des conséquences sur la mise en place des filtres corticaux passe-bande orientés. En effet, un même filtre avec une fonction de transfert donnée en zone fovéale, va voir sa fonction de transfert modifiée suivant la localisation spatiale.

Pour étudier cela, nous allons d'abord créer une fonction de compression implémentant le sous-échantillonnage SV, puis nous allons étudier les modifications des fonctions de transfert des filtres corticaux LogNormaux suivant le taux de compression, et enfin, nous allons étudier l'impact sur les fonctions d'interaction avant la création de la carte de saillance.

3.1 Fonction de compression

Pour construire la fonction de compression, nous distinguons les indices spatiaux définissant d'une part l'image initiale "dans le monde" avant sa projection sur la rétine et d'autre part, après projection sur la rétine. En une dimension, on note x l'indice spatial décrivant l'image initial et u l'indice spatial décrivant l'image sur les photorécepteurs de la rétine. Ces indices sont notés à partir de la position zéro au centre de l'image et sur la fovéa. La fonction de compression est la fonction $u = \rho(x)$ reliant la position u sur la rétine à partir de la position x sur l'image initiale. Par la suite, la position x sera appelée position "dans le monde".

Le taux de sous-échantillonnage est alors défini par :

$$R(x) = \frac{dx}{du} = \frac{1}{\rho'(x)},$$

avec x la position en une dimension par rapport au centre de l'image, et u la position sur la rétine par rapport à la fovéa. Au centre, on a $R(0) = 1$ et donc $\rho'(0) = 1$.

La densité des photorécepteurs et des cellules ganglionnaires présentée au chapitre 5 est également liée à cette fonction de compression :

$$d(x) = d_{max}.\frac{du}{dx},$$

avec $d(0) = d_{max}$ la densité maximale au centre, dans la fovéa. La dérivée de la fonction de compression est alors :

$$\rho'(x) = \frac{du}{dx} = \frac{d(x)}{d_{max}} \tag{6.16}$$

et en intégrant, on obtient :

$$\rho(x) = \int_0^x \frac{d(\xi)}{d_{max}} d\xi \tag{6.17}$$

La courbe de densité est connue en fonction de l'excentricité en degré angulaire. Nous convertissons cette fonction selon des abscisses en pixel [2]. La fonction obtenue est présentée à la figure 6.11a. L'intégrale de cette fonction normalisée par la densité maximale est présentée à la figure 6.11b. Au centre de l'image, le taux de sous-échantillonnage est à 1, il n'y a pas de compression. En s'éloignant vers la périphérie, le taux de sous-échantillonnage augmente, la compression est plus importante.

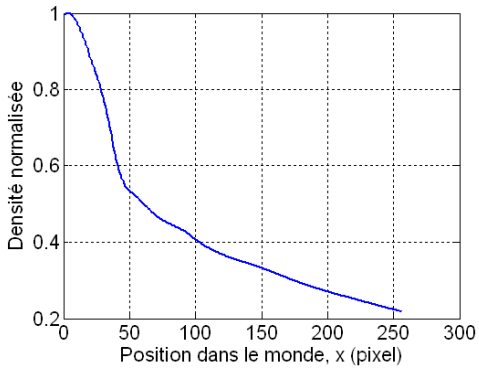
Ce sous-échantillonnage SV est illustré en l'appliquant à l'image "Léna" (Fig. 6.11c). Pour implémenter cette fonction à deux dimensions, nous l'appliquons successivement dans la direction horizontale et puis dans celle verticale. Préalablement à cette opération de sous-échantillonnage afin de respecter le théorème de Shannon, l'image est filtrée par un filtre passe-bas SV selon les directions horizontale et verticale, comme expliqué au chapitre 5. La fréquence de coupure de ce filtre varie selon la position sur l'image et est égale à la moitié de la fréquence de sous-échantillonnage $f_c(x) = \frac{f_e(x)}{2} = \frac{1}{2R(x)}$. A la figure 6.11d, nous observons une compression unitaire dans la zone centrale de l'image et une compression importante à la périphérie.

3.2 Influence du sous-échantillonnage SV sur les filtres corticaux

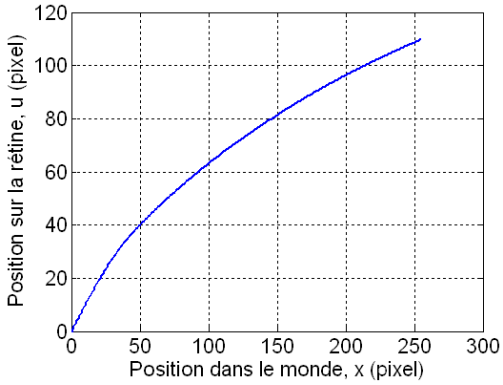
3.2.1 Filtre cortical SV

Dans le modèle de saillance présenté au chapitre 2, le banc de filtres corticaux est formé par des filtres de type passe-bande (filtre LogNormal). Leurs caractéristiques sont prédéfinies et ne dépendent pas des positions spatiales. Ils sont spatialement invariants (SI) à l'opposé des filtres SV dont les caractéristiques (fréquence centrale, bande passante, ...) vont dépendre des positions spatiales. En appliquant un filtre SI à une image sous-échantillonnée SV dont la fréquence d'échantillonnage est spatialement variante (comme par exemple, vu précédemment), on va obtenir un filtre SV. Nous allons étudier cette transformation spatiale pour les filtres corticaux

[2]Selon notre expérience, un degré angulaire correspond à 25 pixels



(a) Fonction de densité



(b) Fonction de compression

(c) Image initiale

(d) Image sous-échantillonnée

FIG. 6.11 – Illustration du sous-échantillonnage spatialement variant : (a) Fonction de densité $d(x)$; (b) Fonction de compression ; (c) Image "Léna" initiale ; (d) Image "Léna" sous-échantillonnée SV selon cette fonction de compression. Le point de vue est au centre de l'image dont la taille est 512×512 pixels (soit $20° \times 20°$ d'angle visuel). La fonction de compression est appliquée successivement suivant les directions horizontale et verticale pour obtenir l'image sous-échantillonnée SV.

LogNormaux appliqués aux images rétiniennes sous-échantillonnée SV. Dans un premier temps, nous nous intéressons à un filtre LogNormal quelconque, puis ensuite au banc de filtres.

Cette question a été abordée par Croll [Croll, 1998] dans le contexte de l'étude des projections rétinotopiques. Trois étapes sont alors définies. La première est l'étape de projection des coordonnées rétiniennes aux coordonnées corticales. La deuxième étape est l'implémentation d'un filtre cortical SI. Enfin la troisième étape, à titre de vérification, est la projection inverse de retour dans les coordonnées rétiniennes du filtrage SV équivalent. Cette méthodologie s'applique à notre contexte. La première étape correspond à l'étape de sous-échantillonnage SV vue au paragraphe précédent. La seconde étape est l'implémentation d'un filtre LogNormal SI sur cette image sous-échantillonnée SV et on obtient alors un filtre SV dont la fréquence centrale, l'orientation et la bande passante varient en fonction de la position spatiale. La troisième

étape de validation correspondrait ici à une étape de sur-échantillonnage réciproque.

Nous allons étudier l'influence du sous-échantillonnage SV sur un filtre LogNormal et donc trouver la relation entre le filtre LogNormal SI appliqué sur la rétine échantillonnée SV et le filtre LogNormal SV équivalent avec un échantillonnage uniforme.

Nous désignons maintenant les notations qui seront utilisées dans la suite :
$\mathbf{x} = (x_1, x_2)$ la position d'un pixel d'une image "dans le monde",
$\mathbf{u} = (u_1, u_2)$ la position sur la rétine,
P la fonction de compression en deux dimensions, et :

$$\mathbf{u} = P(\mathbf{x}) = \begin{bmatrix} P_1(\mathbf{x}) \\ P_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \rho(x_1) \\ \rho(x_2) \end{bmatrix}$$

avec $\rho(x)$ la fonction de compression en une dimension construite à la section §3.1. L'image "dans le monde" est notée $e(\mathbf{x})$, l'image sous-échantillonnée SV est notée $r(\mathbf{u})$. Ainsi on a : $r(\mathbf{u}) = e(P^{-1}(\mathbf{u}))$ avec P^{-1} la fonction inverse de P représentant le sur-échantillonnage.

Soit $g(\mathbf{u})$ la réponse impulsionnelle d'un filtre SI sur la rétine (dans notre contexte, le filtre SI est un filtre LogNormal). Alors, la sortie du filtre SI sur la rétine est :

$$\begin{aligned} q(\mathbf{u}) &= r(\mathbf{u}) * g(\mathbf{u}) \\ &= \int r(\boldsymbol{\nu}) g(\mathbf{u} - \boldsymbol{\nu}) d\boldsymbol{\nu} \\ &= \int e(\boldsymbol{\xi}) g(P(\mathbf{x}) - P(\boldsymbol{\xi})) \Big| \det J_{P(\boldsymbol{\xi})} \Big| d\boldsymbol{\xi} \end{aligned} \tag{6.18}$$

avec $\boldsymbol{\nu} = P(\boldsymbol{\xi})$, $\boldsymbol{\xi}$ position "dans le monde", $\boldsymbol{\nu}$ position sur la rétine et $J_{P(\boldsymbol{\xi})}$ est le Jacobien de $P(\boldsymbol{\xi})$:

$$J_{P(\boldsymbol{\xi})} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial P_1}{\partial \xi_1} & \frac{\partial P_1}{\partial \xi_2} \\ \\ \frac{\partial P_2}{\partial \xi_1} & \frac{\partial P_2}{\partial \xi_2} \end{bmatrix} = \begin{bmatrix} \rho'(\xi_1) & 0 \\ 0 & \rho'(\xi_2) \end{bmatrix}$$

Or, "dans le monde", la sortie du filtre SV est :

$$s(\mathbf{x}) = e(\mathbf{x}) * h_{\mathbf{x}}(\mathbf{x}) = \int e(\boldsymbol{\xi}) h_{\mathbf{x}}(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} \tag{6.19}$$

avec $h_{\mathbf{x}}$ le filtre SV à la position $\mathbf{x}$.

La sortie du filtre SV "dans le monde" et la sortie du filtre SI sur la rétine doivent être identiques, c'est-à-dire : $q(\mathbf{u}) = s(P^{-1}(\mathbf{u})) = s(\mathbf{x})$. Ainsi par identification, on obtient :

$$h_{\mathbf{x}}(\mathbf{x} - \boldsymbol{\xi}) = g(P(\mathbf{x}) - P(\boldsymbol{\xi})) \Big| \det J_{P(\boldsymbol{\xi})} \Big| \tag{6.20}$$

A partir de cette équation, en linéarisant $P(\mathbf{x})$, on peut trouver la relation entre $H_{\mathbf{x}}(\mathbf{f})$, transformée de Fourier de $h_{\mathbf{x}}(\boldsymbol{\xi})$, et $G(\mathbf{f})$, transformée de Fourier de $g(\mathbf{u})$ [Croll, 1998] :

$$H_{\mathbf{x}}(\mathbf{f}) = G\left(\left[J_{P(\mathbf{x})}^{-1}\right]^T \mathbf{f}\right) = G\left(J_{P(\mathbf{x})}^{-1}\mathbf{f}\right) \tag{6.21}$$

en notant que $J_{P(\mathbf{x})}^{-1}$ est symétrique et "T" signifie la transposée d'une matrice. Les détails du calcul se trouvent en annexe G.

En se donnant une fonction de compression $P(\mathbf{x})$, nous allons maintenant examiner la relation entre le filtre SI, $G(\mathbf{f})$, sur la rétine et le filtre SV, $H_{\mathbf{x}}(\mathbf{f})$, "dans le monde" selon différentes positions $\mathbf{x}$. La réponse fréquentielle $G(\mathbf{f})$ du filtre SI sur la rétine est celle d'un filtre LogNormal (cf. chapitre 2), qui est caractérisée par une fréquence centrale f_0, une orientation θ_0 et des variances σ_f, σ_θ :

$$G(\mathbf{f}) = \exp\left\{-\frac{\left[\log\left(\frac{\|\mathbf{f}\|}{f_0}\right)\right]^2}{2\sigma_f^2}\right\} \exp\left\{-\frac{\left[\Phi(\mathbf{f}) - \theta_0\right]^2}{2\sigma_\theta^2}\right\}$$

où

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \|\mathbf{f}\| = \sqrt{f_1^2 + f_2^2}, \Phi(\mathbf{f}) = \arctan\left(\frac{f_2}{f_1}\right)$$

En appliquant l'équation 6.21, la réponse fréquentielle du filtre SV correspondant "dans le monde" est alors représentée par l'équation suivante :

$$H_{\mathbf{x}}(\mathbf{f}) = \exp\left\{-\frac{\left[\log\left(\frac{\|J_{P(\mathbf{x})}^{-1}\mathbf{f}\|}{f_0}\right)\right]^2}{2\sigma_f^2}\right\} \exp\left\{-\frac{\left[\Phi(J_{P(\mathbf{x})}^{-1}\mathbf{f}) - \theta_0\right]^2}{2\sigma_\theta^2}\right\} \tag{6.22}$$

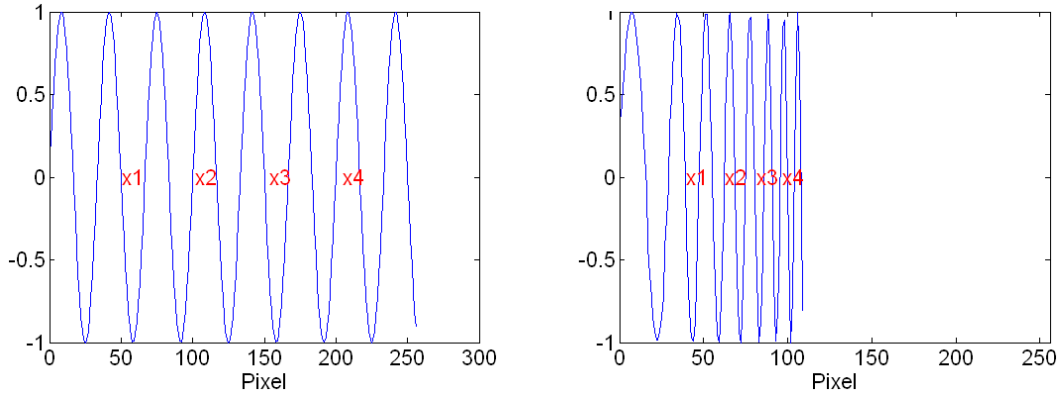
C'est un filtre LogNormal de fréquence centrale et d'orientation variante en fonction de la position spatiale. On note f_m cette fréquence centrale, une fonction de la position spatiale et θ_m l'orientation, fonction également de la position spatiale. Après développement, on trouve que f_m et θ_m suivent les relations suivantes :

$$\begin{cases} \theta_m = \arctan\left(\frac{J_{22}}{J_{11}}\tan(\theta_0)\right) \\ f_m = f_0\sqrt{J_{11}^2\cos(\theta_0)^2 + J_{22}^2\sin(\theta_0)^2} \end{cases} \tag{6.23}$$

Les détails du calcul se trouvent en annexe H. La fréquence et l'orientation du filtre SV dépendent de la position $\mathbf{x}$ qui est présente dans J_{11} et J_{22}. Autrement dit, lorsqu'on filtre une image sous-échantillonnée SV par un filtre LogNormal SI à une fréquence f_0 et une orientation θ_0 spécifiques, cela est équivalent à un filtrage de l'image initiale "dans le monde" par un filtre LogNormal SV dont la fréquence et l'orientation varient selon la position sur l'image. Alors que le filtre LogNormal SI est configuré par une fréquence centrale f_0 et une orientation θ_0 indépendamment l'une de l'autre, on obtient pour le filtre SV une configuration bien sûr spatialement variante, mais aussi un couplage entre la fréquence f_m et l'orientation θ_m à travers le paramètre commun θ_0 ($\theta_m(\mathbf{x}, \theta_0)$, $f_m(\mathbf{x}, \theta_0, f_0)$). Les bandes passantes varient également ; nous n'avons pas développé les équations analytiques. Les conséquences de cette caractéristique seront analysées au paragraphe §3.2.2 à propos du banc de filtres SV et au paragraphe §3.3.1 à propos des interactions pour former la carte de

saillance.

Nous voulons d'abord présenter un exemple de filtres SV obtenus à partir du sous-échantillonnage SV et d'un filtre SI pour un signal en une dimension (Fig. 6.12a). Le sous-échantillonnage SV effectué à partir de la position 0 donne un signal sous-échantillonné à la figure 6.12b. Nous avons examiné 4 positions différentes ; ces positions (x1, x2, x3, x4) sont indiquées sur le signal initial et celui sous-échantillonné SV. Plus la position est éloignée de la position 0, plus la compression est importante. La figure 6.13 montre les filtres SV aux 4 positions différentes. Plus la position est en périphérie, plus le filtre SV correspondant est en basses fréquences. Ainsi, si la fréquence centrale du filtre SI correspond à la fréquence du signal "dans le monde", on obtiendra un signal de sortie dont l'amplitude décroît de x1 à x4, pour les positions de plus en plus périphériques.



(a) Signal initial (b) Signal sous-échantillonné SV

FIG. 6.12 – Exemple d'un signal en une dimension sous-échantillonné SV et illustration pour 4 positions spatiales x1, x2, x3, x4. Elles sont notées à la fois sur le signal initial et sur le signal sous-échantillonné SV. (a) Signal initial ; (b) Signal sous-échantillonné SV.

De même nous avons testé le filtre SV en deux dimensions pour différentes positions spatiales. Ces positions spatiales (x1, x2, x3, x4) sont indiquées à la figure 6.14a. A titre d'illustration, le résultat du sous-échantillonnage SV est montré à la figure 6.14b. La figure 6.15 montre alors pour ces 4 positions spatiales, le filtre SV obtenu à partir d'un seul filtre SI. En x1, les deux filtres sont identiques. Le déplacement en x4 s'effectue sur la diagonale, alors les composantes J_{11} et J_{22} de la matrice Jacobienne sont identiques, et donc le filtre noté SV-x4 est à la même orientation que le filtre SI. Il y a seulement une diminution de la fréquence $f_m = f_0 \, |J_{11}|$. Pour la position x2, le déplacement s'effectue horizontalement, alors uniquement la composante horizontale de la fréquence centrale est impactée par une diminution, le filtre SI qui était sensible aux orientations obliques devient à cette position, un filtre sensible aux orientations plus horizontales. Le phénomène est identique pour la position x3, vis-à-vis des verticales.

Ainsi, lorsque l'on applique un filtre SI sur une image sous-échantillonnée SV, les

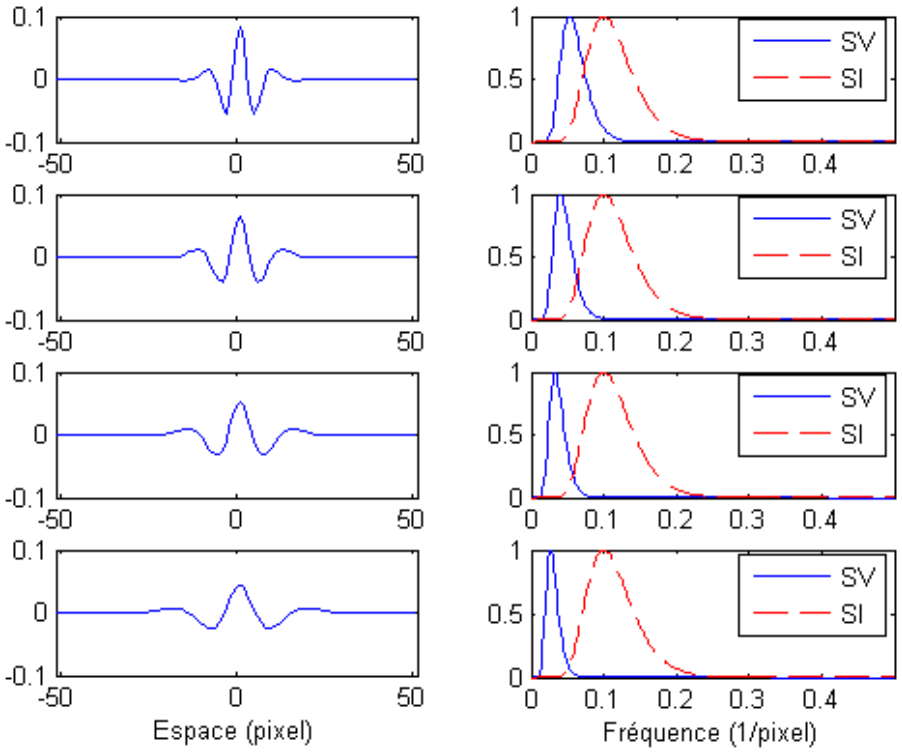
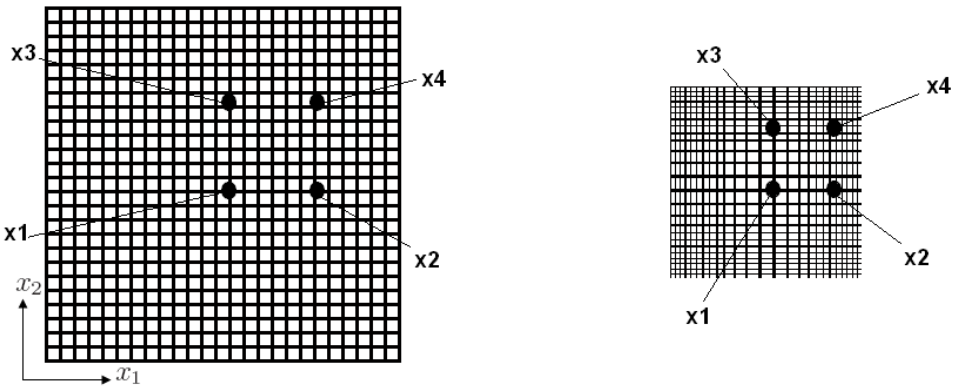


FIG. 6.13 – Exemple de filtres SV obtenus à partir du sous-échantillonnage SV et d'un filtre SI pour le signal en une dimension à la figure 6.12. Les filtres SV aux 4 positions (x1, x2, x3, x4) correspondent aux 4 lignes de haut en bas, à gauche : la partie réelle de la réponse impulsionnelle du filtre SV, à droite : la fonction de transfert du filtre SV.



(a) Image initiale (b) Image sous-échantillonnée SV

FIG. 6.14 – Exemple d'une image sous-échantillonnée SV et illustration pour 4 positions spatiales x1, x2, x3, x4. Elles sont notées à la fois sur l'image initiale et sur l'image sous-échantillonnée SV. (a) Image initiale de 300×300 pixels (soit $11.7° \times 11.7°$ d'angle visuel), le point de vue est au centre de l'image ; (b) Image sous-échantillonée SV.

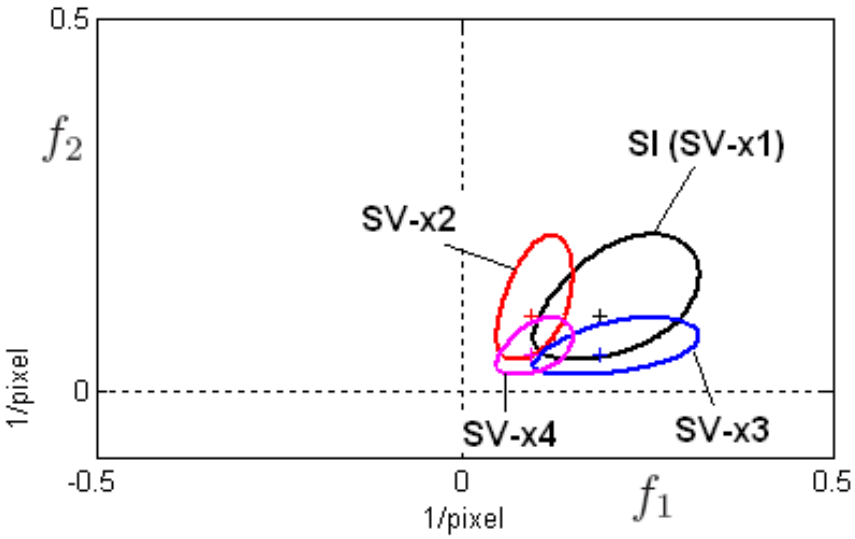


FIG. 6.15 – Exemple de filtres SV obtenus à partir du sous-échantillonnage SV et d'un filtre SI pour l'image à la figure 6.14. Les filtres SV sont représentés par leurs contours à mi-hauteur des fonctions de transfert aux 4 positions.

informations (fréquences et orientations) extraites au centre sont différentes de celles en périphérie en comparaison avec les informations extraites de l'image initiale si l'on applique également le même filtre SI sur cette image. En fait, les informations prises au centre de l'image sous-échantillonnée SV coïncident avec celles de l'image initiale. En revanche, les informations extraites en périphérie de l'image sous-échantillonnée SV ne coïncident plus ; en particulier, elles sont extraites en plus basses fréquences. De plus, sur l'image sous-échantillonnée SV, le contraste en périphérie est moins fort que celui au centre à cause du filtrage passe-bas SV de l'image initiale avant sous-échantillonnage. Ainsi, après le filtrage LogNormal SI de l'image sous-échantillonnée, l'énergie au centre est plus importante que celle en périphérie. Selon les différentes positions, cela va influencer la saillance évaluée par la carte de saillance que nous verrons au paragraphe §3.4.

3.2.2 Influence du sous-échantillonnage SV sur un banc de filtres SI

Nous savons ci-dessus que le filtrage d'une image sous-échantillonnée SV (sur la rétine) par un filtre LogNormal SI correspond au filtrage de l'image originale ("dans le monde") par un filtre SV dont l'orientation et la fréquence varient selon la position sur cette image. Ainsi, si un banc de filtres LogNormaux SI est appliqué à une image sous-échantillonnée SV pour la décomposer en différentes orientations et différentes fréquences comme au chapitre 2, le banc de filtres SI devient un banc de filtres SV.

Nous allons maintenant étudier la variation des orientations et des fréquences de ce banc de filtres SV constitué de 32 filtres (8 orientations et 4 fréquences). La variation de chaque filtre en fréquence centrale et en orientation est représentée par l'équation 6.23. La figure 6.16 illustre ces décalages de positions fréquentielles pour tous les filtres du banc pour une position spatiale avec coordonnées (30, 20) en

pixel. Nous observons que les filtres SI de même orientation deviennent les filtres SV de même orientation mais cette dernière est différente de celle des filtres SI. En revanche, les filtres SI de même fréquence correspondent aux filtres SV de différentes fréquences. De plus, la fréquence d'un filtre SV est inférieure ou égale à celle du filtre SI correspondant.

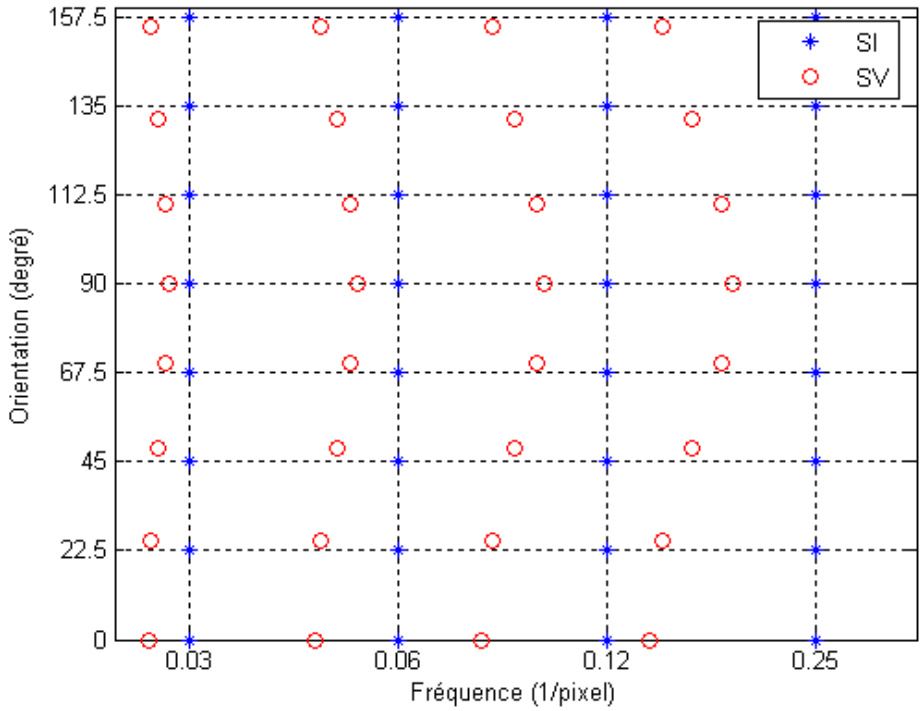


FIG. 6.16 – La variation de la fréquence et de l'orientation des filtres SV pour le banc de filtres LogNormaux SI à la position spatiale $\mathbf{x}$ avec coordonnées (30, 20) en pixel ($J_{11} = 0.8$ et $J_{22} = 0.9$).

3.3 Influence du sous-échantillonnage SV sur les interactions

3.3.1 Interactions courtes

Puisque les orientations et les fréquences des filtres SV varient selon la position sur l'image, il est nécessaire de modifier les interactions courtes qui font intervenir les cartes d'énergie de fréquences et orientations voisines. En effet, nous avons vu au chapitre 2 la relation d'interaction à coefficients constants qui tend à renforcer l'orientation pour des fréquences voisines et mettre en compétition des orientations voisines à la même fréquence. Ici en conséquence, il faudrait envisager d'étendre cette relation d'interaction avec des coefficients dont les valeurs dépendent des positions spatiales, car les fréquences et les orientations des filtres SV voisins changent pour chaque position. Comme nous avons vu ci-dessus, pour une position spatiale donnée (J_{11} et J_{22} fixées), des filtres SI de même orientation restent des filtres SV de même orientation, mais à une orientation commune différente des filtres SI. De plus, les fréquences centrales des filtres SV sont proportionnelles à celles des filtres SI

(Eq. 6.23) et gardent donc la relation dyadique établie entre eux lors de la construction du banc de filtres. Ainsi, les coefficients d'interaction entre les cartes de même orientation et de fréquences voisines ne changent pas et restent à 0.5. En revanche, les coefficients d'interaction entre les cartes de même fréquence et d'orientations voisines doivent bien être évaluées suivant les positions spatiales, car la différence d'orientation des filtres SV concernés n'est plus constante. Alors, nous proposons de modifier les coefficients de manière à être inversement proportionnels aux différences des orientations des filtres SV voisins. Plus deux filtres SV sont proches en orientation, plus l'inhibition entre eux est importante et inversement. Cela est résumé par l'équation décrivant les interactions courtes, qui est modifiée de celle présentée au chapitre 2 :

$$e^s_{ij} = e^{norm}_{ij} + 0.5e^{norm}_{i,j-1} + 0.5e^{norm}_{i,j+1} - C_{ant} \cdot e^{norm}_{i-1,j} - C_{pos} \cdot e^{norm}_{i+1,j} \tag{6.24}$$

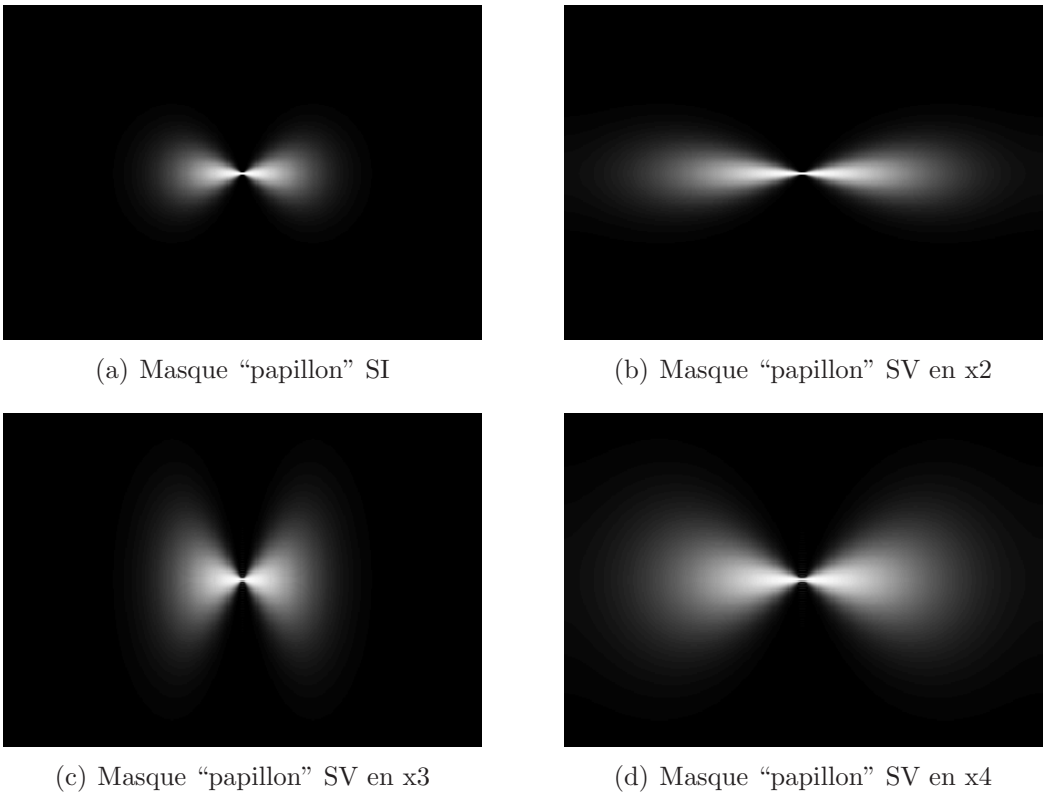
où e^{norm}_{ij} est une carte d'énergie après les normalisations comme au chapitre 2, e^s_{ij} une carte d'énergie après les interactions courtes. Les coefficients d'interactions SV sont notés C_{ant}, C_{pos}, et on a :
$C_{ant} = \min\left\{0.5\frac{\Delta\theta^{SI}}{\Delta\theta^{SV}_{ant}}, 1\right\}$ avec $\Delta\theta^{SI} = \frac{180}{M}$ (M est le nombre d'orientations) et $\Delta\theta^{SV}_{ant} = \theta^{SV}_i - \theta^{SV}_{i-1}$, θ^{SV}_i l'orientation du filtre SV correspondant au filtre SI de l'orientation θ_i
$C_{pos} = \min\left\{0.5\frac{\Delta\theta^{SI}}{\Delta\theta^{SV}_{pos}}, 1\right\}$ avec $\Delta\theta^{SV}_{pos} = \theta^{SV}_{i+1} - \theta^{SV}_i$

3.3.2 Interactions longues

Les interactions longues concernent les pixels voisins sur chaque carte d'énergie indépendamment les unes des autres. Or, les positions spatiales voisines ont des valeurs J_{11}, J_{22} proches et les fréquences et les orientations sont donc également proches. Ainsi en première approximation, nous pouvons appliquer les interactions longues comme au chapitre 2. Comme les filtres LogNormaux, à cause du sous-échantillonnage SV, le masque "papillon" devient également spatialement variant selon la position sur l'image. La figure 6.17 montre un masque "papillon" SI et les masques SV correspondant à des positions distinctes à la périphérie. Nous observons qu'à la périphérie, la taille du masque "papillon" SV est plus importante que celle du masque SI. Cela correspond à une fréquence spatiale plus faible, perçue à la périphérie de l'image.

3.4 Carte de saillance d'une image sous-échantillonnée SV

Nous appliquons les différentes étapes vues au chapitre 2 et conduisant au modèle d'attention visuelle, mais en intégrant en plus le sous-échantillonnage SV de la rétine et les modifications à apporter en conséquence sur les interactions courtes. L'étape d'interactions longues reste identique (voir le paragraphe précédent), de même que les étapes de normalisation et de fusion pour combiner les différentes cartes d'énergie et obtenir la carte de saillance finale. L'objectif de cette implémentation est double. Il faut, d'une part, s'assurer de la validité des modifications apportées, et d'autre



(a) Masque “papillon” SI

(b) Masque “papillon” SV en x2

(c) Masque “papillon” SV en x3

(d) Masque “papillon” SV en x4

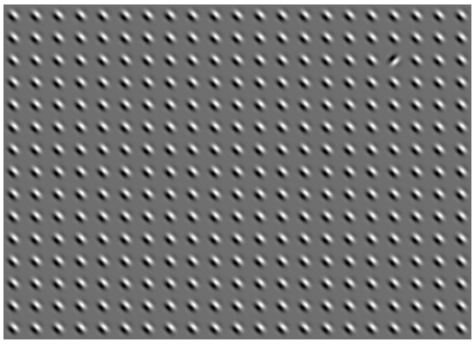
Fig. 6.17 – Exemple du masque “papillon” SV pour différentes positions comme dans le figure 6.14a : (a) Masque “papillon” SI, mais aussi masque “papillon” SV en x1 ; (b) Masque “papillon” SV en x2 ; (c) Masque “papillon” SV en x3 ; (d) Masque “papillon” SV en x4.

part, on s’attend à obtenir une diminution des niveaux de saillance en périphérie qui est sous-résolue par rapport à la zone centrale.

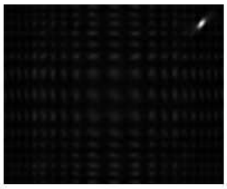
Les exemples de cartes de saillance pour les images sous-échantillonnées SV sont montrés à la figures 6.18. Le résultat obtenu correspond au résultat attendu. Il est intéressant de noter que le stimulus “distracteur” au centre est plus saillant que celui en périphérie selon la carte de saillance calculée à partir de l’image sous-échantillonnée SV (Fig. 6.18d). Cela est logique parce que le contraste de l’image sous-échantillonnée SV est moins fort en périphérie qu’au centre. Il y a moins d’information visuelle par sous-échantillonnage SV, pour les traitements à la périphérie.

3.5 Synthèse et Perspectives

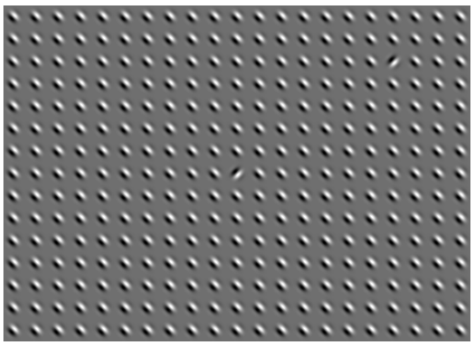
Nous avons modélisé dans cette partie la résolution variable de la rétine, représentée par la densité non uniforme des photorécepteurs et des cellules ganglionnaires. Cette résolution SV a été implémentée en utilisant un sous-échantillonnage à taux variable. Ce sous-échantillonnage modifie à son tour le comportement des filtres corticaux. Ainsi, lorsqu’on applique un banc de filtres SI sur une image sous-échantillonnée SV, cela correspond à un banc de filtres SV, dont la fréquence spatiale et l’orientation varient en fonction de l’excentricité par rapport au point de vue (correspondant à la fovéa). Plus on va à la périphérie, plus la déformation du filtre est importante.



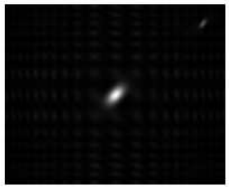
(a) Image initiale



(b) Carte de saillance



(c) Image initiale



(d) Carte de saillance

FIG. 6.18 – Exemples de carte de saillance avec des images sous-échantillonnées SV : (a) Image initiale avec le "distracteur" à la périphérie ; (b) Carte de saillance avec le "distracteur" à la périphérie ; (c) Image initiale avec un "distracteur" à la périphérie et l'autre au centre ; (d) Carte de saillance avec un "distracteur" à la périphérie et l'autre au centre. Nous avons utilisé 8×4 filtres LogNormaux comme dans le modèle décrit au chapitre 2. Le point de vue est au centre de l'image (fovéa). La taille de l'image initiale est de 300×418 pixels (soit $11.7° \times 16.3°$ d'angle visuel).

La variation des filtres corticaux avec la position sur l'image nécessite des modifications sur les traitements du modèle d'attention visuelle. Nous avons proposé de modifier les interactions courtes à partir des interactions courtes au chapitre 2 en tenant compte des relations entre les fréquences spatiales et les orientations voisines. Les autres étapes du modèle d'attention visuelle semblent peu influencées par le sous-échantillonnage SV et restent donc les mêmes comme au chapitre 2.

En suivant les travaux de Croll [Croll, 1998], les études sur l'échantillonnage spatialement variant de la rétine et son influence sur le filtrage cortical fournissent une démarche méthodologique pour modéliser les traitements au niveau du cortex visuel. En effet, le stimulus visuel est projeté sur le cortex suivant une projection rétinotopique dans laquelle le sous-échantillonnage à taux variable est une partie essentielle. Pour le modèle d'attention visuelle, le sous-échantillonnage à taux va-

riable permet d'affiner la construction d'une carte de saillance dans le cortex visuel en intégrant dans la modélisation la projection rétinotopique. Ces transformations d'espace fortement non linéaires vont avoir des implications importantes dans les différentes étapes de formation de la carte de saillance, comme on l'a déjà esquissé avec le seul sous-échantillonnage à taux variable. En particulier il sera nécessaire de revoir les étapes de normalisation et de fusion qui sont actuellement implémentées de manière spatialement uniforme et qui devraient tenir compte des positions spatiales.

Conclusions et Perspectives

Les études que nous avons menées durant cette thèse ont porté sur un modèle d'attention visuelle et se sont concentrées sur la question : "comment, et surtout, où les gens regardent-ils lorsqu'ils explorent le monde visuel environnant ?". Ces études suivent deux approches très complémentaires : l'une centrée sur la modélisation et l'autre sur l'expérience comportementale. D'autre part, ces études ont été menées dans un contexte de laboratoire en utilisant des images statiques présentées sur un écran d'ordinateur.

Dans un premier temps, nous avons proposé un modèle d'attention visuelle biologiquement plausible. Inspiré par la biologie du système visuel, notre modèle d'attention visuelle modélise les premiers étages de traitement depuis le fonctionnement des cellules rétiniennes jusqu'aux cellules du cortex visuel primaire. A partir de certaines caractéristiques visuelles de bas niveau, ce modèle permet pour une image particulière de lui associer sa carte de saillance ; autrement dit de faire ressortir les régions les plus susceptibles d'attirer le regard. Sous le guidage de certaines caractéristiques visuelles, le modèle permet de prédire les premières fixations d'un "sujet moyen" après la présentation d'une image.

Le modèle d'attention visuelle que nous avons proposé nous a permis d'étudier plus en détails un certain nombre de points dans l'étude des mouvements oculaires.

Premièrement, nous avons évalué les contributions relatives de plusieurs caractéristiques visuelles de bas niveau sur les mouvements oculaires, obtenus grâce à une expérience comportementale utilisant l'oculométrie, d'un grand nombre de sujets sur différentes images. Cette étude s'est faite selon deux méthodes statistiques ; elles permettent d'étudier et de quantifier la contribution à une carte de densité de fixations oculaires, de différentes cartes de caractéristique visuelle de bas niveau. Nous avons ainsi pu montrer que l'influence de la couleur est faible par rapport à celle de la luminance lors de l'exploration libre de scènes naturelles. Nous avons également montré que la luminance seule (et plus particulièrement, les contrastes en luminance) est suffisante pour prédire les premiers mouvements oculaires lors de l'exploration libre de scènes naturelles. De plus, nous avons observé une contribution importante du "biais de centralité" (les sujets ont tendance à regarder au centre des images), en particulier au début de la présentation d'une image (ce qui avait déjà été souvent observé dans la littérature).

Deuxièmement, le modèle d'attention visuelle est utilisé pour étudier "la programmation de saccade" lors de l'exploration de scènes naturelles. En effet, il a été

montré dans la littérature qu'à partir d'un même point de vue les deux saccades suivantes étaient programmées en parallèle ; cela a été montré dans des tâches bien particulière sur des stimuli simples. Nous avons voulu ici étudier cette question toujours dans le cadre de l'exploration libre des scènes naturelles. Dans cette étude, nous avons testé plusieurs stratégies de programmation de saccade en prédisant, à partir d'un point de vue, un nombre varié de saccades : soit les 4 saccades suivantes à partir d'un même point de vue, soit uniquement les 2 saccades suivantes, soit une seule saccade. Notre étude n'a pas confirmé une stratégie de programmation de deux saccades en parallèle pour des scènes naturelles. Nous avons ainsi montré que programmer une seule saccade à partir d'un point de vue est la stratégie la plus efficace et celle qui permet d'expliquer au mieux les données expérimentales réelles.

Troisièmement, à travers cette étude de programmation de saccade nous avons étudié l'influence de la décroissance de la densité des photorécepteurs ou des cellules ganglionnaires sur la distribution des amplitudes de saccades. Le modèle de programmation de saccade tenant compte de la décroissance de l'échelle des stimuli, liée à la décroissance de la densité des cellules ganglionnaires, montre une distribution des amplitudes de saccades plus proche de celle expérimentale. Il est donc indispensable d'intégrer la décroissance de l'échelle des stimuli dans les études de perception visuelle.

Enfin, nous avons approfondi le modèle d'attention visuelle en s'appuyant plus encore sur l'anatomie du système visuel. En effet, notre perception visuelle n'est pas immédiate mais suit un certain décours temporel. Nous avons donc modélisé le filtrage spatio-temporel de la rétine et testé son impact sur le modèle d'attention visuelle. Le nouveau modèle permet de tenir compte du décours temporel différent des sorties parvocellulaire et magnocellulaire de la rétine. Ainsi les caractéristiques spatiales de ces deux voies sont étroitement liées aux caractéristiques temporelles de l'information visuelle. Nous avons donc étudié ces deux voies lors de l'exploration d'une scène naturelle en nous servant des données expérimentales. Nous avons montré que pendant la fixation ces deux voies fonctionnent différemment : la voie parvocellulaire véhicule les informations en basses fréquences spatiales au début et puis en hautes fréquences spatiales en fin tandis que la voie magnocellulaire véhicule les informations en plus basses fréquences spatiales au début et s'atténue complètement pour ne plus rien transmettre en fin de fixation. Pendant la saccade ces deux voies véhiculent les informations en basses fréquences spatiales, représentant une décroissance de la sensibilité au contraste. De plus, le contrôle du gain pourrait expliquer l'inhibition saccadique sur la voie magnocellulaire. Ainsi, seule la voie parvocellulaire serait préservée durant la saccade.

Nous sommes allés également plus loin dans la modélisation de la rétine en examinant à nouveau l'échantillonnage spatialement variant causé par la densité non uniforme des cellules ganglionnaires. Nous avions dans un premier temps tenu compte de cette densité spatialement variante en la modélisant uniquement par un filtrage passe-bas spatialement variant. Nous avons voulu ici en plus ajouter l'échantillonnage spatialement variant. Un tel échantillonnage à taux variable conduit à la déformation de l'image, en particulier à la périphérie. Cette déformation in-

fluence à son tour les analyses en orientations et en fréquences spatiales réalisées au niveau du banc de filtres corticaux. Pour résoudre ce problème nous avons mené une étude théorique pour retrouver à partir du banc de filtres original le banc de filtres modifié permettant de tenir compte d'un échantillonnage à taux variable. La déformation d'un filtre augmente progressivement du centre à la périphérie de l'image. En examinant la déformation d'un filtre en fonction de la position sur l'image, nous avons également modifié les interactions. Ces études théoriques pourront fournir des éléments en vue de construire un modèle d'attention visuelle tenant compte de l'échantillonnage spatialement variant.

Nouveautés de la thèse

Notre thèse apporte certaines nouveautés par rapport à la littérature :
- Un traitement rétinien complet qui permet de prétraiter le signal pour le modèle d'attention visuelle.
- Une normalisation des cartes d'attributs en s'appuyant sur les énergies des différentes orientations pour faire ressortir les zones saillantes.
- Une évaluation quantitativement des contributions des caractéristiques visuelles de bas niveau aux mouvements oculaires avec une expérience oculométrique permettant de tester l'influence de la couleur vis-à-vis celle de la luminance.
- Une étude approfondie de différentes stratégies de programmation de saccade lors d'une exploration libre de scènes naturelles.
- Une étude théorique des traitements corticaux pour une image issue d'un échantillonnage spatialement variant.

Perspectives

Notre modèle d'attention visuelle s'appuie sur des caractéristiques visuelles de bas niveau. Il permet donc de ne prédire que les premières fixations, celles majoritairement guidées par des processus "Bottom-Up", après la présentation d'une image et lors de l'exploration libre. Ainsi ce modèle pourrait être amélioré en y intégrant des facteurs de haut niveau. Il pourrait relativement simplement être modifié dans des tâches bien précises par exemple de recherche de cible pour lesquelles la cible est définie par un certain nombre d'attributs et cette connaissance serait prises en compte au niveau du modèle d'attention visuelle. Navalpakkam a ainsi proposé une approche méthodologique pour intégrer les facteurs de haut niveau dans le modèle d'attention visuelle en modélisant la mémoire de travail et la mémoire à court terme du système visuel, combinée avec la détection et la reconnaissance d'objets. Cette méthode pourrait permettre d'étendre le modèle d'attention visuelle pour une tâche précise et donc d'augmenter la capacité de ce modèle tout en s'approchant du fonctionnement du système visuel.

En outre, dans un modèle tenant compte à la fois des facteurs de bas niveau et des facteurs de haut niveau, la question concernant la combinaison de ces facteurs reste encore ouverte. Des études de la littérature prônent souvent la multiplication pixel par pixel d'une carte de saillance guidée par des facteurs de bas niveau et

d'une carte de relevance guidée par les facteurs de haut niveau. Cette méthode ne fait ressortir que les zones qui sont saillantes à la fois dans la carte de saillance et dans la carte de relevance et pourrait omettre les zones saillantes qui n'existent que dans une des deux cartes même si la saillance de ces zones est importante. Au lieu de la combinaison multiplicative, nous pouvons également envisager une combinaison additive des facteurs de bas niveau et des facteurs de haut niveau. Dans ce cas, les pondérations des facteurs peuvent être déterminées à partir du modèle "EM carte". Ainsi, le modèle "EM carte" comporte maintenant un ensemble des facteurs de bas niveau et de haut niveau dont les pondérations sont trouvées en confrontant le modèle à des données réelles obtenues à partir d'une expérience avec un but précis.

La modélisation du traitement spatio-temporel de la rétine peut ouvrir plusieurs pistes sur les études de perception visuelle et sur les mouvements oculaires. Cela permet de simuler de manière plus complète l'évolution temporelle des voies rétiniennes suivant une succession de fixations et de saccades lors de l'exploration de scènes naturelles. De plus, en reprenant le traitement rétinien pendant la saccade, nous pourrions également aller plus loin et tenter de modéliser les traitements pendant les micro-saccades. Le rôle des micro-saccades ne fait pas l'unanimité dans la littérature. Certaines études ont avancé que les micro-saccades ont pour objectif de rafraîchir les stimuli visuels tandis que les autres ont supposé que les micro-saccades permettent d'égaliser le spectre des scènes naturelles dans les hautes fréquences spatiales. Ainsi une modélisation du fonctionnement biologique de la rétine pendant les micro-saccades pourrait permettre d'éclairer le rôle des micro-saccades lors de l'exploration de scènes naturelles. Une telle modélisation permet d'ouvrir des pistes d'études sur ces mouvements.

Nous avons enfin abordé l'échantillonnage spatialement variant, lié à la densité non uniforme des cellules ganglionnaires, et puis son impact sur le filtrage cortical. Cet échantillonnage peut maintenant être intégré dans la modélisation du traitement spatio-temporel de la rétine. En effet, on sait qu'il y a environ 105 millions photorécepteurs recevant le flux visuel entrant dans la rétine mais qu'il n'y a que 1,2 à 1,5 millions cellules ganglionnaires à la sortie de la rétine, permettant de transmettre l'information visuelle au cortex visuel. Ainsi, une cellule ganglionnaire reçoit en moyenne 100 photorécepteurs. En combinant l'échantillonnage spatialement variant dans le traitement spatio-temporel de la rétine pour une image, on peut obtenir en sortie de la rétine des voies visuelles de taille plus petite que celle de l'image en entrée ; cela permet de modéliser de manière plus réaliste le fonctionnement de la rétine.

Ce traitement de la rétine peut être poursuivi par le traitement au niveau du cortex visuel primaire que nous avons introduit au chapitre précédent afin de construire un modèle d'attention visuelle assez complète, inspiré du fonctionnement biologique du système visuel. De plus, en considérant l'aspect dynamique du modèle, nous pourrons également envisager une carte de saillance spatio-temporelle même lors de l'exploration des scènes statiques. Dans ce cas, la carte de saillance sera construite au cours du temps à partir du traitement spatio-temporel de la rétine. Cette carte devrait prendre en compte les évolutions temporelles de l'information visuelle véhiculée

différemment par la voie parvocellulaire et par la voie magnocellulaire. Un tel modèle d'attention visuelle pourrait permettre d'expliquer la saillance présente à chaque instant dans le cortex visuel lors de l'exploration de scènes et de comprendre au mieux les mécanismes sous jacents à l'attention visuelle.

Annexe

A Masque “papillon”

Le masque “papillon” B_{ij} pour l'orientation i et la fréquence j comporte deux parties : l'une pour modéliser les phénomènes d'excitation (B_{ij}^{+}) et l'autre pour l'inhibition (B_{ij}^{-}) (Eq. A1). L'orientation de la partie excitatrice constitue l'orientation du masque. La partie inhibitrice se situe pour les orientations perpendiculaires. La partie excitatrice ou inhibitrice s'effectue de manière similaire en combinant deux fonctions séparées B_i^{θ} et B_j^{r}. A titre d'exemple, l'équation A2 montre la partie excitatrice.

$$B_{ij} = B_{ij}^{+} - B_{ij}^{-} \tag{A1}$$

$$B_{ij}^{+} = B_j^{r}(r) \cdot B_i^{\theta}(\theta) \tag{A2}$$

où

$$B_j^{r}(r) = \begin{cases} 1 & \text{si } r \leq r_j \\ \exp\left\{-0.5\frac{r^2}{\sigma_j^2}\right\} & \text{sinon} \end{cases}$$

$$B_i^{\theta}(\theta) = \begin{cases} \cos\left(\frac{\theta-\theta_i}{\alpha}\pi\right) & \text{si } |\theta-\theta_i| \leq \frac{\alpha}{2} \\ 0 & \text{sinon} \end{cases}$$

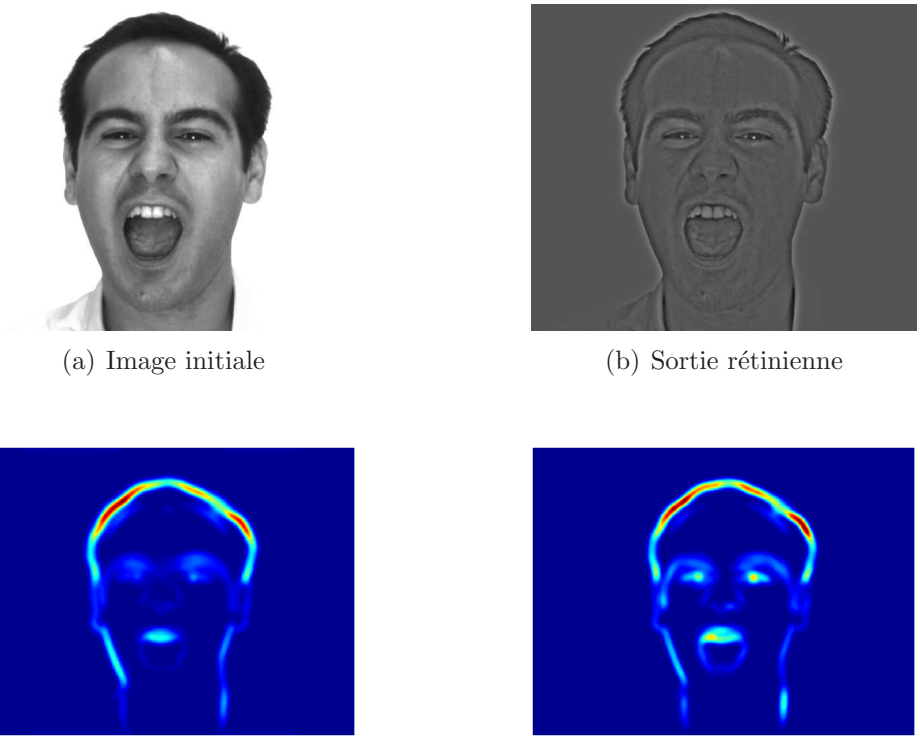
avec r_j le rayon du masque B_{ij}, $\sigma_j = \frac{r_j}{4}$, θ_i l'orientation du masque, $\alpha = \frac{180}{N}$, N le nombre d'orientations.

Le masque se fait par la combinaison de ces deux parties : excitatrice et inhibitrice. Elles sont fusionnées de sorte que : $\sum_{x,y} B_{ij}^{-}(x,y) = \frac{1}{5}\sum_{x,y} B_{ij}^{+}(x,y)$. Le masque est normalisé pour que la somme de toutes les valeurs dans le masque soit égale à 1.

La taille d'un masque varie selon les fréquences spatiales du banc de filtres. Concrètement, la taille du masque “papillon” utilisé pour la convolution avec une carte d'énergie est égale à l'inverse de la fréquence spatiale de cette carte.

B L'avantage du traitement rétinien dans le modèle d'attention visuelle

Alors que le traitement de la rétine permet de renforcer les contrastes des stimuli, il peut faire ressortir des objets saillants qui ont de faibles contrastes dans l'image originale. La figure B1 montre l'avantage de l'intégration du traitement rétinien dans le modèle d'attention visuelle. La carte de saillance utilisant le traitement rétinien (Fig. B1d) permet de faire ressortir les zones saillantes liées aux yeux qui ont de faibles contrastes dans l'image initiale. Ce n'est pas le cas pour la carte de saillance sans l'utilisation du traitement rétinien (Fig. B1c).



(a) Image initiale (b) Sortie rétinienne

(c) Carte de saillance sans traitement rétinien (d) Carte de saillance avec traitement rétinien

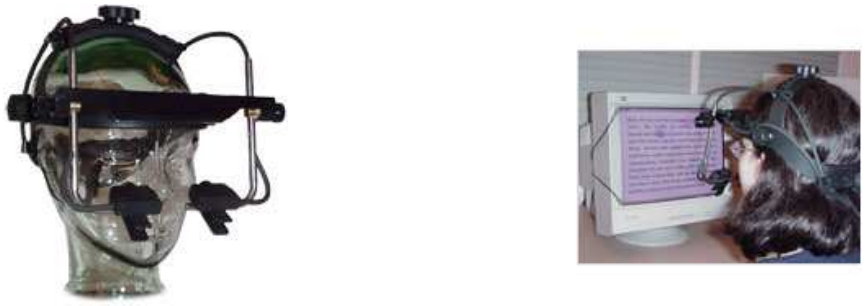
FIG. B1 – Le rôle du traitement rétinien dans le modèle d'attention visuelle. Le traitement rétinien permet de mettre en exergue les deux yeux qui sont de faibles contrastes dans l'image initiale.

C Dispositif d'enregistrement

Nous avons utilisé l'oculomètre Eyelink II de SR Research pour enregistrer les mouvements oculaires des sujets (*http ://sr-research.com/EL_II.html*). L'oculomètre est un casque muni de deux caméras permettant d'enregistrer l'image de chaque œil. De plus, à côté de ces deux caméras des lumières infra-rouge permettent d'envoyer un faisceau infra-rouge et de récupérer la réflexion de ces signaux infra rouge par la cornée. Les images de deux yeux sont enregistrées et grâce à une phase de calibration et à la segmentation de la pupille nous arrivons à récupérer la position des yeux sur l'écran. De plus, le logiciel fourni avec l'oculomètre permet à partir des positions oculaires enregistrées toutes les 4 ms de calculer les fixations et les saccades correspondantes. C'est ces données que nous analysons.

D Images en couleur

Les images utilisées dans l'expérience au chapitre 4 proviennent de la base d'images naturelles de Kodak (`http://www.cipr.rpi.edu/resource/stills/kodak.html`) et sont présentées à la figure D1.



(a) (b)

FIG. C1 – Un casque permet d'enregistrer des fixations lors de l'expérience de mouvements oculaires (*http ://sr-research.com/EL_II.html*). (a) Un casque ; (b) Un sujet muni d'un casque lors d'une expérience oculométrique.

E Implémentation d'un filtre récursif

Donné un filtre récursif de type passe-bas sous la forme :

$$H(z) = (1-a)^2 \cdot \frac{1}{1-az} \cdot \frac{1}{1-az^{-1}} \tag{E1}$$

Il est réécrit :

$$H(z) = \frac{1-a}{1+a}\left[\frac{1}{1-az} + \frac{1}{1-az^{-1}} - 1\right] \tag{E2}$$

Dans le domaine spatial, la sortie $y(n)$ du filtrage pour le signal d'entrée $x(n)$ est calculée comme suivant :

$$y(n) = \frac{1-a}{1+a}\left[y_1(n) + y_2(n) - x(n)\right] \tag{E3}$$

avec $y_1(n) = x(n) + a \cdot y_1(n+1)$
et $y_2(n) = x(n) + a \cdot y_2(n-1)$
$y_1(n)$ et $y_2(n)$ correspondent respectivement à un filtre anticausal et un filtre causal.

F Modélisation de la distribution des amplitudes de saccades

F1 La distribution exponentielle

La modélisation de la distribution des amplitudes de saccades a été effectuée par Bahill [Bahill et al., 1975]. Les auteurs ont montré que la distribution des amplitudes de saccades a la forme d'une distribution exponentielle dont la valeur moyenne correspond à une amplitude de 7.6°. Dans [Andrews and Coppola, 1999], ils ont retrouvé une amplitude moyenne de saccade également proche de 7.6°.

FIG. D1 – Les images en couleur utilisées dans l'expérience au chapitre 4.

Nous modélisons par une distribution exponentielle la distribution des amplitudes de saccades obtenues à partir des données expérimentales. Une distribution exponentielle dépend d'un seul paramètre λ :

$$f(y) = \frac{1}{\lambda} \exp\left(-\frac{y}{\lambda}\right) \tag{F1}$$

avec y l'amplitude de saccade
λ paramètre de la distribution représentant l'amplitude moyenne de saccade.

En utilisant la méthode de maximum de vraisemblance pour une distribution exponentielle, on trouve $\lambda = \frac{1}{N}\sum_{i=1}^{N} y_i$ où $\{y_i\}_{i=1..N}$ est un échantillon de taille N d'amplitudes de saccades. A partir de nos données expérimentales, nous obtenons $\lambda = 6.89$. Cette valeur est très proche de celle trouvée par Bahill [Bahill et al., 1975].

Cependant, la distribution exponentielle ne peut pas refléter complètement la forme de la distribution des amplitudes de saccades obtenues à partir des données expérimentales. La distribution exponentielle a un maximum à zéro et décroît avec l'excentricité. Quant à la distribution des amplitudes de saccades expérimentales, elle augmente pour atteindre un maximum vers une amplitude d'environ 2° avant de décroître (Fig. F1). C'est pourquoi nous allons utiliser une autre distribution théorique à 2 paramètres, la distribution Gamma, pour modéliser la distribution des amplitudes de saccades.

F2 La distribution Gamma

La forme de la distribution des amplitudes de saccades expérimentales peut faire penser à une distribution Gamma. Cette dernière dépend de deux paramètres : k représentant l'allure de la distribution et θ l'échelle :

$$f(y, k, \theta) = y^{k-1} \frac{e^{-\frac{y}{\theta}}}{\theta^k \Gamma(k)} \tag{F2}$$

avec $y, k, \theta > 0$
et $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$, $\Gamma(k) = (k-1)!$ si k entier.

Nous remarquons que $f(y, k = 1, \theta)$ correspond à la distribution exponentielle.

De la même manière que pour la distribution exponentielle, nous utilisons l'estimation du maximum de vraisemblance pour estimer les paramètres de la loi Gamma, k et θ. Après calculs, nous obtenons :

$$\ln(k) - \Psi(k) = \ln\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) - \frac{1}{N}\sum_{i=1}^{N} \ln(y_i) \tag{F3}$$

avec $\Psi(k) = \frac{\Gamma'(k)}{\Gamma(k)}$

Selon [Muqattash and Yahdi, 2006], on peut approximer $\Psi(k) = \ln(k) - \frac{1}{2k} - \frac{1}{12k^2}$

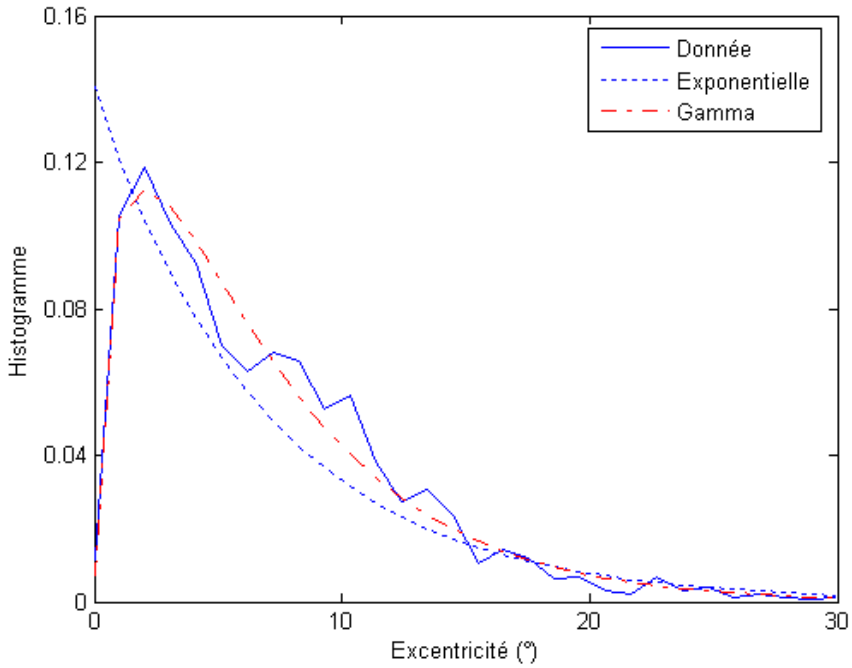


FIG. F1 – Distribution des amplitudes de saccades à partir des données expérimentales et sa modélisation par une distribution exponentielle et par une distribution Gamma.

TAB. F1 – Erreur quadratique des deux distributions

	Distribution exponentielle	Distribution Gamma
Erreur quadratique	0.0198	0.0011

En posant $s = \ln\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) - \frac{1}{N}\sum_{i=1}^{N}\ln(y_i)$, nous trouvons :

$$k = \frac{3 + \sqrt{9 + 12s}}{12s} \tag{F4}$$

puis,

$$\theta = \frac{1}{kN}\sum_{i=1}^{N} y_i \tag{F5}$$

Avec nos données expérimentales, nous obtenons $k = 1.4146$ et $\theta = 4.8707$. Selon la figure F1, la distribution Gamma permet de mieux modéliser les données expérimentales que la distribution exponentielle. Si nous regardons l'erreur quadratique présentée dans le tableau F1, la distribution Gamma donne une erreur bien moins importante.

La distribution des amplitudes de saccades expérimentales est caractérisée par son mode et son étendue. Alors que la distribution exponentielle a un seul paramètre qui joue sur l'étendue, elle ne peut pas modéliser complètement la distribution des

amplitudes de saccades. En revanche, la distribution Gamma dispose de deux paramètres qui peuvent ajuster à la fois la position du mode et l'étendue de la distribution. Ainsi, la distribution Gamma permet de mieux modéliser la distribution des amplitudes de saccades.

G Transformée de Fourier de la réponse impulsionnelle du filtre SV

Ici nous voulons chercher la transformée de Fourier de la réponse impulsionnelle du filtre SV. En reprenant la réponse impulsionnelle du filtre SV à l'équation 6.20, on a :

$$h_{\mathbf{x}}(\mathbf{x}-\boldsymbol{\xi}) = g(P(\mathbf{x})-P(\boldsymbol{\xi}))\Big|\det J_{P(\boldsymbol{\xi})}\Big| \tag{G1}$$

La fonction de compression $P(\boldsymbol{\xi})$ peut se linéariser en appliquant la formule de Taylor autour de $\mathbf{x}$:

$$P(\boldsymbol{\xi}) = P(\mathbf{x}) + J_{P(\mathbf{x})}(\boldsymbol{\xi}-\mathbf{x}) \tag{G2}$$

La réponse implusionnelle SV est alors :

$$h_{\mathbf{x}}(\mathbf{x}-\boldsymbol{\xi}) = g(J_{P(\mathbf{x})}(\mathbf{x}-\boldsymbol{\xi}))\Big|\det J_{P(\boldsymbol{\xi})}\Big| \tag{G3}$$

Si on considère le Jacobien $J_{P(\boldsymbol{\xi})}$ constante sur le support de la réponse impulsionnelle SV, alors on peut poser $J_{P(\boldsymbol{\xi})} = J_{P(\mathbf{x})}$. Ainsi, la réponse impulsionnelle SV devient :

$$h_{\mathbf{x}}(\mathbf{x}-\boldsymbol{\xi}) = g(J_{P(\mathbf{x})}(\mathbf{x}-\boldsymbol{\xi}))\Big|\det J_{P(\mathbf{x})}\Big| \tag{G4}$$

En changeant $(\mathbf{x}-\boldsymbol{\xi})$ par $\boldsymbol{\xi}$, on obtient :

$$h_{\mathbf{x}}(\boldsymbol{\xi}) = g(J_{P(\mathbf{x})}\boldsymbol{\xi})\Big|\det J_{P(\mathbf{x})}\Big| \tag{G5}$$

Or, la transformée de Fourier de $g(J_{P(\mathbf{x})}\boldsymbol{\xi})$ est la suivante :

$$\mathcal{F}\left\{g(J_{P(\mathbf{x})}\boldsymbol{\xi})\right\} = \frac{G\Big(\Big[J_{P(\mathbf{x})}^{-1}\Big]^T\mathbf{f}\Big)}{\Big|\det J_{P(\mathbf{x})}\Big|} \tag{G6}$$

D'où, la transformée de Fourier $H_{\mathbf{x}}(\mathbf{f})$ de $h_{\mathbf{x}}(\boldsymbol{\xi})$ est alors :

$$H_{\mathbf{x}}(\mathbf{f}) = G\Big(\Big[J_{P(\mathbf{x})}^{-1}\Big]^T\mathbf{f}\Big) \tag{G7}$$

H Fréquence centrale et orientation du filtre SV

Avec un filtre LogNormal SI, la réponse fréquentielle du filtre SV est la suivante en reprenant l'équation 6.22 :

$$H_{\mathbf{x}}(\mathbf{f}) = \exp\left\{-\frac{\Big[\log\big(\frac{\|J_{P(\mathbf{x})}^{-1}\mathbf{f}\|}{f_0}\big)\Big]^2}{2\sigma_f^2}\right\}\exp\left\{-\frac{\big[\Phi(J_{P(\mathbf{x})}^{-1}\mathbf{f})-\theta_0\big]^2}{2\sigma_\theta^2}\right\} \tag{H1}$$

La fréquence centrale et l'orientation du filtre SV peuvent être déterminées en se basant sur la localisation dans le domaine fréquentiel où la réponse fréquentielle $H_{\mathbf{x}}(\mathbf{f})$ du filtre SV atteint l'amplitude maximale. $H_{\mathbf{x}}(\mathbf{f})$ est maximale, alors :

$$\begin{cases} \|J_{P(\mathbf{x})}^{-1}\mathbf{f}\| = f_0 \\ \Phi(J_{P(\mathbf{x})}^{-1}\mathbf{f}) = \theta_0 \end{cases} \tag{H2}$$

ou :

$$\begin{cases} \frac{f_1^2}{J_{11}^2} + \frac{f_2^2}{J_{22}^2} = f_0^2 \\ \frac{f_2}{f_1} = \frac{J_{22}}{J_{11}}\tan(\theta_0) \end{cases} \tag{H3}$$

D'où, on peut trouver :

$$\begin{cases} f_1 = f_0 J_{11}\cos(\theta_0) \\ f_2 = f_0 J_{22}\sin(\theta_0) \end{cases} \tag{H4}$$

Ainsi, la fréquence centrale f_m(son module) et l'orientation θ_m du filtre SV sont les suivantes :

$$\begin{cases} \theta_m = \arctan\left(\frac{f_2}{f_1}\right) = \arctan\left(\frac{J_{22}}{J_{11}}\tan(\theta_0)\right) \\ f_m = \sqrt{f_1^2 + f_2^2} = f_0\sqrt{J_{11}^2\cos(\theta_0)^2 + J_{22}^2\sin(\theta_0)^2} \end{cases} \tag{H5}$$

Bibliographie

D. Alleysson. *Le traitement du signal chromatique dans la rétine : un modèle de base pour la perception humaine des couleurs*. PhD thesis, Université Joseph Fourier, 1999.

T. J. Andrews and D. M. Coppola. Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. In *Vis Res*, volume 39, pages 2947–2953, 1999.

J. R. Antes. The time course of picture viewing. In *Journal of Experimental Psychology*, volume 103(1), pages 62–70, 1974.

R. Baddeley. Searching for filters with 'interesting' output distributions : An uninteresting direction to explore ? In *Network Computation in Neural Systems*, volume 7(2), pages 409–421, 1996.

R. Baddeley. The correlational structure of natural images and the calibration of spatial representations. In *Cogn. Sci.*, volume 21, pages 351–372, 1997.

R. J. Baddeley and B. W. Tatler. High frequency edges (but not contrast) predict where we fixate : a bayesian system identification analysis. In *Vision Research*, volume 46, pages 2824–2833, 2006.

A. T. Bahill, D. Adler, and L. Stark. Most naturally occurring human saccades have magnitudes of 15 degrees or less. In *Investigative Ophthalmology & Visual Science*, volume 14, pages 468–469, 1975.

W. Beaudot, P. Palagi, and J. Hérault. Realistic simulation tool for early visual processing including space, time and colour data. In *International Workshop on Artificial Neural Networks, LNCS*, volume 686, pages 370–375, Barcelona, 1993. Springer-Verlag.

W. H. A. Beaudot. *Le traitement neuronal de ll'information dans la rétine des vertébrés : Un creuset d'idées pour la vision artificielle*. PhD thesis, INPG, 1994.

W.H.A. Beaudot and K.T. Mullen. Orientation selectivity in luminance and color vision assessed using 2-d bandpass filtered spatial noise. In *Vision Research*, volume 45(6), pages 687–696, 2005.

W. Becker and R. Jürgens. An analysis of the saccadic system by means of double step stimuli. In *Vision Research*, volume 19, pages 967–983, 1979.

C. Biernacki and R. Mohr. Indexation et appariement d'images par modèle de mélange gaussien des couleurs. In *GRETSI'99, 17ème colloque GRETSI*, pages 291–294, 1999.

J. Braun. Contour salience and striate cortex : A new model matches human sensitivity. In *Investigative Ophthalmology & Visual Science*, volume 40(4), pages S780–S780, 1999.

J. Bullier. Neural basis of vision. In H. Pashler, editor, *Steven's handbook of experimental psychology*. John Wiley & Sons. Inc., 2002.

G. T. Buswell. *How people look at pictures : A study of the psychology of perception in art*. Chicago :University of Chicago Press, 1935.

G.T. Buswell. *Fundamental reading habits : A study of their development*. Chicago, IL : University of Chicago Press, 1922.

G.T. Buswell. *How adults read*. Chicago, IL : University of Chicago Press, 1937.

B. Cassin and S. Solomon. *Dictionary of Eye Terminology*. Gainsville, Florida : Triad Publishing Company, 1990.

M. S. Castelhano and J. M. Henderson. The influence of color on the activation of scene gist. In *Journal of Experimental Psychology : Human Perception and Performance*, volume 34, pages 660–675, 2008.

M. S. Castelhano, M. L. Mack, and J. M. Henderson. Viewing task influences eye movement control during active scene perception. In *Journal of Vision*, volume 9(3) :6, pages 1–15, 2009.

S. Chatterjee and E. M. Callaway. Parallel colour-opponent pathways to primary visual cortex. In *Nature*, volume 426, pages 668–671, 2003.

A. Chauvin. *Perception des scènes naturelles : Etude et simulation du rôle de l'amplitude, de la phase et de la saillance dans la catégorisation et l'exploration des scènes naturelles*. PhD thesis, Université Pierre Mendès-France, Grenoble, 2003.

C. Croll. *Synthèse d'un banc de filtres spatialement variants, inspiré des colonnes d'orientations du système visuel*. PhD thesis, INPG, Grenoble, France, 1998.

C. A. Curcio, K. R. Sloan, R. E. Kalina, and K. E. Hendrickson. Human photoreceptor topography. In *J Comp Neurol*, volume 292, pages 497–523, 1990.

D. M. Dacey. Circuitry for color coding in the primate retina. In *Proc Natl Acad Sci USA*, volume 93, pages 582–588, 1996.

D. M. Dacey and O. S. Packer. Colour coding in the primate retina. diverse cell types and cone-specific circuitry. In *Curr Opin Neurobiol*, volume 13, pages 421–427, 2003.

A. Das and C. Gilbert. Topography of contextual modulations mediated by short-range interactions in primary visual cortex. In *Nature*, volume 399(6737), pages 655–661, 1999.

I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

J. G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. In *Vision Research*, volume 20, pages 847–856, 1980.

G. DeAngelis, J. Robson, I. Ohzawa, and R. Freeman. Organization of suppression in receptive-fields of neurons in cat visual-cortex. In *Journal of Neurophysiology*, volume 68(1), pages 144–163, 1992.

A. Delorme and M. Flückiger. *Perception et réalité - Une introduction à la psychologie des perceptions*. De Boeck, 2003.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. In *Journal of the Royal Statistical Society. Series B (Methodological)*, volume 39(1), pages 1–38, 1977.

A. M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. In *The Journal of Physiology*, volume 357, pages 241–265, 1984.

R. L. DeValois, D. G. Albrecht, and L. G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. In *Vision Research*, volume 22, pages 545–559, 1982.

N. Dhavale and L. Itti. Saliency-based multi-foveated mpeg compression. In *Proc. IEEE Seventh International Symposium on Signal Processing and its Applications*, pages 229–232, Paris, France, 2003.

M.R. Diamond, J. Ross, and M.C. Morrone. Extraretinal control of saccadic suppression. In *J. Neurosci.*, volume 20, 2000.

N. Drasdo and C. W. Fowler. Non-linear projection of the retinal image in a wide-angle schematic eye. In *Br J Ophthalmol*, volume 58(8), pages 709–714, 1974.

B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. Chapman and Hall, New York, 1993.

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

D. J. Field. What is the goal of sensory coding ? In *Neural Comput.*, volume 6, pages 559–601, 1994.

J. M. Findlay and R. Walker. A model of saccade generation based on parallel processing and competitive inhibition. In *Behav. Brain Sci.*, volume 22, pages 661–721, 1999.

H. P. Frey, C. Honey, and P. König. What's color got to do with it ? the influence of color on visual attention in different categories. In *Journal of Vision*, volume 8(14) :6, pages 1–17, 2008.

D. A. Gajewski, A. M. Pearson, M. L. Mack, F. N. Bartlett, and J. M. Henderson. Human gaze control in real world search. In *Attention and Performance in Computational Vision*, volume 3368, pages 83–99, 2005.

W. S. Geisler, J. S. Perry, and J. Najemnik. Visual search : The role of peripheral information measured using gaze-contingent displays. In *Journal of Vision*, volume 6(9), pages 858–873, 2006.

W.S. Geisler and J.S. Perry. A real-time foveated multi-resolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging, SPIE Proceedings*, volume 3299, pages 294–305, 1998.

A. K. Goodchild, K. K. Ghosh, and P. R. Martin. Comparison of photoreceptor spatial density and ganglion cell morphology in the retina of human, macaque monkey, cat, and the marmoset callithrix jacchus. In *The Journal of Comparative Neurology*, volume 366, pages 55–75, 1996.

J. P. Gottlieb, M. Kusunoki, and M. E. Goldberg. The representation of visual salience in monkey parietal cortex. In *Nature*, volume 391, pages 481–484, 1998.

A. Guérin-Dugué, C. Biernacki, and J. Hérault. Statistical modelling for image retrieval using a biological model of the perceptive colour space. In *International Conference on Image Processing*, volume 1, pages 209–212, 2001.

N. Guyader. *Scènes visuelles : Catégorisation basée sur des modèles de perception.* PhD thesis, Université Joseph Fourier, Grenoble, France, 2004.

T. Hansen, W. Sepp, and H. Neumann. Recurrent long-range interactions in early vision. In S. Wermter, J. Austin, and D. Willshaw, editors, *Emergent Neural Computational Architectures Based on Neuroscience*, volume 2036, pages 139–153, 2001.

J. M. Henderson. Visual attention and saccadic eye movements. In G. d'Ydewalle and J. Van Rensbergen, editors, *Perception and cognition : Advances in eye movement research*, pages 37–50. North-Holland, Amsterdam, 1993.

J. M. Henderson. Human gaze control during real-world scene perception. In *Trends in Cognitive Science*, volume 7(11), pages 498–504, 2003.

J. M. Henderson and A. Hollingworth. High-level scene perception. In *Annual Review of Psychology*, volume 50, pages 243–271, 1999.

J. M. Henderson, Jr. Weeks, P. A., and A. Hollingworth. Effects of semantic consistency on eye movements during scene viewing. In *Journal of Experimental Psychology : Human Perception and Performance*, volume 25, pages 210–288, 1999.

T. Ho-Phuoc, A. Guérin-Dugué, and N. Guyader. A biologically-inspired visual saliency model to test different strategies of saccade programming. In A. Fred, J. Filipe, and H. Gamboa, editors, *BIOSTEC 2009, CCIS*, volume 52, pages 187–199, 2010.

J. E. Hoffman. Visual attention and eye movements. In H. Pashler, editor, *Attention*, pages 119–154. University College London Press, London, 1998.

I.T. Hooge and C.J. Erkelens. Adjustment of fixation duration in visual search. In *Vision Research*, volume 38(9), pages 1295–1302, 1998.

J. Hérault. De la rétine biologique aux circuit neuromorphiques. In J. M. Jolion, editor, *Les Systèmes de Vision*. Hermès, 2001.

J. Hérault. Retinal sampling. Rapport interne, GIPSA-Lab, 2009.

D. H. Hubel. Evolution of ideas on the primary visual cortex, 1955-1978 : a biased historical account. 1981. Paper presented at the Nobel Lecture.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. In *J. Physiol. (Lond.)*, volume 160, pages 106–154, 1962.

D. H. Hubel, T. N. Wiesel, and M. P. Stryker. Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. In *Nature*, volume 269(5626), pages 328–330, 1977.

L. Itti. Quantitative modeling of perceptual salience at human eye position. In *Visual Cognition*, volume 14(4-8), pages 959–984, 2006.

L. Itti and P. F. Baldi. Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems*, volume 19, pages 547–554, 2006.

L. Itti and C. Koch. Feature combination strategies for saliency-based visual attention systems. In *Journal of Electronic Imaging*, volume 10(1), pages 161–169, 2001.

L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 20(11), pages 1254–1259, 1998.

E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of Neural Science*. McGraw-Hill, New York, 4th edition, 2000. ISBN 0-8385-7701-6.

R. M. Klein. Inhibition of return. In *Trends in Cognitive Sciences*, volume 4, pages 138–147, 2000.

H. Knutsson, C. F. Westin, and G. Granlund. Local multiscale frequency and bandwidth estimation. In *IEEE international conference on image processing (ICIP94)*, Austin, TX, 1994.

C. Koch and S. Ullman. Shifts in selective visual attention : Towards the underlying neural circuitry. In *Human Neurobiology*, volume 4(4), pages 219–227, 1985.

K. Koch, J. McLean, R. Segev, M. A. Freed, M. J. Berry, V. Balasubramanian, and P. Sterling. How much the eye tells the brain. In *Current Biology*, volume 16(14), pages 1428–1434, 2006.

P. Kowaliski. *Vision et mesure de la couleur*. Masson, 2 edition, 1990. actualisée par F. Viénot et R. Sève.

G. Krieger, I. Rentschler, G. Hauske, K. Schill, and C. Zetzsche. Object and scene analysis by saccadic eye-movements : an investigation with higher-order statistics. In *Spatial Vision*, volume 13(2-3), pages 201–214, 2000.

S. Kullback and R. A. Leibler. On information and sufficiency. In *The Annals of Mathematical Statistics*, volume 22(1), pages 79–86, 1951.

M.F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities ? In *Vision Res*, volume 41, pages 3559–3565, 2001.

F. Landragin. Saillance physique et saillance cognitive. In *Corela*, volume 2(2), 2004.

P. Le Callet. *Critères objectifs avec référence de qualité visuelle des images couleur*. PhD thesis, Université de Nantes, 2001.

O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28(5), pages 802–817, 2006.

D. Lee, L. Itti, C. Koch, and J. Braun. Attention activates winner-take-all competition among visual filters. In *Nature Neuroscience*, volume 2(4), pages 375–381, 1999.

M. W. Levine and J. M. Shefner. *Fundamentals of sensation and perception*. Brooks/Cole, Pacific Grove, CA, 2nd edition, 1991.

J. Lupianez, R. M. Klein, and P. Bartolomeo. Inhibition of return : Twenty years after. In *Cognitive neuropsychology*, volume 23(7), pages 1003–1014, 2006.

M. A. Mahowald and C. A. Mead. Une rétine en silicium. In *Pour la Science*, volume 165, pages 50–57, 1991.

S. K. Mannan, K. H. Ruddock, and D. S. Wooding. Fixation sequences made during visual examination of briefly presented 2d images. In *Spatial Vision*, volume 11(2), pages 157–178, 1997.

J. L. Mannos and D. J. Sakrison. The effects of a visual fidelity criterion on the encoding of images. In *IEEE Transactions of Information Theory*, volume 20(4), pages 525–535, 1974.

C. Massot and J. Hérault. Model of frequency analysis in the visual cortex and the shape from texture problem. In *IJCV*, volume 76(2), pages 165–182, 2008.

J. A. Mazer and J. L. Gallant. Goal-related activity in v4 during free viewing visual search. evidence for a ventral stream visual salience map. In *Neuron*, volume 40, pages 1241–1250, 2003.

W. H. McIlhagga and K. T. Mullen. Contour integration with colour and luminance contrast. In *Vision Research*, volume 36(9), pages 1265–1279, 1996.

R. M. McPeek, V. Maljkovic, and K. Nakayama. Saccades require focal attention and are facilitated by a short-term memory system. In *Vision Research*, volume 39 (8), pages 1555–1566, 1999.

R.M. McPeek, A.A. Skavenski, and K. Nakayama. Concurrent processing of saccades in visual search. In *Vision Research*, volume 40(18), pages 2499–2516, 2000.

C. A. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, 1989.

C. A. Mead and M. A. Mahowald. A silicon model for early visual processing. In *Neural Networks*, volume 1(1), pages 91–97, 1988.

F. Miau and L. Itti. A neural model combining attentional orienting to object recognition : Preliminary explorations on the interplay between where and what. In *Proc. IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 789–792, Istanbul, Turkey, 2001.

R. E. Morrison. Manipulation of stimulus onset delay in reading : evidence for parallel programming of saccades. In *Journal of Experimental Psychology : Human Perception and Performance*, volume 10, pages 667–682, 1984.

B. C. Motter and E. J. Belky. The guidance of eye movements during active visual search. In *Vision Research*, volume 38(12), pages 1805–1815, 1998.

J. A. Movshon. The hypercomplexities in the visual cortex. In *Nature*, volume 272(23), pages 305–306, 1978.

D. P. Munoz, J. R. Broughton, J. E. Goldring, and I. T. Armstrong. Age-related performance of human subjects on saccadic eye movement tasks. In *Exp Brain Res*, volume 121, pages 391–400, 1998.

I. Muqattash and M. Yahdi. Infinite family of approximations of the digamma function. In *Mathematical and Computer Modelling*, volume 43, pages 1329–1336, 2006.

K. Naka and W. Rushton. S-potentials from colour units in the retina of fish (cyprinidae). In *J Physiol (Lond)*, volume 185, pages 536–555, 1966.

V. Navalpakkam and L. Itti. Modeling the influence of task on attention. In *Vision Research*, volume 45(2), pages 205–231, 2005.

D. Navon. Forest before the trees : The precedence of global features in visual perception. In *Cognitive Psychology*, volume 9(3), pages 353–383, 1977.

T. Ogawa and H. Komatsu. Target selection in area v4 during a multidimensional visual search task. In *J. Neurosci.*, volume 24(28), pages 6371–6382, 2004.

A. Oliva and A. Torralba. Modeling the shape of the scene : a holistic representation of the spatial envelope. In *Int. J. Comput. Vis.*, volume 42, pages 145–75, 2001.

A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson. Top-down control of visual attention in object detection. In *IEEE Proceedings of the International Conference on Image Processing*, volume 1, pages 253–256, 2003.

O. Osterberg. Topography of the layer of rods and cones in the human retina. In *Acta Ophthalmologica*, volume 13, pages 1–97, 1935. (Supplement 6).

S. Pannasch, J. R. Helmert, K. Roth, A-K. Herbold, and H. Walter. Visual fixation durations and saccade amplitudes : Shifting relationship in a variety of conditions. In *Journal of Eye Movement Research*, volume 2(2) : 4, pages 1–19, 2008.

D. J. Parkhurst and E. Niebur. Scene content selected by active vision. In *Spatial Vision*, volume 16(2), pages 125–154, 2003.

D. J. Parkhurst and E. Niebur. Texture contrast attracts overt visual attention in natural scenes. In *European Journal of Neuroscience*, volume 19, pages 783–789, 2004.

D. J. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocation of overt visual attention. In *Vision Research*, volume 42(1), pages 107–123, 2002.

J. B. Pelz and R. Canosa. Oculomotor behavior and perceptual strategies in complex tasks. In *Vision Research*, volume 41(25-26), pages 3587–3596, 2001.

J. S. Perry. http ://svi.cps.utexas.edu/software.shtml, 2002.

J.S. Perry and W. S. Geisler. Gaze-contingent real-time simulation of arbitrary visual fileds. In B. Rogowitz and T. Pappas, editors, *Human Vision and Electronic Imaging, SPIE Proceedings*, 2002.

R. J. Peters and L. Itti. Beyond bottom-up : Incorporating task-dependent influences into a computational model of spatial attention. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

R. J. Peters and L. Itti. Applying computational tools to predict gaze direction in interactive visual environments. In *ACM Transactions on Applied Perception*, volume 5(2), 2008.

R. J. Peters, A. Iyer, L. Itti, and C. Koch. Components of bottom-up gaze allocation in natural images. In *Vision Research*, volume 45(8), pages 2397–2416, 2005.

M. I. Posner and Y. Cohen. Components of visual orienting. In H. Bouma and D. Bouwhuis, editors, *Attention and performance*, volume X, pages 531–556, Hove, UK, 1984. Lawrence Erlbaum Associates Ltd.

M. I. Posner, R. D. Rafal, L. S. Choate, and J. Vaughan. Inhibition of return : Neural basis and function. In *Cognitive Neuropsychology*, volume 2, pages 211–228, 1985.

C. M. Privitera and L. W. Stark. Algorithms for defining visual regions-of-interest : Comparison with eye fixations. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 22, pages 970–982, 2000.

P. Reinagel and A. M. Zador. Natural scene statistics at the centre of gaze. In *Network-Computation in Neural Systems*, volume 10(4), pages 341–350, 1999.

M. Rucci. Fixational eye movements, natural image statistics, and fine spatial vision. In *Network : Computation in Neural Systems*, volume 19(4), pages 253–285, 2008.

A. G. Samuel and D. Kat. Inhibition of return : A graphical meta-analysis of its time course and an empirical test of its temporal and spatial properties. In *Psychonomic Bulletin & Review*, volume 10, pages 897–906, 2003.

G. Schwarz. Estimating the dimension of a model. In *Annals of Statistics*, volume 6(2), pages 461–464, 1978.

R. Shapley, E. Kaplan, and R. Soodak. Spatial summation and contrast sensitivity of x and y cells in the lateral geniculate nucleus of the macaque. In *Nature*, volume 292, pages 543–545, 1981.

D. Stettler, A. Das, J. Bennett, and C. Gilbert. Lateral connectivity and contextual interactions in macaque primary visual cortex. In *Neuron*, volume 36(4), pages 739–750, 2002.

G. Strang and T. Nguyen. *Wavelets and filter banks*. Wellesley-Cambridge Press, 1996.

E. Switkes, M. J. Mayer, and J. A. Sloan. Spatial frequency analysis of the visual environment : anisotropy and the carpentered environment hypothesis. In *Vis. Res.*, volume 18, pages 1393–1399, 1978.

G. Tassinari, S. Aglioti, L. Chelazzi, C. A. Marzi, and G. Berlucchi. Distribution in the visual field of the costs of voluntarily allocated attention and of the inhibitory after-effects of covert orienting. In *Neuropsychologia*, volume 25, pages 55–71, 1987.

B. W. Tatler. The central fixation bias in scene viewing : Selecting an optimal viewing position independently of motor biases and image feature distributions. In *Journal of Vision*, volume 7(14) :4, pages 1–17, 2007.

B. W. Tatler and B. T. Vincent. Systematic tendencies in scene viewing. In *Journal of Eye Movement Research*, volume 2(2), pages 1–18, 2008.

B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist. Visual correlates of fixation selection : effects of scale and time. In *Vision Research*, volume 45(5), pages 643–659, 2005.

B. W. Tatler, R. J. Baddeley, and B. T. Vincent. The long and the short of it : spatial statistics at fixation vary with saccade amplitude and task. In *Vision Research*, volume 46, pages 1857–1862, 2006.

J. Theeuwes, A. F. Kramer, S. Hahn, and D. E. Irwin. Our eyes do not always go where we want them to go : capture of the eyes by new objects. In *Psychological Science*, volume 9 (5), pages 379–385, 1998.

D. J. Tolhurst, Y. Tadmor, and C. Tang. The amplitude spectra of natural images. In *Ophthalmic Physiol. Opt.*, volume 12, pages 229–232, 1992.

R. B. Tootell, M. S. Silverman, and R. L. De Valois. Spatial frequency columns in primary visual cortex. In *Science*, volume 214, pages 813–815, 1981.

A. Torralba. *Architectures analogiques pour le traitement d'images : réseaux cellulaires neuronaux et circuits neuromorphiques*. PhD thesis, INPG, 1999.

A. Torralba. Modeling global scene factors in attention. In *Journal of Optical Society of America A. Special Issue on Bayesian and Statistical Approaches to Vision*, volume 20(7), pages 1407–1418, 2003a.

A. Torralba. Contextual priming for object detection. In *International Journal of Computer Vision*, volume 53(2), pages 169–191, 2003b.

A. Torralba and J. Hérault. Circuits neuromorphiques pour l'estimation du mouvement. In *Seizième Colloque GRETSI*, pages 639–642, 1997.

A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson. Contextual guidance of eye movements and attention in real-world scenes : the role of global features in object search. In *Psychol. Rev.*, volume 113(4), pages 766–786, 2006.

A. M. Treisman and G. Gelade. A feature-integration theory of attention. In *Cognitive Psychology*, volume 12(1), pages 97–136, 1980.

C. B. Trevarthen. Two mechanisms of vision in primates. In *Psychologische Forschung*, volume 31(4), pages 299–337, 1968.

D. Ts'o, C. D. Gilbert, and T. N. Wiessel. Relationships between horizontal interactions and functionnal architecture in cat striate cortex as revealed by cross-correlation analysis. In *Jour. Neurosci.*, volume 6, pages 1160–1170, 1986.

K.A. Turano, D.R. Geruschat, and F.H. Baker. Oculomotor strategies for the direction of gaze tested with a real-world activity. In *Vision Res.*, volume 43, pages 333–346, 2003.

B. M. Velichkovsky, M. Joos, J. R. Helmert, and S. Pannasch. Two visual systems and their eye movements : Evidence from static and dynamic scene perception. In B. G. Bara, L. Barsalou, and M. Bucciarelli, editors, *Proceedings of the XXVII Conference of the Cognitive Science Society*, pages 2283–2288, Mahwah, NJ, 2005.

B. T. Vincent, R. Baddeley, A. Correani, T. Troscianko, and U. Leonards. Do we look at lights ? using mixture modelling to distinguish between low- and high-level factors in natural image viewing. In *Visual Cognition*, 2009.

P. Viviani and R. G. Swensson. Saccadic eye movements to peripherally discriminated visual targets. In *Journal of Experimental Psychology : Human Perception and Performance*, volume 8 (1), pages 113–126, 1982.

N. J. Wade and B. W. Tatler. The moving tablet of the eye : the origins of modern eye movement research. New York, 2005. Oxford University Press.

D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch. Attentional selection for object recognition - a gentle way. In *Lecture Notes in Computer Science*, volume 2525, pages 472–479, 2002.

B. A. Wandell. *Foundations of Vision*. Sinauer Associates, Incorporated, 1995.

J. M. Wolfe. Guided search 2.0 : A revised model of visual search. In *Psyonomic Bulletin and Review*, volume 1(2), pages 202–238, 1994.

Q. Yang, M. P. Bucci, and Z. Kapoula. The latency of saccades, vergence, and combined eye movements in children and in adults. In *Investigative Ophthalmology & Visual Science*, volume 43, pages 2939–2949, 2002.

I. A. Yarbus. *Eye movements and vision*. Plenum Press, New York, 1967.

G.J. Zelinsky. Using eye saccades to assess the selectivity of search movements. In *Vision Research*, volume 36(14), pages 2177–2187, 1996.

B. Zenger, J. Braun, and C. Koch. Attentional effects on contrast detection in the presence of surround masks. In *Vision Research*, volume 40(27), pages 3717–3724, 2000.

L. Zhaoping. A saliency map in primary visual cortex. In *Trends in Cognitive Sciences*, volume 6(1), pages 9–16, 2002.

L. Zhaoping. The primary visual cortex creates a bottom-up saliency map. In L. Itti, G. Rees, and J.K. Tsotsos, editors, *Neurobiology of Attention*, pages 570–575. Elsevier, 2005.