



Implémentation d'un système préattentionnel avec des neurones impulsionnels

Sylvain Chevallier

► To cite this version:

Sylvain Chevallier. Implémentation d'un système préattentionnel avec des neurones impulsionnels. Modélisation et simulation. Université Paris Sud - Paris XI, 2009. Français. NNT : . tel-00472849

HAL Id: tel-00472849

<https://theses.hal.science/tel-00472849>

Submitted on 13 Apr 2010

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

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

Implémentation d'un système préattentionnel avec des neurones impulsionnels

Sylvain Chevallier

LIMSI-CNRS

Université Paris-Sud 11 - Orsay

Soutenue le 25 juin 2009 devant le jury composé de :

<i>Président du jury</i>	Jean-Paul Sansonnet
<i>Rapporteurs</i>	Philippe Gaussier
	Pierre Kornprobst
<i>Examineur</i>	Laurent Perrinet
<i>Directrice</i>	Hélène Paugam-Moisy
<i>Directeur</i>	Philippe Tarroux

Résumé

Les neurones impulsionnels prennent en compte une caractéristique fondamentale des neurones biologiques : la capacité d’encoder l’information sous forme d’événements discrets. Nous nous sommes intéressés à l’apport de ce type de modèles dans le cadre de la vision artificielle, dont les contraintes nous ont orienté vers le choix de modèles simples, adaptés à la rapidité de traitement requise.

Nous décrivons une architecture de réseaux pour encoder et extraire des saillances utilisant la discrétisation induite par les neurones impulsionnels. La carte de saillances est obtenue à partir de la combinaison, spatiale et temporelle, de différentes cartes de modalités visuelles (contrastes, orientations et couleurs) à différentes échelles spatiales. Nous proposons une méthode de filtrage neuronal pour construire les cartes de modalité visuelle. Cette méthode réalise le filtrage de façon graduelle : plus le temps de traitement alloué à l’algorithme est important, plus le résultat est proche de celui obtenu avec un filtrage par convolution.

L’architecture proposée donne en sortie les saillances triées temporellement dans l’ordre de leur importance. Nous avons placé en aval de cette architecture un autre réseau de neurones impulsionnels, s’inspirant des champs neuronaux, qui permet de sélectionner la zone la plus saillante et de maintenir une activité constante sur cette zone. Les résultats expérimentaux montrent que l’architecture proposée est capable d’extraire des saillances dans une séquence d’images, de sélectionner la saillance la plus importante et de maintenir la focalisation sur cette saillance, même dans un contexte bruité ou quand la saillance se déplace.

Abstract

Spiking neurons seem to capture important characteristics of biological neurons with relatively simple models. Their main interest is their ability to code information as discrete events. We have investigated the possible contributions of these models to artificial vision. Due to vision constraints, particularly for processing time, we use simple spiking neural models.

A spiking neuron network architecture is described for encoding and extracting saliencies, thanks to the discretization induced by spiking neurons. The saliency map is a combination, in space and temporal domain, of visual modality maps (contrasts, orientations and colors) for different spatial scales. We propose a neural filtering method for building the modality maps. This method implements a process, similar to convolution filtering, by gradual approximation : the more processing time is allowed to the algorithm, the better is the approximation.

The architecture ranks the saliencies in the order of their interest, with the most important encoded first. Another spiking neuron network, inspired by dynamic neural fields, is attached on top of the architecture, allowing to select the most salient region and to focus on. Experimental results point out that the architecture is able to extract saliencies in a sequence of images, to select the most salient region and to focus on this region, in presence of noise or when the saliency moves.

Table des matières

Introduction	1
1 Du neurone à l'attention visuelle	5
1.1 Les neurones biologiques	6
1.1.1 Propos liminaire sur la complexité du système nerveux central . .	6
1.1.2 La morphologie d'un neurone	7
1.1.3 Le potentiel électrique transmembranaire	7
1.1.4 Le potentiel de repos	9
1.1.5 Le potentiel d'action	9
1.1.6 Les synapses	10
1.2 L'architecture du système visuel	12
1.2.1 L'œil	13
1.2.2 La rétine	13
1.2.3 Le corps genouillé latéral	19
1.2.4 Les aires visuelles	21
1.3 L'attention visuelle	23
1.3.1 Être attentif pour percevoir	25
1.3.2 Mise en évidence expérimentale de la préattention	27
1.3.3 Les théories de la préattention	30
1.4 Synthèse	32

2	Les modèles de neurones impulsionnels	35
2.1	Modèles explicatifs ou modèles <i>per se</i> ?	36
2.1.1	Du réel au modèle	36
2.1.2	Différents niveaux de représentation	38
2.1.3	Des inspirations pluridisciplinaires	40
2.1.4	Des modèles à taux de décharge	40
2.2	Les modèles de neurones impulsionnels	42
2.2.1	Caractéristiques des modèles de neurones impulsionnels	42
2.2.2	Le modèle d'intégrateur à fuite	43
2.2.3	Les variantes des modèles intègre-et-décharge	44
2.2.4	Le modèle SRM	46
2.3	Les modèles de synapses	48
2.3.1	Les courants synaptiques	49
2.3.2	La plasticité synaptique	50
2.4	Synthèse	51
3	Les implémentations de mécanismes d'attention visuelle	53
3.1	Quels intérêts pour la vision artificielle ?	54
3.2	Les modèles attentionnels	55
3.2.1	Les modèles neuropsychologiques	56
3.2.2	Les systèmes de traitement d'images	57
3.2.3	Les approches bio-inspirées	59
3.3	Une architecture de la vision préattentive avec des neurones impulsionnels	64
3.3.1	Intérêts d'une implémentation impulsionnelle	65
3.3.2	Un système <i>anytime</i>	68
3.3.3	Méthode d'implémentation du système	70
4	Filtrage neuronal d'image	73
4.1	Le filtrage neuronal	74
4.1.1	La transduction	74

4.1.2	L'intégration	76
4.2	Les codages impulsionnel et fréquentiel	76
4.2.1	Codage des entrées	77
4.2.2	Caractérisation du codage impulsionnel pour le filtrage	79
4.2.3	Utilisation du codage fréquentiel	83
4.2.4	Une approche <i>anytime</i>	86
4.3	Résultats expérimentaux	87
4.3.1	Méthodes	87
4.3.2	Comparaison des méthodes de filtrage	89
4.3.3	Études de paramètres	101
4.4	Conclusion	107
5	Architecture préattentionnelle	109
5.1	Description de l'architecture	110
5.1.1	Transduction artificielle	110
5.1.2	Décomposition en traits caractéristiques	114
5.2	Extraction des saillances	116
5.2.1	Mécanismes mis en jeu	116
5.2.2	Résultats de l'extraction de saillances	118
5.3	Focalisation préattentionnelle	127
5.3.1	Sélection et suivi d'une région saillante	128
5.3.2	Évaluation de la robustesse de la focalisation	131
5.3.3	Focalisation sur une séquence d'images naturelles	134
5.4	Conclusion	136
6	Conclusions et perspectives	139
6.1	Conclusions	139
6.2	Perspectives	143
	Publications	145

Bibliographie	147
A Implémentation du réseau de neurones	165
A.1 Conception du système	165
A.2 La propagation des potentiels d'action	166
A.3 La mise à jour des cartes neuronales	169
B Modèles synaptiques implémentés dans le système	173
B.1 Conductance synaptique instantanée	173
B.2 Fonction α	174
B.3 Modèle en différence d'exponentielles	174
C Générateur de potentiels d'action pour les cartes d'entrée	177
D Résultats complémentaires pour l'extraction de saillances	181

Table des figures

1.1	Microphotographie d'un neurone	6
1.2	Équilibre électrochimique de la membrane plasmique	8
1.3	Microphotographie d'une synapse	11
1.4	Schéma d'une synapse	12
1.5	Différents types d'œil	14
1.6	Schéma de la rétine	15
1.7	Distribution des cônes et des bâtonnets sur la rétine	15
1.8	Cellule ganglionnaire <i>on-off</i>	17
1.9	Microphotographie du corps genouillé latéral (CGL)	19
1.10	Mise en évidence de la rétinotopie de l'aire visuelle primaire	21
1.11	Hypercolonne de V1	23
1.12	Les aires corticales impliquées dans la vision	24
1.13	Cécité aux changements	26
1.14	Phénomène de <i>pop out</i>	27
1.15	Recherche visuelle sérielle	28
1.16	Traits caractéristiques	29
1.17	<i>Feature integration theory</i>	31
1.18	<i>Guided search theory</i>	33
2.1	Organisation schématique du système nerveux central	39
2.2	Simple circuit électrique modélisant la membrane d'un neurone	43
2.3	Classifications des différents motifs de décharge des neurones corticaux	45
2.4	Illustration du <i>Spike Response Model</i> (SRM)	47

2.5	Illustration du modèle SRM_0	48
2.6	Fenêtre de <i>Spike Timing-Dependent Plasticity</i> (STDP)	51
3.1	Descripteurs d'image <i>Scale Invariant Feature Transform</i> (SIFT)	59
3.2	Architecture d'Itti & Koch	61
3.3	Ondelettes de Gabor et champs récepteurs	62
3.4	Les modèles attentionnels et leurs corrélats neuronaux	66
3.5	Neurones intégrateurs et détecteurs de coïncidence	67
3.6	Caractérisation de la réponse d'un algorithme <i>anytime</i>	69
4.1	Filtrage neuronal d'image	75
4.2	Schéma de l'influence de la luminance sur les fréquences de décharge	77
4.3	Temps de latence et fréquence de décharge des neurones d'entrée	78
4.4	Approximations réalisées par le codage impulsionnel	80
4.5	Décours des potentiels d'action présynaptiques	81
4.6	Filtrage avec un codage fréquentiel	84
4.7	Ajout de bruit au niveau du potentiel membranaire	86
4.8	Filtre différence de gaussiennes (DOG) utilisé	88
4.9	Comparaison des méthodes de filtrage pour l'image "Rectangles 1"	90
4.10	Comparaison des méthodes de filtrage pour l'image "Rectangles 2"	91
4.11	Comparaison des méthodes de filtrage pour l'image "Dégradés"	91
4.12	Comparaison des méthodes de filtrage pour l'image "Stries dégradées"	92
4.13	Comparaison des méthodes de filtrage pour l'image "Stries 1"	92
4.14	Comparaison des méthodes de filtrage pour l'image "Stries 2"	93
4.15	Comparaison des méthodes de filtrage pour l'image "Stries 3"	93
4.16	Comparaison des méthodes de filtrage pour l'image "Filtres 1"	94
4.17	Comparaison des méthodes de filtrage pour l'image "Filtres 2"	95
4.18	Histogramme des erreurs quadratiques moyennes mesurées	98
4.19	Ordre d'apparition des résultats du filtrage	100
4.20	Influence de $\tau_{\text{entrée}}$ sur le filtrage	102

4.21	Résultats du filtrage pour $\tau_{\text{entrée}} \in \{0.02, 0.1\}$	103
4.22	Influence de τ_{SDOG} sur le filtrage	104
4.23	Résultats du filtrage pour $\tau_{\text{SDOG}} \in \{0.01, 1\}$	105
4.24	Influence de w_{max}^D sur le filtrage	106
4.25	Résultats du filtrage pour $w_{\text{max}}^D \in \{0.3, 0.5\}$	106
5.1	Schéma de l'architecture de vision attentionnelle	110
5.2	Séparations des canaux	111
5.3	Neurone soumis à un courant d'entrée variable	112
5.4	Décomposition en traits caractéristiques et extraction de saillance	113
5.5	Filtre de Gabor	115
5.6	Double opposition de couleurs	116
5.7	Comparaison des saillances extraites sur une image pour les luminances	120
5.8	Mise en évidence du caractère <i>anytime</i> de l'extraction des saillances	121
5.9	Mise en évidence de l'ordre d'apparition des saillances selon leur importance	122
5.10	Influence de τ_S sur la détection de saillances	123
5.11	Saillances extraites avec les luminances	124
5.12	Saillances extraites avec les couleurs	125
5.13	Focus d'attention	127
5.14	Modèle proposé pour évaluer la focalisation attentionnelle	128
5.15	Construction des images pour l'évaluation du focus	130
5.16	Suivi d'un stimulus dans un environnement bruité	132
5.17	Comparaison avec les champs neuronaux	133
5.18	Séquence d'images naturelles utilisée pour l'évaluation	135
5.19	Erreur moyenne calculée pour la séquence d'images naturelles	136
A-1	Schéma présentant les liens entre les différents <i>packages</i> du système.	166
A-1	Fonction α et différence d'exponentielles	174
A-1	Intégrateur à fuite soumis à un courant d'entrée constant	178

A-2	Intégrateur à fuite soumis à un courant d'entrée variable	178
A-1	Détails des résultats intermédiaires pour l'image "helis"	182
A-2	Résultats intermédiaires de l'image "sailboats"	183

Liste des tableaux

4.1	Les différents types de codages temporels	77
4.2	Paramètres utilisé de la comparaison des méthodes de filtrages.	90
4.3	Tableau récapitulatif des valeurs d'erreur mesurées	97
5.1	Paramètres des masques de connexion utilisés pour l'extraction de saillances.	119

Remerciements

Je tiens à remercier chaleureusement Pierre Kornprobst et Philippe Gaussier pour avoir accepté d'être les rapporteurs de cette thèse, pour leurs commentaires précis et pour leurs remarques constructives sur ce travail. Je souhaite remercier Laurent Perrinet pour la justesse de ses commentaires et pour toutes les idées judicieuses dont il m'a fait part. Je tiens à remercier Jean-Paul Sansonnet, qui a accepté d'examiner ce travail et de faire partie de mon jury.

Je souhaite remercier très largement mes deux directeurs de thèse, Hélène Paugam-Moisy et Philippe Tarroux. Au travers de longues et constructives discussions, au cours desquelles j'ai sans doute éprouvé leur patience, ils m'ont aidé à formuler et à élaborer une approche pour tenter de répondre aux questions sur lesquelles je me suis penché durant cette thèse. Chacun à leur façon, ils m'ont montré comment exercer ce métier si particulier et que j'apprécie tant. Merci beaucoup.

La conclusion de ces remerciements va logiquement à tous mes amis, collègues du laboratoire et compagnons de projets et de conférences. Leur compagnie a été appréciable et très appréciée.

Introduction

Comprendre le codage neuronal

S'il est possible de considérer le cerveau comme un calculateur, comprendre comment les neurones encodent l'information et comment les assemblées de neurones traitent cette information peut permettre de développer de nouvelles approches en traitement de l'information. Cette rencontre entre neurosciences et informatique constitue un courant appelé neurosciences computationnelles [[Rumelhart et McClelland, 1986](#); [Churchland, 1986](#); [Schwartz, 1990](#); [Sejnowski *et al.*, 1988](#); [Churchland et Sejnowski, 1994](#); [Feng, 2004](#)]. La comparaison entre cerveau et ordinateur n'est pas une idée nouvelle ; elle est présente dès les balbutiements de l'informatique. Pourtant, les recherches en neurobiologie avancent vite et la puissance de calcul disponible sur un simple poste informatique croît exponentiellement. Il devient donc de plus en plus aisé de simuler, avec plus ou moins de réalisme, l'évolution de l'activité dans des réseaux de neurones de plus en plus grands et de plus en plus complexes.

Ces avancées rapides, tant du point de vue du réalisme des modèles que du nombre de neurones que l'on peut simuler, permettent d'étudier le codage et le traitement de l'information sous un jour nouveau. En particulier, les modèles de neurones impulsifs constituent une nouvelle vague connexionniste [[Maass, 1997](#)], comme nous le verrons dans le chapitre 2, et peuvent nous amener à envisager de nouvelles solutions. En effet, ce type de modèle prend en compte le décours temporel du potentiel membranaire d'un neurone et modélise explicitement l'émission de ses potentiels d'action. Les conséquences sont importantes car les neurones discrétisent ainsi l'information : les flux d'information entrant dans le réseau sont ainsi transformés en séquence temporelle d'événements discrets. L'étude de ce codage temporel et des possibilités qu'il offre constitue, dans le contexte de la vision artificielle, l'objet de cette thèse.

Ce codage sous forme de trains de potentiels d'action, qui est largement utilisé chez les animaux invertébrés et vertébrés, doit présenter un avantage pour les organismes biologiques. Bien que les contraintes des neurones biologiques soient bien

différentes de celles des neurones simulés, le codage impulsionnel offre lui aussi des intérêts. Il est possible de considérer l'émission d'un potentiel d'action comme la décision d'un neurone, ce neurone s'activant seulement quand il a reçu suffisamment d'informations pour déclencher lui-même l'envoi d'une nouvelle information. On peut donc se représenter la propagation des flux d'information dans le réseau comme un enchaînement permanent de décisions, où seuls les neurones qui reçoivent des informations pertinentes les propagent à leurs voisins. Les biologistes ont observé un effet intéressant : il semble que dans bien des cas l'activité soit très loin d'être intense et repose seulement sur l'activation simultanée d'un petit nombre de neurones. Ce phénomène, connu sous le nom de codage éparé ou *sparse coding*, a été observé dans le système visuel [Vinje et Gallant, 2000; Frégnac *et al.*, 2005; Olshausen et Field, 2005], dans le système olfactif [Theunissen, 2003] et de façon générale dans toute l'activité corticale. Le codage éparé offre des perspectives intéressantes pour les réseaux de neurones artificiels [Perrinet, 2004]. D'autre part, la représentation des informations sous forme de trains de potentiels d'action permet de développer des méthodes efficaces pour les calculs distribués [Mouraud *et al.*, 2006; Grassmann et Anlauf, 1998; Brette *et al.*, 2007] : il est en effet possible de simuler des réseaux de très grande taille puisque l'information échangée entre les nœuds peut être représentée sous forme d'événements discrets.

Contexte de travail

Nous nous sommes intéressés aux apports des neurones impulsionnels dans le domaine de la vision artificielle. Ce contexte impose des contraintes fortes sur les possibilités envisageables, les traitements effectués devant rester peu coûteux en temps de calcul pour permettre, à échéance, d'extraire les régions de la scène visuelle les plus susceptibles de contenir des informations intéressantes. Ces choix ont eu une forte incidence sur le positionnement que nous avons adopté pour ce travail ; c'est-à-dire que, tout en s'inspirant des observations et des mécanismes décrits dans les travaux de neurosciences et de psychologie, le travail que nous présentons ne cherche ni à reproduire des observations biologiques, ni à valider des modèles théoriques. Il s'agit essentiellement de proposer un compromis entre des modèles biologiquement plausibles et des algorithmes de vision artificielle. Ce manuscrit s'appuie donc sur des travaux pluridisciplinaires pour proposer des méthodes efficaces de traitement de l'information visuelle en informatique.

Cette approche bio-inspirée s'inscrit dans l'axe de recherche en robotique autonome développées au LIMSI, dans le groupe Perception Située. Le système de vision artifi-

cielle utilisant des neurones impulsionnels doit être intégré, à terme, sur une plateforme robotique.

Problématique

Les mécanismes auxquels nous nous sommes intéressés sont ceux qui se situent au tout début de la chaîne des traitements visuels. En particulier, nous abordons ici l'étude des mécanismes *bottom-up*, c'est-à-dire dirigés uniquement par les informations entrantes, et qui ne prennent aucunement en compte les connaissances préalables (mécanismes dits *top-down*). Ces mécanismes sélectionnent les informations importantes et sont inspirés par les processus préattentionnels décrits dans certains travaux de psychologie cognitive [Treisman et Gelade, 1980; Pylyshyn, 1994; Wolfe, 2000; Heinke et Humphreys, 2005]. Outre l'intérêt de réduire l'information transmise, et donc d'améliorer la rapidité des traitements, les mécanismes préattentionnels permettent de fusionner des informations qualitativement différentes. En effet, si l'on considère l'hypothèse selon laquelle ces traitements permettent d'accéder à différentes modalités visuelles – comme par exemple la luminance, les oppositions de couleurs ou les contours orientés – il faut ensuite combiner ces informations multimodales pour trouver les régions les plus intéressantes, celles qui attirent l'attention visuelle. Ces régions reçoivent le nom de saillances.

Principales contributions

Pour étudier les apports des neurones impulsionnels à certains traitements de vision artificielle, nous proposons une architecture neuronale, c'est-à-dire un ensemble de réseaux de neurones connectés entre eux, qui permet d'extraire différentes modalités visuelles, de les combiner en une carte de saillances, de sélectionner la saillance la plus forte et de la suivre à l'aide d'un mécanisme de focalisation. Cette architecture a été complètement implémentée et les résultats expérimentaux sont détaillés dans ce manuscrit. Pour extraire les informations des différentes modalités visuelles, nous proposons d'utiliser des réseaux de neurones pour réaliser des opérations de filtrage. Ce filtrage "neuronal" est réalisé différemment du filtrage par convolution car il calcule des approximations successives, ce qui permet d'obtenir les résultats de façon graduelle et de plus en plus complète.

Plan de la thèse

Dans un premier temps, nous verrons comment fonctionnent les neurones biologiques, le rôle des différentes structures impliquées dans la perception visuelle et les différentes théories explicatives de l'attention visuelle (chapitre 1). Nous verrons ensuite comment les modèles neuronaux, initialement bio-inspirés, s'étaient progressivement éloignés de la biologie, puis les différentes raisons qui justifient actuellement un retour vers des modèles biologiquement réalistes (chapitre 2). Nous définirons le positionnement bio-inspiré de ce travail en revenant sur différentes implémentations de modèles de l'attention visuelle, en biologie, en psychologie et en informatique (chapitre 3) et nous expliquerons les avantages qu'apportent les modèles de neurones impulsionnels pour la modélisation de l'attention visuelle.

Les chapitres suivants décrivent le fonctionnement d'une architecture bio-inspirée de la vision préattentionnelle, implémentée avec des réseaux de neurones impulsionnels.

Nous verrons tout d'abord comment filtrer des images avec des réseaux de neurones impulsionnels (chapitre 4), puis nous expliquerons dans le chapitre 5 comment combiner ces réseaux de neurones pour former une architecture de vision préattentionnelle.

Chapitre 1

Du neurone à l'attention visuelle

Sommaire

1.1 Les neurones biologiques	6
1.1.1 Propos liminaire sur la complexité du système nerveux central	6
1.1.2 La morphologie d'un neurone	7
1.1.3 Le potentiel électrique transmembranaire	7
1.1.4 Le potentiel de repos	9
1.1.5 Le potentiel d'action	9
1.1.6 Les synapses	10
1.2 L'architecture du système visuel	12
1.2.1 L'œil	13
1.2.2 La rétine	13
1.2.3 Le corps genouillé latéral	19
1.2.4 Les aires visuelles	21
1.3 L'attention visuelle	23
1.3.1 Être attentif pour percevoir	25
1.3.2 Mise en évidence expérimentale de la préattention	27
1.3.3 Les théories de la préattention	30
1.4 Synthèse	32

Dans ce premier chapitre nous allons voir ce qui caractérise biologiquement les cellules nerveuses, ou neurones, et ce qui les distingue du reste des autres cellules de l'organisme. Ces explications, nécessaires à la compréhension des mécanismes qui gouvernent ces cellules, vont nous permettre de mieux comprendre les différents aspects des modèles de neurones présentés dans le chapitre 2. Nous aborderons ensuite la description de différentes structures neuronales mises en jeu dans la perception visuelle,

ainsi que les différentes théories de psychologie cognitive concernant l'attention visuelle. Ceci nous permettra de mieux cerner les modèles d'attention visuelle présentés dans le chapitre 3.

1.1 Les neurones biologiques

1.1.1 Propos liminaire sur la complexité du système nerveux central

En propos liminaire, il est important de préciser que le Système Nerveux Central (SNC) abrite un grand nombre de cellules très différentes, toutes plongées dans un bouillonnement de substances chimiques et de corps circulants. Les cellules nerveuses, ou neurones (comme illustré sur la figure 1.1), ne représentent qu'une petite partie du SNC et au sein même de cette catégorie de cellule, il est possible de différencier un large éventail de types différents de neurones [Arbib, 1998]. Même si tous les types de neurones pourraient être précisément décrits, la richesse de cet environnement comporte encore de nombreuses zones d'ombre. Par exemple les cellules gliales, qui sont jusqu'à dix fois plus nombreuses que les neurones, constituent la *glie*, ou « colle », dont on a longtemps cru que le seul rôle était de maintenir la cohésion physique du SNC. Récemment, des travaux ont montré qu'elles avaient bien plus d'importance que cela et que différents types de cellules gliales intervenaient aussi bien dans le développement du SNC [Pfrieger, 2002] que dans une partie des transmissions synaptiques [Volterra et Steinhäuser, 2004]. Ainsi, il paraît important de préciser que même si l'on disposait d'une modélisation capable de capturer tous les aspects fonctionnels et les différents comportements d'un neurone, cela ne serait sans doute pas suffisant pour reproduire l'ensemble des mécanismes mis en jeu dans le cerveau.

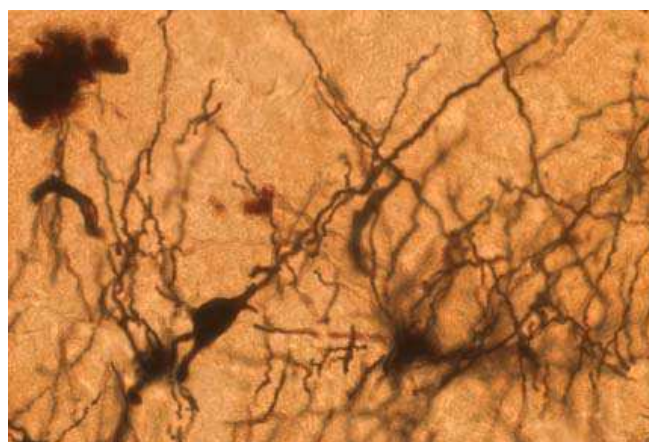


FIGURE 1.1 Microphotographie de neurones après une coloration de Golgi. [Source : www.ihcworld.com]

Toutefois, les neurones semblent réaliser une partie prépondérante des traitements qui prennent place dans le SNC. La richesse des comportements et des traitements réalisés par les neurones, qu'ils soient observés au niveau unitaire ou à celui d'une population, constitue un enjeu à la fois pour les modélisateurs et pour les théoriciens. L'étude des modèles neuronaux suppose certaines connaissances sur la physiologie et le fonctionnement des neurones que nous allons détailler ci-après.

1.1.2 La morphologie d'un neurone

Les neurones sont des cellules et possèdent donc une structure commune à celle de la plupart des cellules. Le corps cellulaire, ou soma, contient le noyau et les organites que l'on retrouve dans toute cellule (mitochondrie, appareil de Golgi, réticulum endoplasmique, ribosome, etc). La différence principale entre les neurones et les autres cellules tient à leur spécialisation qui consiste à intégrer et propager des signaux électriques. Ainsi, la différence la plus notable concerne la morphologie des neurones (figure 1.1). Ils sont pourvus de ramifications plus ou moins importantes qui peuvent, dans les cas les plus extrêmes, dépasser un mètre de long. Ces ramifications sont constituées par les dendrites, qui reçoivent les courants entrants, et par l'axone, unique, qui propage la dépolarisation sortante¹.

1.1.3 Le potentiel électrique transmembranaire

Dendrites et axones convoient l'information sous forme de variations de potentiel et, comme la communication entre neurones est au cœur des modèles présentés ci-après, il est important d'expliquer en détail les mécanismes mis en jeu. Ces mécanismes permettent de propager des dépolarisations sur de longues distances, sans diminution d'intensité.

Pour dépasser cette limite physiologique, la propagation du potentiel utilise un système d'amplification qui s'explique par la perméabilité sélective des neurones à certains ions et par la distribution de ces ions entre l'intérieur et l'extérieur de la membrane. Le milieu intracellulaire est séparé du milieu extracellulaire par la membrane plasmique qui est semi-perméable aux ions, à cause des canaux qui la traversent. Ces canaux peuvent transporter passivement ou activement certains ions, c'est-à-dire qu'un type de canal donné est spécialisé dans le transport d'un type d'ion physiologique donné (Na^+ , K^+ , Ca^{2+} ou Cl^-). La concentration des ions est différente de part et d'autre de

1. On parle aussi du sens afférent pour désigner les connexions entrantes et du sens efférent pour les connexions sortantes.

la membrane et comme cette membrane n'est perméable qu'à certains ions, le système atteint un état d'équilibre électrochimique. Le gradient chimique qui pousse les ions à traverser la membrane est compensé par le gradient électrique qui les empêche de la traverser, comme le montre la figure 1.2 [Purves *et al.*, 1997]. L'état d'équilibre électrochimique atteint en l'absence de stimulation extérieure est appelé potentiel de repos du neurone. Une remarque importante concernant cet équilibre : il suffit que très peu d'ions traversent la membrane pour modifier de façon importante le potentiel de membrane.

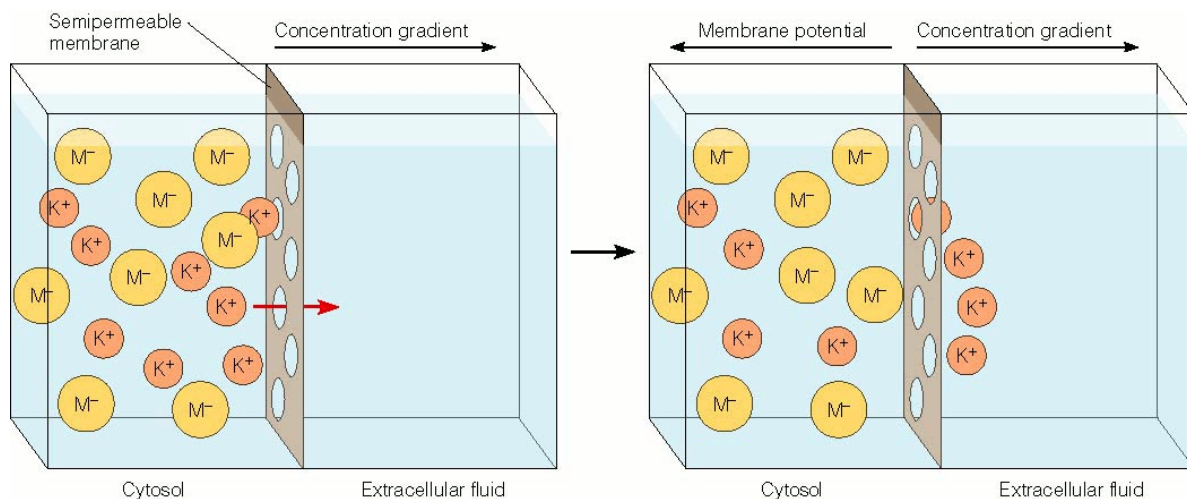


FIGURE 1.2 À gauche : aux conditions initiales, une membrane perméable uniquement aux ions potassium K^+ sépare le cytosol, ou milieu intracellulaire, et le milieu extracellulaire (*Extracellular fluid*). Les ions M^- représentent ici un anion, comme par exemple le chlorure Cl^- . La différence de concentration pousse les ions à s'équilibrer de part et d'autre de la membrane. À droite : aux conditions d'équilibre, une petite quantité d'ions K^+ a traversé la membrane jusqu'au moment où la force électrostatique agissant sur K^+ équilibre le gradient de concentration et annule le mouvement d'ions. [Source : *Principles of Cell Biology*, copyright Pearson Education, Inc 2005]

Les canaux ioniques qui percent la membrane plasmique du neurone sont des protéines membranaires intrinsèques, qui peuvent être plus ou moins complexes. On peut cependant distinguer les canaux passifs des pompes à ions, ces dernières permettant de rééquilibrer les concentrations ioniques de part et d'autre de la membrane [Catterall, 1988]. Les canaux passifs peuvent être séparés en trois grandes catégories, en fonction de ce qui cause l'ouverture de ces canaux : les canaux ouverts par des ligands intracellulaires, ceux ouverts par des ligands extracellulaires et ceux commandés par le voltage. On parle de canaux ouverts par des ligands intracellulaires car ce sont des protéines dont le site fonctionnel est situé sur la face interne de la membrane. Les ligands extracellulaires font principalement référence aux neuromédiateurs que l'on détaillera dans la partie 1.1.6. Les canaux activés par le voltage (dits *voltage-gated*) s'ouvrent en fonction

du potentiel membranaire ou, plus exactement, la probabilité que ces canaux laissent passer des ions dépend du voltage.

1.1.4 Le potentiel de repos

Ce sont les canaux activés par le voltage qui expliquent l'origine du potentiel d'équilibre, en particulier ceux qui gèrent les flux de potassium. En effet, la membrane est plus perméable aux ions potassium qu'aux autres ions autour du potentiel de repos et il y a plus de potassium à l'intérieur qu'à l'extérieur de la cellule, grâce aux pompes à ions [Hodgkin *et al.*, 1952]. Ce potentiel peut s'exprimer par l'équation de Goldman ([Goldman, 1953] cité dans [Purves *et al.*, 1997])² :

$$V = \frac{RT}{F} \log \frac{P_K[K]_{ext} + P_{Na}[Na]_{ext} + P_{Cl}[Cl]_{int}}{P_K[K]_{int} + P_{Na}[Na]_{int} + P_{Cl}[Cl]_{ext}} \quad (1.1)$$

où V est le voltage transmembranaire, R est la constante des gaz parfaits, T la température, F la constante de Faraday, P_X désigne la perméabilité de la membrane à l'ion X et $[X]$ indique la concentration de l'ion X à l'intérieur ou à l'extérieur de la cellule. Comme Cl^- est de valence opposée aux autres ions, les concentrations intérieure et extérieure sont inversées.

1.1.5 Le potentiel d'action

Les changements de perméabilité des canaux activés par le voltage ont été décrits par Hodgkin et Huxley [Hodgkin et Huxley, 1952b; Hodgkin et Huxley, 1952a; Hodgkin et Huxley, 1952c; Hodgkin et Huxley, 1952d]. Les auteurs sont partis de l'hypothèse que les courants ioniques sont dûs à des changements de la conductance membranaire, celle-ci étant l'inverse de la résistance membranaire. Cette conductance capture bien les effets de la perméabilité, bien qu'elle ne soit pas strictement identique à cette quantité. En utilisant la loi d'Ohm, il est possible de décrire la relation entre un le courant d'un ion X et la conductance membranaire avec

$$I_X = g_X(V_m - E_X) \quad (1.2)$$

où I_X est le courant ionique, g_X la conductance membranaire, V_m le potentiel de membrane et E_X le potentiel d'équilibre de l'ion X .

2. Cette équation ne prend en compte que les ions perméants les plus importants, mais il est possible de l'étendre pour intégrer l'influence des ions calcium Ca^{2+} .

Ce potentiel d'équilibre est donné par l'équation de Nernst :

$$E_X = \frac{RT}{zF} \ln \frac{[X]_{ext}}{[X]_{int}} \quad (1.3)$$

avec z la valence de l'ion considéré.

À partir de mesures expérimentales, les auteurs ont pu déterminer les conductances pour les ions sodium g_{Na} et potassium g_K [Hodgkin et Huxley, 1952b]. Ces conductances mettent un certain temps à s'activer, et en particulier l'activation de la conductance potassique est plus lente que l'activation de la conductance sodique.

Si un neurone est soumis à un courant d'entrée suffisamment fort, il déclenche un potentiel d'action ou *spike*. Le potentiel d'action est un mécanisme en tout ou rien, c'est-à-dire que tant que le courant d'entrée ne modifie pas suffisamment le potentiel membranaire, celui-ci revient vers le potentiel d'équilibre. Si le courant dépasse un certain seuil, la conductance sodique s'active et accroît l'entrée de sodium dans la cellule, ce qui dépolarise encore plus la membrane et ouvre donc encore plus les canaux sodiques. Comme l'activation des conductances sodiques est plus rapide que celle des conductances potassiques, la conductance sodique atteint son maximum plus rapidement. E_{Na} étant positif et supérieur à E_K , le potentiel membranaire croît rapidement vers E_{Na} . La conductance potassique, plus lente, fait alors redescendre le potentiel vers E_K , qui est généralement inférieur au potentiel de repos. Cette brève période pendant laquelle le neurone est hyperpolarisé s'accompagne d'une inactivation des canaux sodiques. Cette période est appelée période réfractaire et l'on distingue généralement la période réfractaire absolue, pendant laquelle le neurone ne répond plus aux excitations, de la période réfractaire relative, où le neurone est difficilement excitable car les canaux sodiques sont inactivés et les canaux potassiques ouverts poussent le potentiel vers E_K .

1.1.6 Les synapses

Les variations de potentiel proviennent principalement de l'activité de cellules qui sont connectées par une liaison physique. Cette liaison peut être directe ou passer brièvement par le milieu extracellulaire. Dans le premier cas, on parle de jonction communicante ou *gap junction*, c'est-à-dire que deux cellules partagent une sorte de pont physique qui traverse les membranes et permet un échange non contraint du milieu intracellulaire entre les deux neurones. Ce type de connexion n'est pas le plus fréquent, bien qu'on en trouve dans la rétine ; la plupart des neurones sont connectés entre eux par l'intermédiaire de synapses chimiques.

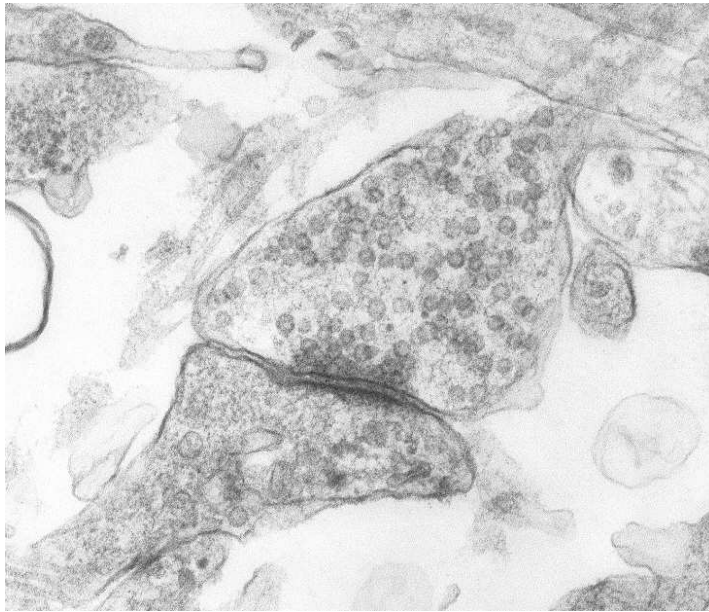


FIGURE 1.3 Microphotographie électronique d'une synapse issue d'une culture de neurones du rat. Les vésicules de neurotransmetteurs, petits cercles sombres dans la partie haute de l'image, sont clairement visibles. [Source : [laboratoire de neurobiologie d'Orléans](#)]

Les synapses chimiques (figure 1.3) permettent de propager unilatéralement les stimulations électriques entre deux neurones qui ne sont pas connectés physiquement. La terminaison présynaptique est remplie de vésicules contenant des neurotransmetteurs : ce sont des signaux chimiques qui peuvent prendre la forme de différentes substances chimiques. Ces neurotransmetteurs sont relâchés dans la fente synaptique qui sépare physiquement les neurones lorsqu'un potentiel d'action arrive dans la partie présynaptique. Ces neurotransmetteurs vont se lier à des récepteurs de la partie postsynaptique et entraîner l'ouverture ou la fermeture de canaux ioniques. Il en résulte une modification locale du potentiel de membrane (le potentiel postsynaptique ou PPS) qui va se propager jusqu'au corps cellulaire. Les caractéristiques de cette variation du potentiel dépendent du type de neurotransmetteurs que la synapse relâche.

Les synapses chimiques mettent en jeu de nombreux mécanismes (figure 1.4), souvent concurrents, et constituent un objet d'étude d'une grande complexité. Parmi tous ces mécanismes, nous aborderons dans le chapitre suivant ceux relatifs à la modification de l'efficacité synaptique : on parle de potentiation à long terme (ou LTP pour *long-term potentiation*) et de dépression à long terme (LTD pour *long-term depression*). Ces modifications à long terme peuvent être provoquées par la co-stimulation des neurones pré- et postsynaptiques. Si l'on stimule de façon répétée le neurone présynaptique, puis 10 ms plus tard le neurone postsynaptique, il en résulte que le PPS généré est plus important [Markram *et al.*, 1995; Markram *et al.*, 1997]. À l'inverse, il est possible

d'induire une LTD en stimulant le neurone postsynaptique en premier. Ces modifications de l'efficacité synaptique ont été largement étudiées (voir par exemple [Bi et Poo, 1998; Bi et Poo, 2001]).

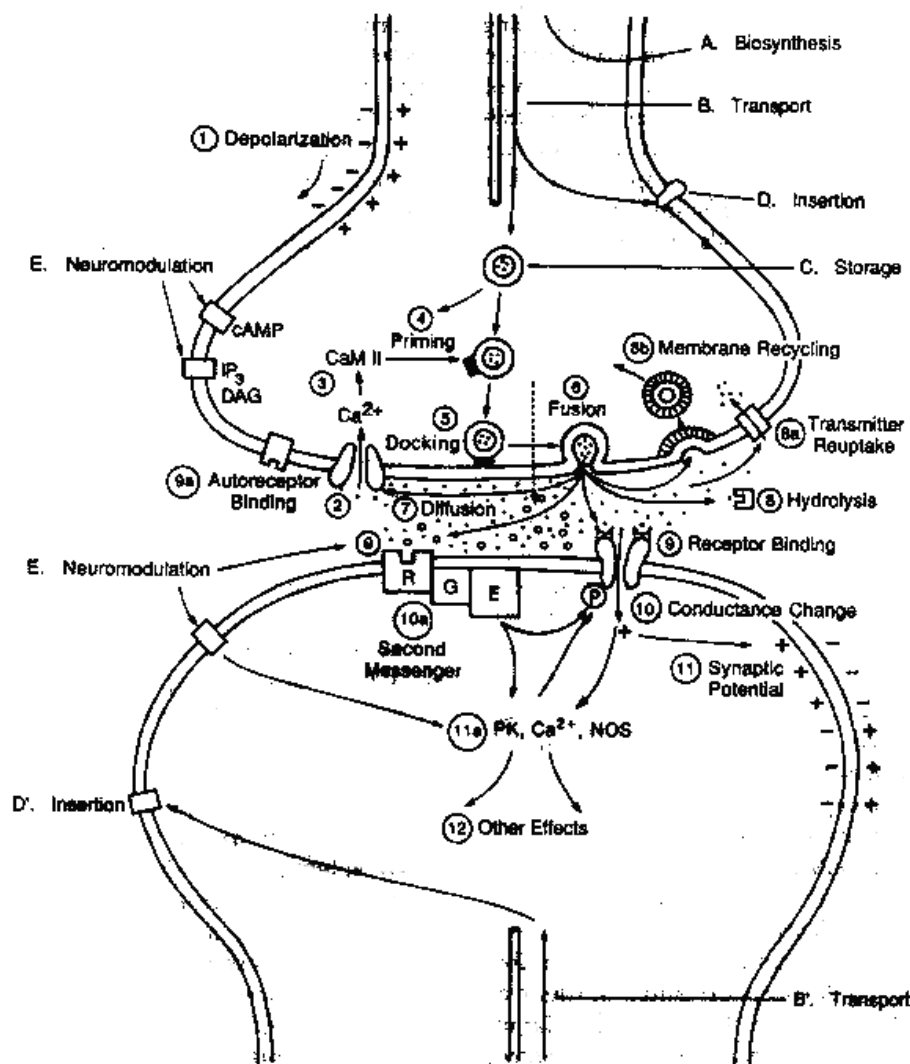


FIGURE 1.4 Schéma d'une vue en coupe de synapse, qui illustre les différents mécanismes prenant part à la propagation électrochimique de l'information. La terminaison présynaptique est en haut et le neurone postsynaptique est en bas. [Source : [Shepherd, 1998]]

1.2 L'architecture du système visuel

Maintenant que nous avons vu le fonctionnement unitaire des cellules nerveuses, nous allons aborder le fonctionnement des neurones dédiés à la perception visuelle. Chez les primates, ce système perceptif est très développé : le tiers de la surface corticale du cerveau peut être considéré comme traitant des informations visuelles [Chalupa et Werner, 2003]. Dans cette partie, après un bref commentaire sur l'œil, nous verrons les princi-

pales structures neuronales mises en jeu et leur rôle respectif : la rétine et ses différentes couches cellulaires (1.2.2), le corps genouillé latéral (1.2.3), l'aire visuelle primaire et les aires corticales impliquées dans le traitement des informations visuelles (1.2.4). Ces explications sont loin d'être exhaustives et seules les informations pertinentes pour la compréhension des modélisations de l'attention visuelle présentées dans le chapitre 3 sont détaillées ici. Cette partie est très largement influencée par les explications que l'on peut trouver dans les ouvrages de référence « Neurosciences » [Purves *et al.*, 1997], « The Visual Neurosciences » [Chalupa et Werner, 2003], « Principles of Neural Science » [Kandel *et al.*, 2000] et le site internet « Webvision » [Kolb *et al.*, 2008].

1.2.1 L'œil

Des différents types d'yeux qu'il est possible d'observer dans le règne animal, les yeux des primates sont parmi les plus complexes et offrent une très bonne acuité visuelle comme le montre la figure 1.5. Pourtant, la structure de l'œil et la qualité des images formées au fond de l'œil est loin d'être sans défaut. Du point de vue d'un ingénieur, la structure de l'œil des primates est même contre-intuitive.

La structure de l'œil humain a de quoi surprendre. La cornée, le cristallin et la pupille permettent de former une image plus ou moins nette sur le fond de l'œil. Mais les rayons lumineux doivent traverser les couches cellulaires de la rétine, sans parler des vaisseaux sanguins, avant d'atteindre les photorécepteurs, qui sont tournés vers le fond de la rétine³ (voir figure 1.6). Du point de vue d'un ingénieur, c'est le contraire de ce qu'il faudrait faire [Purves *et al.*, 1997] : il vaut mieux en effet mettre tous les câbles derrière les capteurs et non pas devant. Tous les axones, les vaisseaux sanguins passent presque en plein milieu de la rétine et il existe donc une zone au fond de chaque œil, appelée tache aveugle, sur laquelle il ne peut pas y avoir de capteurs. L'absence de capteurs sur cette zone est apparente sur la figure 1.7.

1.2.2 La rétine

La rétine désigne l'ensemble des neurones qui sont situés au fond de l'œil et qui réalisent la transduction – le passage du signal lumineux à un signal électrique – et les premiers traitements visuels. La rétine est plus qu'un simple capteur car elle traite déjà l'information visuelle. Il existe une grande diversité de cellules dans la rétine, qui sont structurées en couches réalisant différentes transformations spatio-temporelles de

3. Sauf au centre de la rétine, où il n'y a presque pas de vaisseaux sanguins et où les photorécepteurs sont "couchés" pour laisser passer plus de lumière.

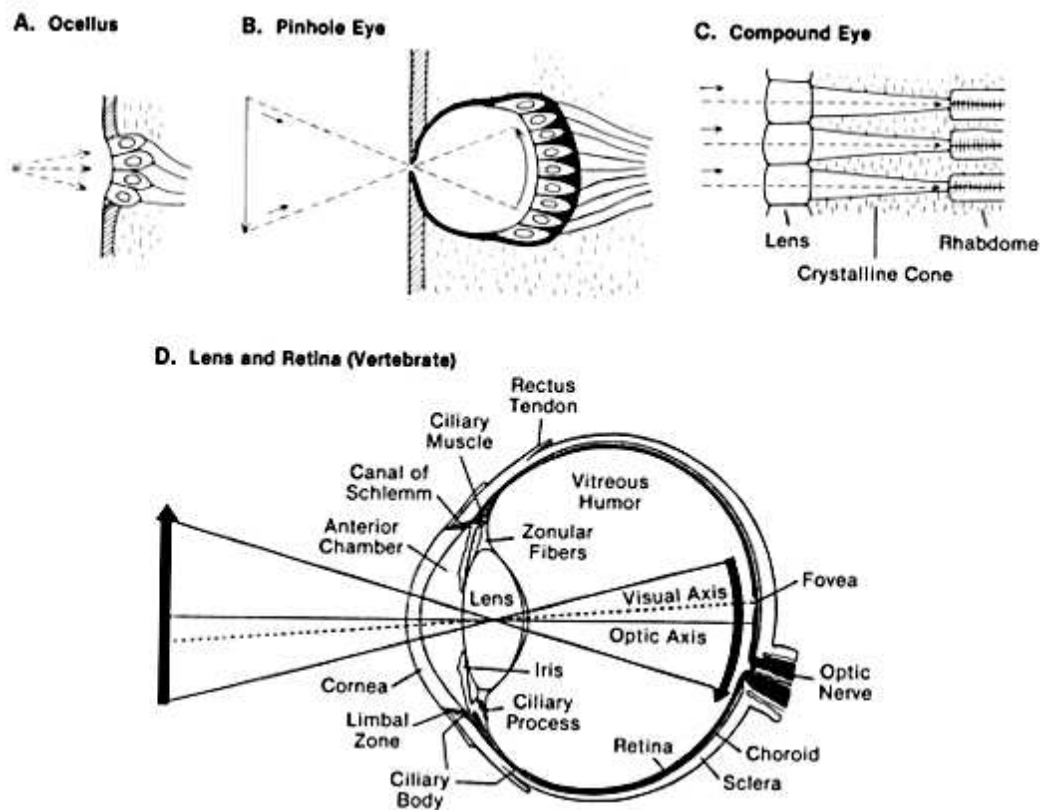


FIGURE 1.5 Illustration résumant les différents types d'yeux existant dans le monde animal. (A) Ocelle (B) Simple pupille, ou « œilleton » (C) Œil composé ou à facettes (D) Cristallin et rétine. [Source : cours de [Paul Avery](#), professeur au département de physique de l'université de Floride]

l'information. Ces différentes étapes de traitement permettent à la rétine d'envoyer vers le cerveau les informations visuelles sous une forme condensée qui implique un processus de sélection des informations pertinentes.

Les fonctions de transduction sont assurées par les photorécepteurs, cellules nerveuses qui contiennent un pigment photosensible. La lumière absorbée par ce pigment produit une cascade d'événements qui conduit à la libération de neurotransmetteurs à destination des cellules postsynaptiques. Le type de pigment et la morphologie de ces photorécepteurs permet de distinguer deux types de cellules : les bâtonnets et les cônes.

D'un point de vue fonctionnel, les bâtonnets sont utilisés pour la vision périphérique et leur convergence vers les cellules des couches supérieures de la rétine est importante. Cette convergence permet d'observer une réponse des cellules des couches supérieures même lorsque l'éclairage est très faible, mais cette convergence diminue la résolution spatiale du réseau des bâtonnets. Les bâtonnets sont généralement 20 fois plus nombreux que les cônes chez les mammifères [[Masland, 2001a](#)].

Les cônes se répartissent en trois catégories selon le pigment qu'ils contiennent et

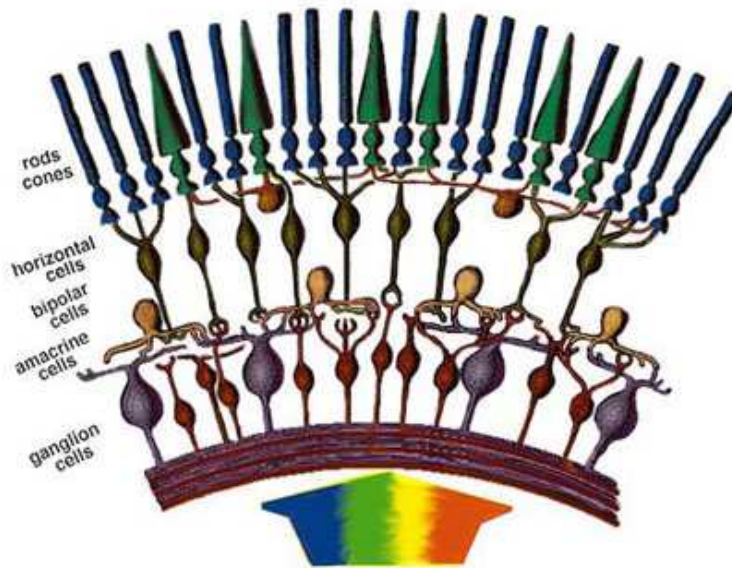


FIGURE 1.6 Schéma en coupe des différentes couches cellulaires qui composent la rétine. La lumière vient d'en bas et traverse d'abord toutes les couches cellulaires avant de provoquer une dépolarisation des cellules photoréceptrices, les cônes (*cones*) et les bâtonnets (*rods*). Les variations de potentiels sont propagées dans les cellules bipolaires (*bipolar cells*) vers les cellules ganglionnaires (*ganglions cells*), dont les axones s'assemblent pour former le nerf optique. Les cellules horizontales (*horizontal cells*) et amacrines (*amacrine cells*) jouent un rôle important dans les interactions latérales. [Source : cours de [Paul Avery](#), professeur au département de physique de l'université de Floride]

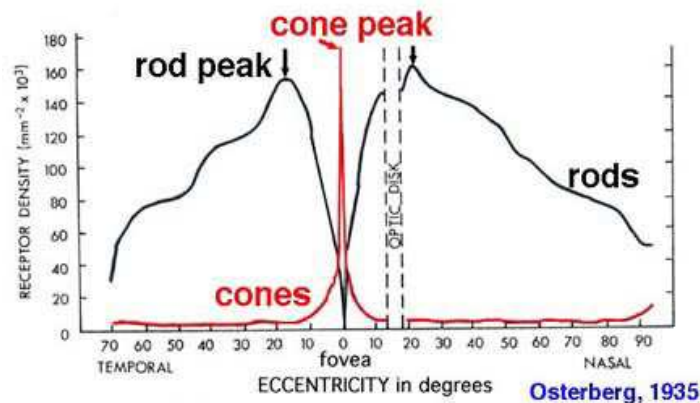


FIGURE 1.7 Nombre de cônes (*cone*) et de bâtonnets (*rod*) en milliers par mm^2 en fonction de l'excentricité par rapport au centre de la rétine. La zone centrale est la fovéa, dans laquelle les cônes sont concentrés. La tache aveugle (*Optic disk*) indique une zone qui ne possède pas de récepteurs car les axones et les vaisseaux sanguins traversent la rétine à cet endroit. [Source : [Osterberg, 1935](#)]

sont sensibles à une longueur d'onde particulière. Le réseau formé par les cônes permet donc de discriminer les couleurs. Sa convergence est faible, c'est-à-dire qu'un cône peut n'être connecté qu'à une seule cellule de la couche supérieure. Ceci permet d'obtenir une bonne résolution spatiale mais cette vision des couleurs ne fonctionne qu'à des niveaux d'éclairement importants. Du point de vue de l'évolution, il semble probable que les cônes soient apparus d'abord et que les bâtonnets se soient ensuite développés en utilisant le réseau des cônes [Masland, 2001a].

Une autre différence remarquable entre ces deux types de photorécepteurs concerne leur répartition sur la rétine. La figure 1.7 fait apparaître deux zones distinctes : une zone centrale, la fovéa, qui concentre la quasi-totalité des cônes, et la région périphérique, dans laquelle se répartissent les bâtonnets. C'est dans la fovéa que l'acuité visuelle est la plus grande et la densité de récepteurs chute rapidement pour les parties excentrées : à seulement 6° d'écart de l'axe du regard, l'acuité a diminué de 75% [Purves *et al.*, 1997].

Les photorécepteurs réalisent la transduction et constituent donc le point d'entrée de l'information de la structure rétinienne (figure 1.6). D'un point de vue fonctionnel, il est possible de segmenter la rétine en deux couches, les couches plexiformes externes et internes. La couche plexiforme externe regroupe les photorécepteurs et les cellules horizontales qui assurent les connexions latérales. La couche plexiforme interne désigne les cellules ganglionnaires, qui sont la porte de sortie de l'information dans le réseau rétinien et les cellules amacrines qui modulent les informations transmises avec leurs connexions latérales. Les cellules bipolaires assurent la transmission de l'information entre ces deux couches, avec d'un part leurs dendrites dans la couche interne et d'autre part leur axone dans la couche externe. Ces cellules bipolaires présentent une activité phasique qui peut être de type *on* ou *off*, selon qu'elles s'activent de façon transitoire lorsqu'une lumière apparaît ou disparaît (pour une description plus détaillée voir le chapitre I.3.2 de [Wohrer, 2008]).

La majorité de l'information transmise dans les couches de la rétine est portée par de simples dépolarisations ou potentiels gradués et non par des potentiels d'action. Fonctionnellement, la rétine est un réseau en avant (ou *feedforward*) relativement compact dans lequel les interactions sont en majorité locales, les cellules qui la composent influent seulement sur les cellules voisines. Le besoin d'un moyen de transmettre l'information sur de longues distances – le potentiel d'action – est donc moins pressant.

D'autre part, bien qu'une grande partie de la communication intercellulaire passe par des synapses, il est possible de trouver des jonctions communicantes dans toutes les couches rétinienne. En particulier, les cellules horizontales utilisent largement ce type de connexion entre elles et propagent donc ainsi localement les activations des

photorécepteurs [Masland, 2001a]. Ce mécanisme permet de réaliser une moyenne locale de l'activité des photorécepteurs et va servir de base aux traitements plus complexes des cellules des couches supérieures.

Au contraire des cellules horizontales, dont la fonction est simple et identique pour toutes ces cellules, les cellules amacrines forment une classe d'une trentaine de types de cellules différentes et réalisent des traitements très variés [Masland, 2001b]. Ces traitements sont plus ou moins bien compris et on peut citer par exemple le fait que certaines cellules amacrines amènent les cellules ganglionnaires à décharger de façon simultanée, ce qui permet un multiplexage des informations en sortie de la rétine [Meister, 1996].

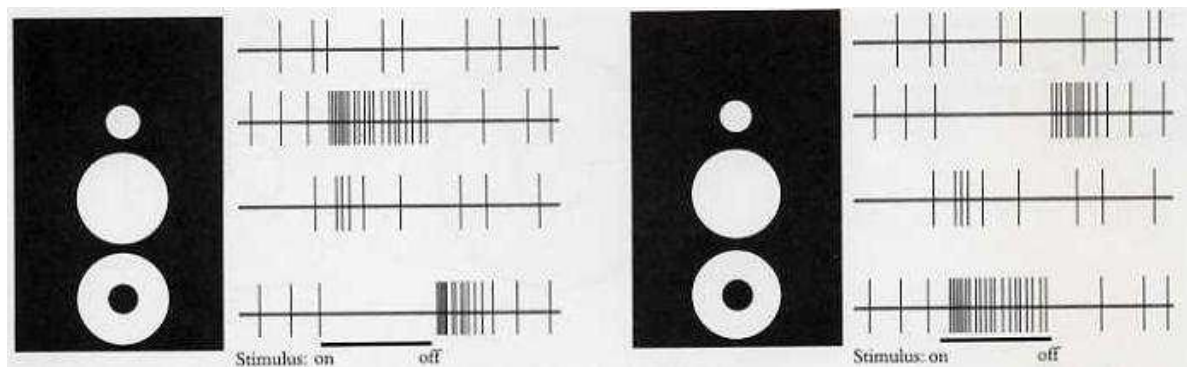


FIGURE 1.8 Illustration de la réponse d'une cellule ganglionnaire lorsque son champ récepteur est stimulé. À gauche : cellule ganglionnaire à centre *on* et à pourtour *off*. Sur la première ligne, on peut voir les potentiels d'action (barres verticales) émis en l'absence de stimulation, appelés activité spontanée. La deuxième ligne montre la réponse impulsionnelle de la cellule quand la stimulation est centrée. Sur la troisième ligne, c'est tout le champ récepteur qui est stimulé. La dernière ligne montre la réponse de la cellule si l'on stimule uniquement le pourtour de la cellule. À droite : *idem* pour une cellule ganglionnaire à centre *off* et à pourtour *on*. [Source : [Hubel, 1988]]

Les cellules ganglionnaires émettent des trains de potentiels d'action qui encodent le résultat de l'ensemble des traitements rétinien. Leur activité est une fonction complexe de l'ensemble des transformations réalisées par les différents types de cellules rétinien [Meister et Berry, 1999]. La contribution de chacun des types cellulaires n'est pas abordée ici. En résumé, la diversité et le fonctionnement des photorécepteurs et des cellules horizontales sont aujourd'hui bien compris, alors que la compréhension des cellules bipolaires, amacrines et ganglionnaires est plus parcelaire et nécessite encore une large exploration systématique [Troy et Shou, 2002; Masland et Martin, 2007].

L'étude des cellules ganglionnaires est très liée à la notion de champ récepteur. Le champ récepteur d'une cellule désigne la partie du champ visuel pour laquelle la cellule réagit quand on applique une stimulation lumineuse. Historiquement, les travaux

de Kuffler [Kuffler, 1953] ont montré qu'il était possible de distinguer les cellules ganglionnaires à centre *on* et à pourtour *off* et vice-versa. Les travaux de Hubel et Wiesel [Hubel et Wiesel, 1960; Hubel et Wiesel, 1962; Hubel et Wiesel, 1968] ont étendu ces observations et ont confirmé l'existence de champs récepteurs concentriques. Les réponses impulsionnelles de ces cellules ganglionnaires sont illustrées sur la figure 1.8.

Ces cellules ont été largement étudiées et il existe plusieurs classifications et dénominations pour les distinguer. L'historique de ces classifications et l'étude approfondie des stimuli qui déclenchent leur activation sortent du cadre de cette thèse et ces aspects sont très bien expliqués dans différents ouvrages, cités dans l'introduction de cette partie. Pour les modèles proposés dans le chapitre 3, il est intéressant de savoir que dès la sortie de la rétine certaines classes de cellules ganglionnaires peuvent réagir à des stimuli relativement complexes comme par exemple à la direction du mouvement ou à des contours orientés locaux [Barlow et Levick, 1965; Levick, 1967]. Nous reparlerons aussi d'une caractéristique structurelle de la rétine : dès le niveau des cellules bipolaires il est possible de distinguer différentes voies de traitement de l'information. Ces voies portent des informations spatio-temporelles distinctes – comme les contours ou les oppositions de couleurs – et sont indépendantes et parallèles [Wässle, 2004]. Dans la partie suivante 1.2.3, nous insisterons sur deux voies particulières, les voies parvo- et magnocellulaire, en sortie de la rétine, qui donnent lieu à des traitements séparés dans les aires corticales.

Conclusion sur la rétine

En conclusion, il est possible de mettre en avant certains points concernant la rétine et les traitements qui y sont réalisés. Tout d'abord, la rétine constitue un ensemble de neurones qui, bien qu'ayant une certaine indépendance vis-à-vis du cortex, réalise une partie importante des traitements visuels. En ce sens, la rétine est bien plus qu'un capteur amélioré, rôle qui lui a été prêté pendant longtemps. Les traitements rétinien représentent une part importante de la perception visuelle et font plus qu'amorcer les traitements corticaux [Masland, 2005; Masland et Martin, 2007]. Ceci est d'autant plus impressionnant que la rétine fonctionne en vase clos, c'est-à-dire qu'elle ne reçoit de projection d'aucune aire corticale. Les traitements complexes qui s'y déroulent sont largement dus aux connexions latérales que l'on peut observer à chaque étage (voir la figure 1.6). Les observations physiologiques montrent que pour chaque type de cellules (photorécepteurs, horizontales, bipolaires, amacrines et ganglionnaires), il existe des sous-types de cellules qui forment des réseaux distincts, assez peu interconnectés, couvrant de façon plus ou moins homogène le champ visuel. La rétine est donc plus

proche d'une superposition de réseaux que d'un seul réseau homogène et complètement connecté. Ainsi, dès le passage de l'information dans les cellules bipolaires, différentes voies parallèles se distinguent et réalisent différentes transformations spatio-temporelles de l'information visuelle. Selon le type d'information portée par ces voies, elles projettent vers différentes structures cérébrales, comme l'hypothalamus pour la régulation des fonctions végétatives ou le colliculus supérieur pour la coordination et le contrôle oculomoteur. La plupart des voies projettent cependant vers une même structure, le corps genouillé latéral puis le cortex et ce chemin qu'emprunte une partie importante de l'information en sortie de la rétine est donc appelé voie rétino-géniculo-striée.

1.2.3 Le corps genouillé latéral

Le corps genouillé latéral (CGL) est une structure sous-corticale qui reçoit une grande partie des informations sortant de la rétine. Cet ensemble de neurones fait partie d'une structure plus importante, qui intervient dans le traitement des informations perceptives : le thalamus. Le thalamus est le relais sensoriel du SNC, car tous les sens hormis l'odorat possèdent un relais thalamique. C'est donc en quelque sorte la porte d'entrée sensoriel du SNC.

Le CGL a longtemps été considéré comme un « simple » relais puisque les cellules qui le composent répondent d'une manière analogue aux cellules ganglionnaires de la rétine. Le terme de relais est cependant à manipuler avec précaution, le CGL est en effet bien plus que cela, comme le montrent par exemple les connexions de ses neurones,

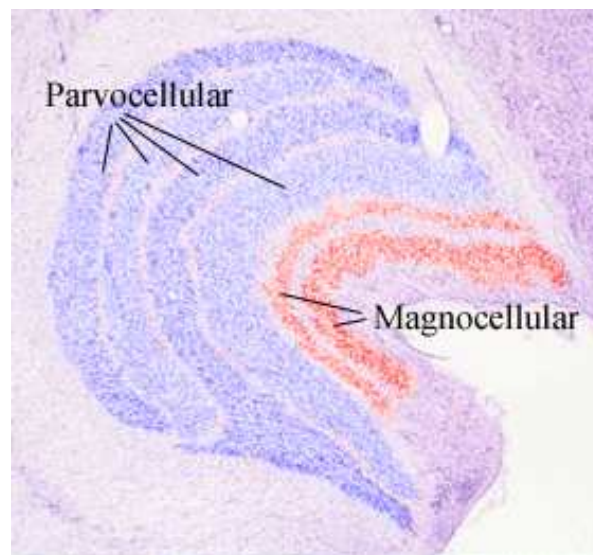


FIGURE 1.9 Microphotographie du corps genouillé latéral où une coloration artificielle distingue les couches parvocellulaire et magnocellulaire. [Source : [Bases of Colour Vision](#), B.T. Wagner & D. Kline, Vision and Aging Lab]

qui reçoivent en majeure partie des projections des aires visuelles supérieures et qui possèdent de nombreuses connexions latérales – c'est-à-dire du CGL sur lui-même. Ces connexions semblent indiquer que le thalamus joue un rôle important dans l'intégration précoce des informations des aires corticales de haut niveau. En particulier, le fait que les neurones du CGL puissent émettre des potentiels d'action de façon tonique ou phasique et qu'il semble que le passage d'un mode de décharge à l'autre puisse être contrôlé par les afférences corticales montre que le CGL fait bien plus que relayer simplement les informations en sortie de la rétine [Guillery et Sherman, 2002; Sherman et Guillery, 2003].

Le CGL possède une organisation anatomique particulière : il est structuré en couches superposées qui sont représentées sur la figure 1.9. Ces couches se distinguent par les projections qu'elles reçoivent de la rétine et en particulier de deux types de cellules ganglionnaires distinctes : les cellules M pour magnocellulaire et P pour parvocellulaire. Ces cellules ont des caractéristiques très différentes : les cellules M ont un large champ récepteur, répondent à des changements rapides de luminance et sont sensibles aux petites variations de contraste alors que les cellules P ont de petits champs récepteurs, au point que dans la fovéa certaines cellules ganglionnaires reçoivent les informations d'un seul cône [Masland, 2001b; Wässle, 2004]. Ceci illustre bien le compromis qui existe entre la précision spatiale et la sensibilité au changement de contraste pour les systèmes biologiques. En effet, plus les champs récepteurs sont larges, plus les cellules peuvent détecter les changements de contrastes fins en sommant l'activité sur l'ensemble du champ récepteur. Évidemment, une cellule avec un grand champ récepteur est moins précise spatialement qu'une cellule qui ne reçoit ses entrées que d'un seul photorécepteur.

Comme des types d'informations différents transitent par les couches magno- et parvocellulaires du CGL et arrivent toujours de façon séparée dans les aires corticales, on emploie souvent les termes de voies magno- et parvocellulaires⁴. La voie parvocellulaire semble porter plutôt des informations d'une grande acuité sur les couleurs et concerne plutôt la vision centrale. La voie magnocellulaire semble traiter plutôt des informations sur les zones périphériques du champ visuel, concerne la détection des mouvements rapides, ne porte pas d'information sur les couleurs et les axones de ses neurones ont des vitesses de conduction plus élevées, ce qui induit que les informations visuelles de la voie magnocellulaire arrivent de façon précoce dans les aires visuelles.

4. Une « troisième » voie, appelé koniocellulaire, existe. Cette voie est beaucoup moins bien connue et étudiée. Les traitements complexes médiés par cette voie sont en partie détaillés dans [Hendry et Reid, 2000].

1.2.4 Les aires visuelles

La plupart des informations visuelles traitées par le CGL arrivent dans l'aire visuelle primaire ou V1, et constituent la voie rétinogéniculostriée – de la rétine au corps genouillé latéral puis vers le cortex strié, c'est-à-dire V1. On retrouve au niveau anatomique et fonctionnel la séparation entre les voies magno- et parvocellulaire que nous avons vue. C'est vers V1 que la plupart des informations visuelles convergent et c'est de loin l'aire corticale la plus étudiée du système visuel. En particulier, l'étude de la structure de V1, de son organisation topologique, a donné lieu à de nombreux travaux. Une des caractéristiques les plus marquantes est que, pour un hémichamp donné, la topologie de V1 est similaire à celle de la rétine ; ainsi si un motif particulier est projeté sur la rétine, les neurones activés par ce stimulus dans V1 ont une disposition similaire au motif projeté sur la rétine, comme illustré sur la figure 1.10. La relation topologique qui lie la rétine à V1 est de type log polaire [Tootell *et al.*, 1982], car la zone fovéale est sur-représentée et les zones périphériques sont d'autant moins représentées que l'on s'éloigne de la fovéa.

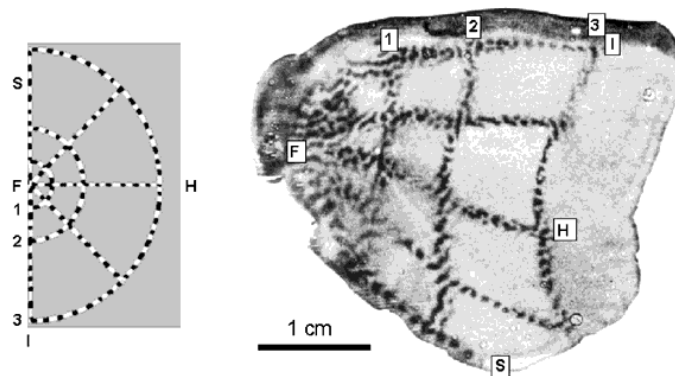


FIGURE 1.10 À gauche, stimulus visuel présenté pendant l'expérience. À droite, les neurones de V1 activés par la présentation du stimulus sont noircis par des marqueurs métaboliques. Il apparaît ainsi que V1 est rétinotopique car il présente la même topologie que la rétine. On voit aussi que la zone fovéale implique beaucoup plus de neurones que les zones périphériques. [Source : [Tootell *et al.*, 1982]]

La notion de champ récepteur est aussi utilisée pour caractériser les neurones de V1 et Hubel et Wiesel [Hubel et Wiesel, 1962] ont montré que de nombreux neurones de V1 répondaient fortement quand on leur présentait des stimulus avec une orientation particulière dans leur champ récepteur. Toutes les cellules ne réagissent pas à la même orientation ni de la même façon ; ainsi on distingue classiquement trois types de cellules. Les cellules simples possèdent un champ récepteur *on-off* allongé et permettent ainsi de détecter les lignes orientées. Les cellules complexes répondent à des lignes de contrastes orientés et sont souvent spécifiques à une direction particulière du mouvement. En effet,

ces cellules ne présentent pas de répartition particulière de leurs zones *on* et *off* et sont sensibles à l'orientation dans une région spatiale donnée. Le dernier type regroupe les cellules hypercomplexes ou *end-stop* donnant une réponse maximale quand un stimulus recouvre seulement une partie du champ récepteur, ce qui les rend sensibles aux fins de lignes ou à la courbure [Gilbert et Wiesel, 1989]. Une revue récente des propriétés des cellules de V1 est proposée dans [Hirsch et Martinez, 2006].

Les neurones de V1 sont donc organisés de façon rétinotopique et pour chaque position spatiale les neurones forment des hypercolonnes pour traiter différents types d'informations visuelles. Ces hypercolonnes (figure 1.11) regroupent des neurones dont les champs récepteurs occupent une même région spatiale mais qui sont sensibles à des informations différentes comme l'orientation, la dominance oculaire⁵, l'opposition de couleur [Kandel *et al.*, 2000], la direction principale du mouvement ou la fréquence spatiale [Purves *et al.*, 1997]. Cette organisation en hypercolonnes permet des interactions latérales et met en compétition les différentes localisations spatiales. On a supposé pendant longtemps que V1 recevait la plupart de ces informations du CGL, alors que les interactions latérales sont extrêmement nombreuses et que V1 reçoit aussi beaucoup d'informations d'autres aires corticales, comme le montre la figure 1.12. Pour donner un ordre de grandeur, environ 10% des neurones corticaux reçoivent des informations sensorielles du thalamus, le reste de l'information provient d'interactions locales ou distantes [Douglas et Martin, 2004].

Malgré cette structure à la topographie apparemment bien organisée, avec des zones de dominance oculaire et des hypercolonnes, bien des résultats expérimentaux restent difficiles à expliquer. Il est en effet assez tentant de se représenter V1 comme une collection de filtres qui entrent en compétition pour trouver les informations les plus pertinentes. La réalité est beaucoup plus complexe. Les neurones répondent par des potentiels d'action et non par des activations graduées, ce qui apporte une forte non linéarité à leurs réponses, en plus du fait qu'ils adaptent ces réponses au cours du temps et qu'ils sont très largement influencés par les interactions locales. Il reste donc encore beaucoup à apprendre de V1 [Olshausen et Field, 2005].

Les aires corticales impliquées dans la vision sont nombreuses et distribuées sur toute la surface corticale. La figure 1.12 propose un schéma de ces aires corticales ainsi que des connexions connues qui les relient. Elle illustre bien le fait qu'il est difficile de comprendre le rôle particulier de chacune d'elle étant donné les influences réciproques qu'elles peuvent exercer les unes sur les autres. Il est cependant couramment admis qu'elles traitent l'information à des niveaux d'abstraction distincts et que certaines

5. c'est-à-dire sur quel œil sont principalement situés les champs récepteurs

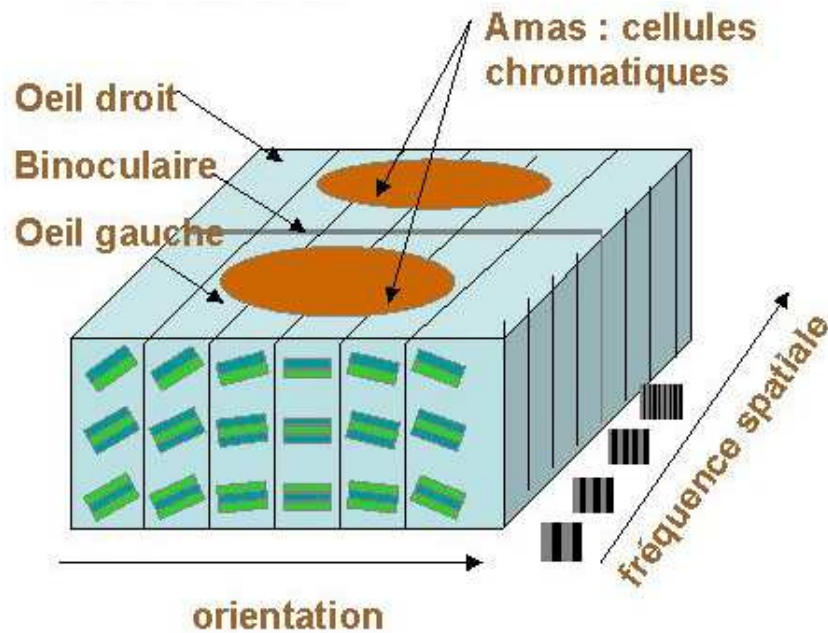


FIGURE 1.11 Schéma représentant la structure d'une hypercolonne de V1.

aires semblent être dévolues à des traitements plus spécifiques, comme par exemple la perception du mouvement, le choix du stimulus le plus pertinent ou la reconnaissance d'objets spécifiques.

1.3 L'attention visuelle

Avant de commencer cette partie, il est important d'établir une distinction entre les sens ascendant et descendant du traitement de l'information. Les processus ascendants, ou *bottom-up*, font référence à tous les traitements qui utilisent uniquement les informations entrantes. Ce sont des processus complètement dirigés par les données (*data-driven*). Les processus descendants, ou *top-down*, désignent les traitements qui intègrent des éléments de décision issus de traitements de plus haut niveau. Par exemple, l'influence d'un but général ou d'un contexte connu durant la réalisation de la tâche en cours. Contrairement aux processus ascendants, ils impliquent généralement l'utilisation de connaissances préalables, comme dans le cas où l'on cherche à retrouver un objet particulier – ses clefs par exemple – dans une scène visuelle.

Lors d'une recherche visuelle, le déplacement du focus d'attention peut prendre deux formes appelées *overt attention* et *covert attention*. La première désigne les cas où l'œil réalise une saccade pour mettre la région désirée dans le focus d'attention, par opposition à la seconde pour laquelle le focus d'attention se déplace sans impliquer de saccades

withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called distraction, and Zerstreutheit in German. »

La mise en évidence expérimentale de certains mécanismes attentionnels à partir des années 1980 [Posner *et al.*, 1980] a permis de proposer différentes théories de l'attention, que nous verrons dans la partie 1.3.3. Ces théories supposent, pour la plupart, que la sélection attentionnelle soit spatiale. Cette sélection spatiale est souvent décrite en utilisant une métaphore, celle du faisceau attentionnel [Crick, 1984] : l'attention se porterait sur différentes régions de l'espace comme le ferait le faisceau d'une lampe qui éclairerait un tableau. Ainsi l'attention pourrait être portée sur une seule région à la fois⁶, appelée focus d'attention, et se déplacerait dans le champ visuel d'une région d'intérêt à l'autre, de façon indépendante des mouvements oculaires [Pylyshyn et Storm, 1988].

1.3.1 Être attentif pour percevoir

Si l'on accepte l'hypothèse de la focalisation attentionnelle unique, ceci implique que l'environnement visuel soit exploré de façon séquentielle, le focus d'attention se déplaçant d'une région spatiale à l'autre sans jamais pouvoir embrasser l'intégralité de la scène visuelle « d'un seul coup d'œil ». Que se passe-t-il si une partie de la scène visuelle est modifiée brutalement pendant une saccade oculaire ou pendant le clignement des yeux ? Est-il possible de détecter ce changement ? Dans [Rensink *et al.*, 1997; Rensink *et al.*, 2000], les auteurs ont proposé une expérience simple, illustrée par la figure 1.13, dans laquelle un sujet voit une image puis la même image modifiée. Si on intercale un écran blanc entre les images⁷, ces changements sont extrêmement difficiles à détecter, même s'ils occupent une partie importante de l'image ou qu'ils sont présentés de façon répétée. Ce phénomène, appelé cécité aux changements ou *change blindness*, semble confirmer qu'une focalisation attentionnelle est nécessaire pour percevoir explicitement les changements [Rensink, 2000].

C'est d'autre part un outil expérimental intéressant pour évaluer différentes théories explicatives de la vision, en particulier concernant la nature des représentations internes. Dans [O'Regan, 1992], puis dans [O'Regan et Noë, 2001; Noë, 2002], les auteurs défendent l'idée que nos représentations internes ne sont pas des reconstructions

6. Plusieurs études ont montré que sous certaines conditions expérimentales l'attention pouvait être « divisée » entre plusieurs cibles [Pylyshyn et Storm, 1988; Pylyshyn, 1998; Pylyshyn, 2001]

7. L'ajout d'un masque entre la présentation des images produit un effet similaire à celui d'un clignement des yeux. Ces résultats sont confirmés par le fait qu'on obtient le même type de résultat en changeant d'images quand le sujet cligne des yeux ou fait une saccade.

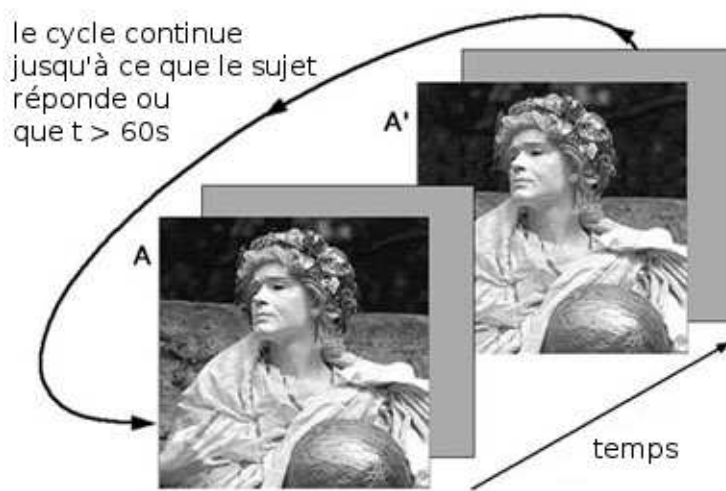


FIGURE 1.13 Le paradigme des images clignotantes (*flicker images*) met en évidence de façon très simple la cécité aux changements. Le dispositif expérimental affiche alternativement les images A et A' en intercalant un rapide écran blanc après la présentation de chaque image. L'image A' comporte une modification par rapport à A , ici c'est le mur derrière le personnage qui est abaissé. La présentation alternée des deux images continue jusqu'à ce que le sujet réponde ou que le temps limite soit dépassé. Ces changements sont très difficiles à détecter et il apparaît que dans beaucoup de cas les sujets ne voient pas ce changement alors même que leur point de fixation oculaire est précisément dessus [O'Regan *et al.*, 2000]. [Source : [Rensink *et al.*, 2000]]

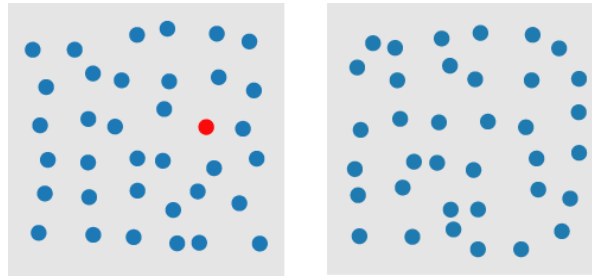


FIGURE 1.14 Exemple d'image utilisée dans une tâche de recherche visuelle préattentive. La cible à trouver possède un trait caractéristique propre (ici la couleur rouge) et les autres objets sont des distracteurs (en bleu). La tâche consiste à trouver si la cible est présente ou non, en répondant le plus vite possible. Sur cette figure, la cible est présente à gauche, mais ne l'est pas à droite. Les résultats montrent que la précision est très bonne – il y a très peu d'erreurs – et que le temps de réponse des sujets ne varie pas quel que soit le nombre de distracteurs. On parle donc de *pop out* pour décrire la rapidité et l'aisance avec lesquelles il est possible d'accomplir cette tâche. [Source : [Healey, 2007]]

complètes de ce que nous voyons, mais plutôt une représentation clairsemée que nous pouvons compléter si nécessaire en déplaçant nos yeux ou notre attention : le monde peut être ainsi vu comme une « mémoire externe » que nous interrogeons au gré de la précision des informations dont nous avons besoin pour interagir avec notre environnement [O'Regan, 1992]. Une présentation du phénomène de cécité aux changements – avec des exemples vidéos – et de ses différentes implications théoriques est disponible sur la page Internet [O'Regan et Noë, 2000] et l'état de l'art de [Simons, 2000] propose une bonne vue d'ensemble des enjeux théoriques.

1.3.2 Mise en évidence expérimentale de la préattention

Neisser a introduit l'idée d'une étape préattentive [Neisser, 1967], c'est-à-dire de vision avant l'attention, avant que n'intervienne l'influence des processus descendants. Cette étape préattentive a été largement étudiée en psychologie – les premiers travaux ont été réalisés par Treisman [Treisman et Souther, 1985; Treisman, 1985; Treisman et Gormican, 1988; Treisman, 1991] – et il existe plusieurs théories explicatives que nous présenterons par la suite (partie 1.3.3). Cette étape préattentive est caractérisée par le fait qu'un ensemble de propriétés visuelles que nous appellerons traits caractéristiques, ou *preattentive features* chez [Treisman, 1985], sont détectés de façon très rapide et très fiable par le système visuel. Ces traits caractéristiques sont bien mis en évidence par le phénomène de *pop out* lors d'une tâche de recherche visuelle, comme illustré sur la figure 1.14.

Ce type de tâche utilise une image avec une cible et des distracteurs qui ne diffèrent

que par un trait caractéristique (la couleur dans la figure 1.14). Le temps nécessaire pour trouver la cible ne dépend pas du nombre de distracteurs. Ainsi pour trouver un cercle rouge parmi des cercles bleus, un sujet est toujours aussi rapide qu'il y ait 5, 10 ou 100 distracteurs. Ces résultats semblent indiquer que les traitements réalisés durant cette étape préattentive sont effectués en parallèle sur l'ensemble de la scène visuelle.

Le phénomène de *pop out* ne concerne que les cas où la cible possède un trait caractéristique unique, qui n'est pas partagé par les distracteurs. Dans le cas où la cible est définie par la conjonction de traits caractéristiques, comme cercle et rouge sur la figure 1.15, et que les distracteurs possèdent seulement l'un ou l'autre de ces traits, par exemple carré rouge ou cercle bleu, la cible n'est plus aussi facilement détectée.

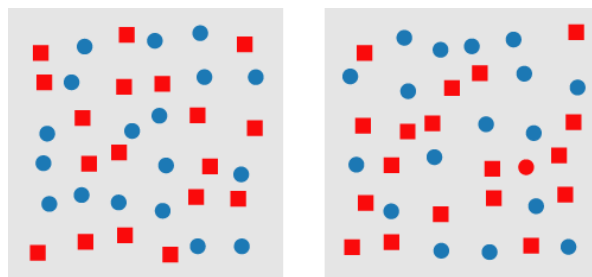


FIGURE 1.15 Exemple d'image utilisée dans une tâche de recherche visuelle attentive. Cette tâche est appelée *conjunction search* car la cible est le seul objet à posséder la réunion de deux traits caractéristiques (ici rond et rouge). Il y a deux types de distracteurs et chaque type ne partage qu'un seul trait caractéristique avec la cible : les ronds bleus et les carrés rouges. Les temps de réponse mesurés montrent que la recherche est d'autant plus longue qu'il y a de distracteurs. Le système visuel préattentif ne permet pas de fournir une réponse rapide, il est donc nécessaire de faire une recherche sérielle jusqu'à trouver le bon objet. [Source : [Healey, 2007]]

Dans cette tâche, appelée *conjunction search*, le sujet ne peut pas utiliser un seul trait pour trouver la cible : s'il cherche les objets rouges, les distracteurs carrés rouges vont rendre plus difficile sa recherche, de même avec les cercles bleus lorsqu'il cherche les objets circulaires. Il est donc nécessaire d'engager l'attention puisque les traitements préattentionnels ne sont pas suffisants pour détecter directement la cible, ce qui impose de faire une recherche sérielle en portant l'attention sur les différents objets⁸. Le temps nécessaire pour trouver la cible est donc nettement plus long et dépend directement du nombre de distracteurs.

Les traits caractéristiques impliqués dans les traitements préattentifs sont nombreux. Nous avons vu que la couleur en est un et d'autres sont présentés sur la figure 1.16. Une liste non exhaustive inclurait par exemple la couleur, l'orientation, la courbure, la taille,

8. Les mécanismes préattentifs jouent cependant un rôle dans la *conjunction search* car ils permettent de réduire l'espace de recherche à un groupe d'objets partageant un trait caractéristique.

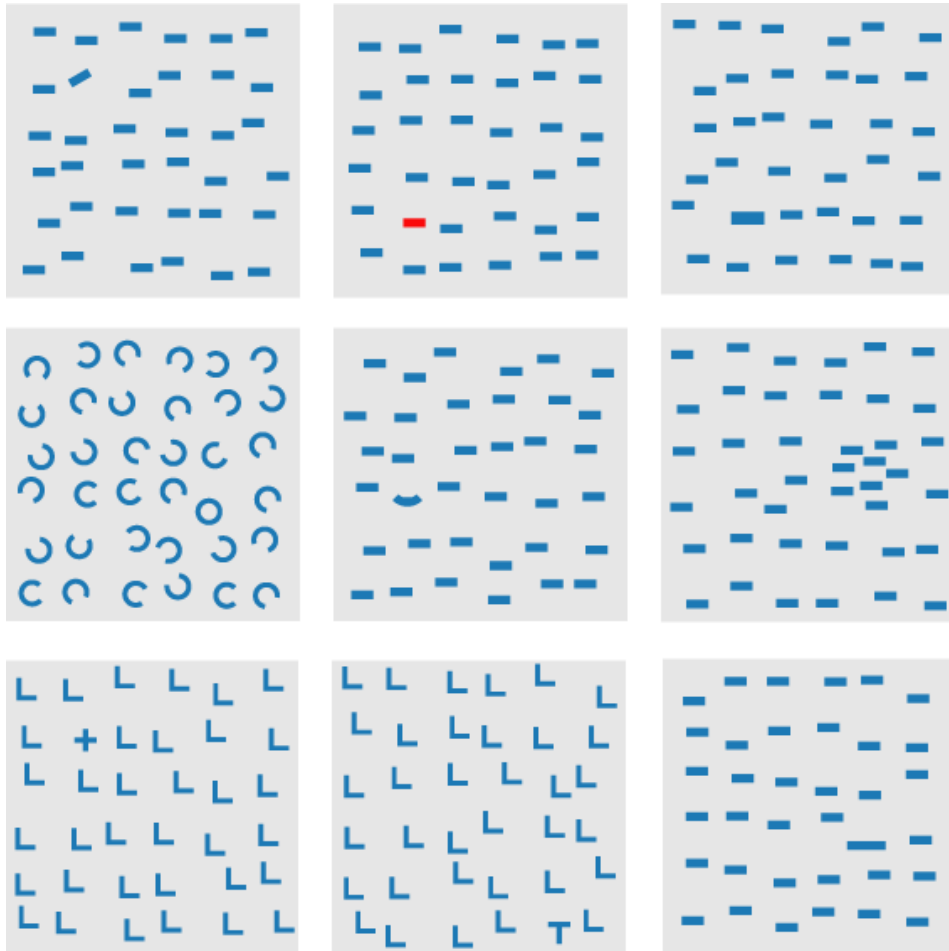


FIGURE 1.16 Illustration de quelques traits caractéristiques (ou *pre-attentive features* [Treisman, 1985]). De haut en bas et de gauche à droite : l'orientation, la couleur, la taille, les figures fermés, la courbure, la densité, l'intersection, la terminaison et la longueur. [Source : [Healey, 2007]]

le mouvement, les indices de profondeur et pour les formes : les terminaisons de lignes, les intersections, les figures fermées ou la présence de trous. Une liste plus complète ainsi qu'un état de l'art relatif aux traits caractéristiques sont proposés dans [Wolfe, 1998; Wolfe et Horowitz, 2004]. Ces traits caractéristiques ne sont pas tous équivalents : certains traits sont détectés plus rapidement que d'autres, comme le mouvement, la couleur ou l'orientation par exemple [Wolfe, 2000]. Pour expliquer comment ces traits caractéristiques sont utilisés dans la vision préattentive, nous allons voir les quatre théories qui ont été proposées : la *feature integration theory* [Treisman et Gelade, 1980], la *similarity theory* [Duncan et Humphreys, 1989], la *guided search theory* [Wolfe, 1994] et la *texton theory* [Jülész, 1986].

1.3.3 Les théories de la préattention

La *feature integration theory* est une théorie majeure de la vision et a influencé de nombreux travaux en perception visuelle et artificielle ; une explication détaillée et chronologique est disponible dans [Quinlan, 2003]. Dans [Treisman et Gelade, 1980; Treisman, 1988], les auteurs proposent un modèle de la vision bas-niveau comportant des cartes de traits caractéristiques et une carte topologique (*map of locations*, voir figure 1.17). Chaque carte de trait caractéristique regroupe les activités relatives à un trait particulier, comme les oppositions de couleurs, les orientations, etc. Ces cartes traitent les informations visuelles de façon parallèle et la disposition spatiale des traits est codée dans un système qui est propre à chaque carte de trait caractéristique [Treisman, 1985]. La carte topologique permet de mettre en relation la présence de traits caractéristiques et une localisation spatiale exprimée dans un référentiel commun. Cette théorie propose donc que si une cible est caractérisée par un trait unique, il est possible de la détecter facilement et comme les cartes fonctionnent en parallèle les unes aux autres, cette détection est quasi instantanée. Dans le cas d'une *conjunction search*, les informations disponibles sur les cartes de traits caractéristiques ne sont pas suffisantes et le focus attentionnel se déplace sur la carte topologique pour trouver la cible. Le traitement effectué en parallèle sur chaque carte de traits caractéristiques peut être réalisé plus ou moins vite en fonction de la différence entre la cible et les distracteurs : par exemple un grand cercle parmi des petits cercles est détecté plus facilement si la différence de taille est importante [Treisman et Gormican, 1988; Treisman, 1991].

Une autre théorie propose une explication plus détaillée de la différence de temps de traitement des traits caractéristiques : dans la *similarity theory* [Duncan, 1989; Duncan et Humphreys, 1989] les auteurs défendent l'idée que seules les similarités cible-distracteurs (C-D) et distracteurs-distracteurs (D-D) importent pour expliquer les temps de réponse. Plus la similarité C-D est importante, plus il est difficile et long de trouver la cible et plus la similarité D-D est faible (c'est-à-dire que les distracteurs sont variés), plus il est difficile et long de trouver la cible. Les auteurs ont proposé une théorie dans laquelle le champ visuel est segmenté en une hiérarchie d'unités structurelles. Les unités structurelles sont des régions de l'espace partageant des traits caractéristiques communs et chaque unité structurelle peut être subdivisée en unités structurelles plus petites. Le champ visuel est ainsi analysé de façon parallèle pour produire une représentation hiérarchique. La recherche visuelle est définie comme un processus aux ressources limitées qui compare le patron de l'information recherchée aux différentes unités structurelles et ce processus affecte d'autant plus de ressources à une unité structurelle qu'elle est

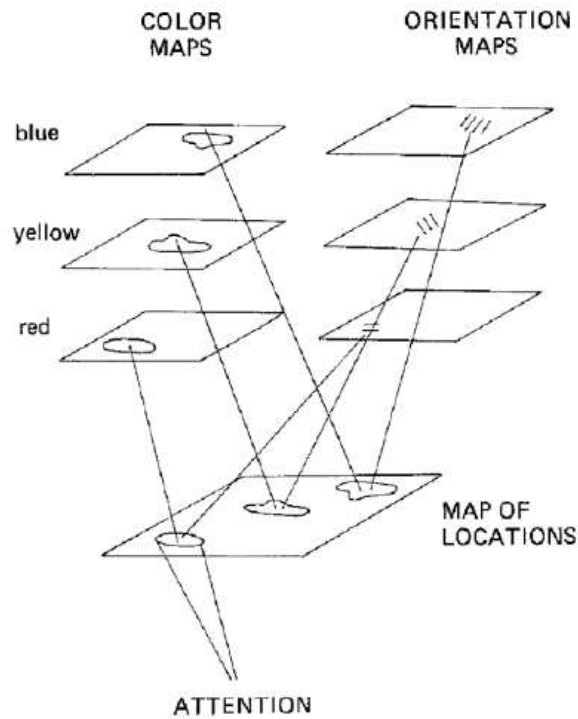


FIGURE 1.17 Schéma explicatif de la *Feature Integration Theory* [Treisman et Gelade, 1980]. Les cartes de traits caractéristiques (ici couleur et orientation) traitent en parallèle les informations de chaque modalité visuelle et peuvent être utilisées en vision préattentive (phénomène de *pop out*). La *Feature Integration Theory* postule que ces cartes sont reliées à une carte topologique (*map of locations*) qui combine les informations disponibles sur les cartes de traits caractéristiques. C'est cette carte qui est utilisée pour la focalisation attentionnelle. [Source : [Treisman et Souther, 1985]]

proche du patron.

La *guided search theory* [Wolfe *et al.*, 1989; Wolfe, 1994] propose quelques modifications par rapport à la *feature integration theory* et développe une explication pour la prise en compte des influences descendantes. Cette théorie postule l'existence d'une carte pour chaque type de caractéristiques, au lieu d'une carte par caractéristique comme dans la *feature integration theory*. Par exemple, sur la carte des couleurs coexistent des représentations indépendantes pour les couleurs bleu, rouge, verte et jaune (voir figure 1.18). Cette étape de catégorisation des différentes composantes de chaque carte est réalisée en parallèle. Les composantes de ces cartes sont ensuite combinées sur une carte d'activation (*activation map* sur la figure 1.18), qui permet de choisir les cibles de la focalisation attentionnelle. La plus grande originalité de cette théorie est de proposer une explication de l'influence des mécanismes ascendants et descendants pour guider la recherche visuelle. La préattention, processus d'activation ascendant, mesurerait les différences locales entre un élément et ses voisins et serait intégrée aux traitements réalisés en parallèle dans les différentes cartes (*local diff.* sur la figure 1.18). Un autre processus, descendant cette fois, biaiserait l'activation de certaines composantes pour trouver des éléments aux caractéristiques spécifiques. Ainsi la recherche visuelle d'un objet d'une couleur et d'une orientation particulière va renforcer l'activation de ces composantes, ce qui va modifier la carte d'activation et prendre en compte ces critères de recherche pour déterminer quelles régions explorer.

Une approche alternative consiste à décrire l'image non plus en termes de traits caractéristiques mais en fonction de la statistique des textures présentes dans les images. Dans [Julész, 1981; Julész, 1986], l'auteur propose une théorie reposant sur les *textons*, qui sont des groupements de traits caractéristiques, définis comme des formes allongées (des morceaux de ligne par exemple) de couleur, de longueur, de largeur et de taille fixe. Dans la théorie proposée, seules les différences entre les densités de *textons* sont détectées préattentivement, la vision préattentive permet donc de définir la position des frontières entre différents *textons* mais l'attention est nécessaire pour identifier les *textons*.

1.4 Synthèse

Nous avons vu dans ce chapitre le fonctionnement physiologique des neurones, puis les structures neuroanatomiques qui traitent les informations visuelles. Les connaissances actuelles sur le fonctionnement du système visuel sont encore assez limitées : en particulier la plupart des études en neurophysiologie utilisent des stimuli artificiels

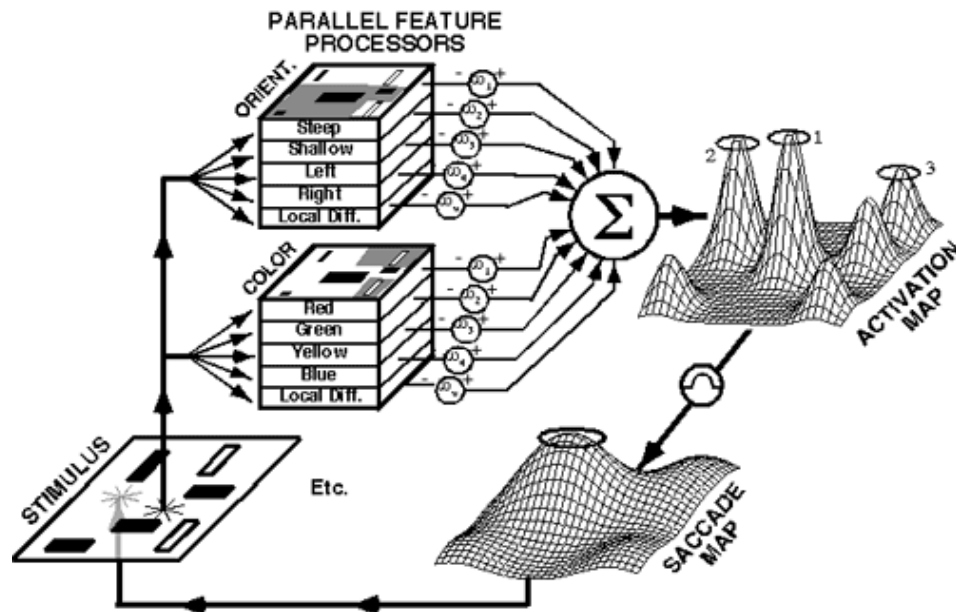


FIGURE 1.18 Schéma récapitulatif de la *guided search theory*. Un stimulus visuel est traité en parallèle par différentes cartes, chaque carte regroupant différents types de traits caractéristiques préattentifs (ici : couleurs et orientations). Les activations des différentes composantes de chaque carte sont combinées sur la carte d'activation, qui permet de sélectionner les régions à explorer. L'influence des mécanismes ascendants et descendants sont directement pris en compte dans les cartes : le mécanisme ascendant de pré-attention mesure les différences locales entre un élément et ses voisins, le mécanisme descendant d'attention modifie les activations des composantes en fonction des caractéristiques de la cible recherchée. [Source : [Scholarpedia](#) [Itti, 2007] d'après [Wolfe, 1994]]

qui induisent des réponses très différentes de celles observées lorsque l'on présente des images naturelles [Frégnac *et al.*, 2005; Marre *et al.*, 2005; Masland, 2005]. La compréhension des mécanismes fins du codage de l'information visuelle nécessite encore un effort expérimental important [Meister et Berry, 1999; Olshausen et Field, 2005; Masland et Martin, 2007], en particulier pour expliquer les aspects dynamiques et adaptatifs de ce codage au niveau de la rétine [Wohrer *et al.*, 2006; Wohrer et Kornprobst, 2008] ou des aires visuelles supérieures [Olshausen et Field, 2004].

La partie 1.3 présente la notion d'attention visuelle et les théories qui y sont consacrées. Les expérimentations en psychologie cognitive ont montré que nous utilisons certains traits caractéristiques – comme la couleur, le mouvement, l'orientation, les terminaisons ou les croisements de lignes – pour aider ou guider l'attention vers les stimulus qui semblent intéressants [Wolfe et Horowitz, 2004]. Ces traits caractéristiques sont nombreux et ne sont pas tous équivalents dans leurs effets sur le processus attentionnel : la couleur, la taille ou l'orientation par exemple guident l'attention de façon significative. Les influences des autres traits caractéristiques sont plus ou moins évidents et nécessitent d'être plus amplement étudiées expérimentalement. Les traits caractéristiques sont utilisés dans les principales théories de l'attention visuelle (hormis celle de [Jülész, 1981]) qui proposent différentes hypothèses sur leur regroupement en cartes de traits caractéristiques, sur la façon dont ces cartes sont combinées pour former une carte de saillances et sur les niveaux d'influence des processus descendants.

Chapitre 2

Les modèles de neurones impulsionnels

Sommaire

2.1	Modèles explicatifs ou modèles <i>per se</i> ?	36
2.1.1	Du réel au modèle	36
2.1.2	Différents niveaux de représentation	38
2.1.3	Des inspirations pluridisciplinaires	40
2.1.4	Des modèles à taux de décharge	40
2.2	Les modèles de neurones impulsionnels	42
2.2.1	Caractéristiques des modèles de neurones impulsionnels	42
2.2.2	Le modèle d'intégrateur à fuite	43
2.2.3	Les variantes des modèles intègre-et-décharge	44
2.2.4	Le modèle SRM	46
2.3	Les modèles de synapses	48
2.3.1	Les courants synaptiques	49
2.3.2	La plasticité synaptique	50
2.4	Synthèse	51

Les liens qui existent entre neurobiologie et informatique sont très étroits : il est en effet très tentant de rapprocher la notion d'opération élémentaire de calcul pour un processeur et pour une cellule nerveuse. Il est alors assez aisé d'étendre ce raisonnement pour comparer cerveau et ordinateur. Cette vision des choses a influencé aussi bien les informaticiens [von Neumann, 1958] que les neurobiologistes [Schwartz, 1990] et comme nous le verrons dans la suite de ce chapitre, l'inspiration biologique a joué un rôle important dans la mise au point d'algorithmes d'apprentissage et de traitement de l'information. Certains de ces algorithmes s'appuient sur des modélisations formelles des neurones.

Bien que les modèles de neurones soient inspirés par la biologie, la notion d’« inspiration biologique » désigne à la fois différentes approches et différents niveaux de proximité avec les observations expérimentales. Au cours de cette thèse, nous avons étudié les neurones impulsionnels en essayant de comprendre s’ils permettaient de capturer les caractéristiques essentielles des cellules du système visuel, celles qui permettent au système de traiter les informations visuelles efficacement et rapidement. En cela, bien que nous soyons restés proche de la biologie et de la compréhension actuelle des mécanismes de la vision biologique, nous avons principalement cherché à évaluer les intérêts que présente l’utilisation de neurones impulsionnels en vision artificielle. De ce fait, si nous avons pris des libertés avec la réalité biologique, c’est pour proposer une adaptation efficace de ces mécanismes sur un ordinateur. L’acception du terme « inspiration biologique » utilisé dans ce manuscrit est définie de façon précise dans le chapitre 3.

Il existe des différences qualitatives importantes entre les neurones biologiques et les modèles formels, comme nous le verrons dans la partie 2.1. Pourtant, il est tout à fait possible de capturer une partie du comportement des neurones au travers de différents modèles, plus ou moins biologiquement réalistes. Nous verrons ensuite comment les modèles neuronaux, qui privilégiaient au départ le réalisme biologique, se sont progressivement éloignés puis rapprochés de l’inspiration biologique. Les modèles de neurones impulsionnels, qui tendent à revenir vers le « biologiquement plausible », permettent d’encoder et de traiter les informations de façon temporelle. Nous discuterons des modèles qui utilisent ce type de codage dans la partie 2.2 et des modèles de synapse dans la partie 2.3, avant de conclure ce chapitre avec la partie 2.4.

2.1 Modèles explicatifs ou modèles *per se* ?

2.1.1 Du réel au modèle

Une fois connu les différents aspects physiologiques des neurones, nous pouvons prendre un peu de distance pour décrire le traitement de l’information réalisé par ces cellules nerveuses. Un premier point de comparaison intéressant entre neurones biologiques et réalisations informatiques concerne leur vitesse de traitement. Les neurones biologiques sont lents, surtout si on les compare à des circuits intégrés : l’intégration des informations prend plusieurs millisecondes et même s’il suffit que très peu d’ions traversent la membrane pour modifier le potentiel, ce mécanisme est lent. D’autre part, la période réfractaire qui suit l’émission d’un potentiel d’action contraint grandement le nombre maximum de potentiels d’action que peut envoyer un neurone en un temps

donné. Il est ainsi rare de trouver des neurones qui dépassent une fréquence de décharge de 100 Hz, c'est-à-dire qui émettent plus de 100 potentiels d'action en 1 seconde.

De plus, une fois un potentiel d'action émis, sa vitesse de propagation dépend principalement de la taille et de la myélinisation¹ de l'axone. Les vitesses de propagation les plus basses sont atteintes dans un axone fin et non myélinisé et sont de l'ordre de 1 m/s. Ce type d'axone « lent » est très fréquent dans les aires corticales pour les connexions locales. Ainsi les neurones mettent un temps non négligeable à intégrer les informations entrantes, ils sont limités dans leur taux de décharge maximum en sortie et les potentiels d'action se propagent lentement entre les différents neurones.

Pourtant, si au niveau unitaire les neurones ne sont pas rapides, quand on les considère à l'échelle d'une population, ils sont beaucoup plus efficaces. L'ensemble des neurones du cerveau permet de traiter de grandes quantités d'information car sa structure est très bien adaptée : différents circuits de traitement spécialisé fonctionnent en parallèle.

L'observation des différents circuits corticaux semble montrer que les neurones d'une aire cérébrale sont généralement affectés à une seule et unique tâche. Les neurones ne sont pas « reprogrammables » à souhait. Bien que le cerveau montre une certaine plasticité, c'est-à-dire que les neurones puissent être « réaffectés » à une autre tâche, cette plasticité est limitée. Elle ne peut pas concurrencer les approches mettant en œuvre un apprentissage en ligne (ou *on-line*) comme celles proposées en intelligence artificielle.

Il est intéressant de constater que, parmi les contraintes auxquelles sont soumis les neurones, la consommation d'énergie, qui est une contrainte importante, est particulièrement basse. Les neurones utilisent en moyenne 10^{-15} Joule par opération [Churchland et Sejnowski, 1994], c'est-à-dire à chaque fois qu'un neurone en active un autre par l'intermédiaire d'une synapse. Si l'on compare cette donnée à celles concernant les microprocesseurs, elle est extraordinairement basse : un microprocesseur consomme en moyenne 10^{-7} Joule par opération, au sens d'opération élémentaire comme l'addition ou la multiplication. La différence est donc de 7 à 8 ordres de grandeur en faveur du cerveau.

Cette faible consommation d'énergie rend possible la mise en parallèle d'un nombre important de neurones. Ainsi quand les ordinateurs de bureau réalisent environ 10^9 opérations par seconde, les neurones d'une mouche domestique effectuent en moyenne 10^{11} calculs par seconde lorsque la mouche est au repos² [Churchland et Sejnowski,

1. un axone myélinisé est partiellement recouvert d'une gaine isolante électriquement, ce qui permet au potentiel d'action de se propager beaucoup plus vite [Purves *et al.*, 1997]

2. il est intéressant de constater que le microprocesseur le plus puissant d'un ordinateur actuel est

1994]. D'autre part, même lorsque le cerveau est en pleine activité, seule une petite partie des neurones est active.

L'observation des neurones a permis de développer des modèles neuronaux pour tester et valider des hypothèses formulées par les neurobiologistes. La mise au point de techniques d'observation de plus en plus performantes, spatialement et temporellement, pour étudier les structures et les fonctions cérébrales ont poussé les modèles à être plus réalistes biologiquement. Il est donc possible de mieux cerner le fonctionnement de ces unités de traitement neuronal et la compréhension de ces mécanismes fins peut inspirer de nouvelles approches pour le traitement de l'information ou la cognition artificielle. Nous allons voir les différents modèles neuronaux qui ont été proposés et la façon dont ils traitent l'information.

2.1.2 Différents niveaux de représentation

La mise au point d'un modèle suppose de choisir un niveau de représentation. Pour le SNC, il est classiquement admis que nous pouvons distinguer différents niveaux d'interactions, depuis les molécules jusqu'au système nerveux dans son ensemble.

Les modèles neuronaux s'intéressent à certains de ces niveaux d'interactions, nous en distinguerons trois : les niveaux microscopique, mésoscopique et macroscopique [Meunier et Paugam-Moisy, 2008] (voir figure 2.1). Le niveau microscopique désigne l'étude des propriétés d'une cellule. Au niveau mésoscopique on s'intéresse aux interactions entre ces cellules, c'est le niveau d'étude d'une population. Le dernier niveau, macroscopique, décrit les interactions entre les populations.

Dans le cadre de ce travail, nous avons pris comme hypothèse que les effets observables aux niveaux méso- et macroscopique sont le résultat de propriétés émergentes, induites par le niveau microscopique. Dans cette optique, nous avons choisi d'utiliser des modèles explicatifs, les neurones impulsionnels, pour évaluer les implications des choix de modélisation au niveau d'un neurone sur l'ensemble du réseau. En particulier, nous avons recherché pour le choix du modèle de neurone un ensemble de propriétés qui soit restreint mais suffisant pour obtenir les résultats souhaités au niveau du réseau.

En effet, même s'il est nécessaire de comprendre le niveau microscopique pour proposer une explication complète des observations expérimentales, la seule connaissance des mécanismes locaux ne suffit pas nécessairement à expliquer les observations macroscopiques. L'exemple le plus célèbre provient sans doute de l'étude des ganglions

celui de la carte graphique et qu'il est hautement parallèle. Cette puissance de calcul peut être mise à profit pour réaliser des simulations neuronales [Chariot *et al.*, 2006]

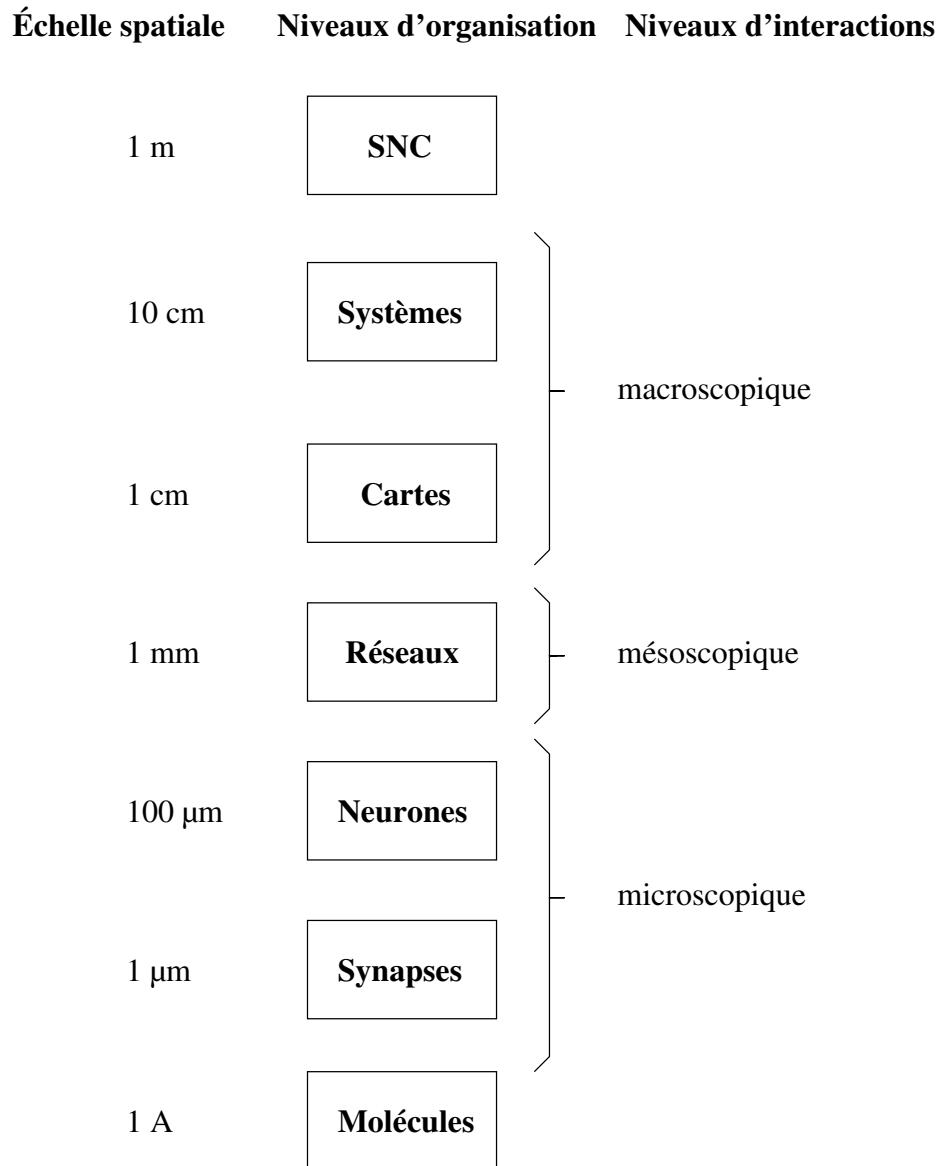


FIGURE 2.1 Schéma des différents niveaux d'organisation [Churchland et Sejnowski, 1994], avec leurs échelles spatiales respectives, ainsi que les niveaux d'interactions correspondants [Meunier et Paugam-Moisy, 2008].

stomatogastriques de la langouste : ce réseau de 28 neurones contrôle les contractions de l'estomac et sa sortie est une oscillation qui rythme les muscles gastriques. Les caractéristiques de chaque neurone sont très bien détaillées, car on peut retrouver les mêmes neurones et les mêmes connexions sur tous les individus. Pourtant, même une modélisation très complète de ce réseau ne permet pas forcément de reproduire l'activité globale observée. En effet, aucun neurone en particulier n'est responsable de l'oscillation, c'est la dynamique du réseau qui produit ce comportement [Selverston, 1988]. Cet exemple montre qu'une analyse uniquement ascendante ne permet pas toujours d'expliquer la totalité du phénomène. La dynamique d'un réseau dépend à la fois des paramètres de chaque unité et des interactions entre les unités qui le composent.

2.1.3 Des inspirations pluridisciplinaires

Les modélisations de neurones nécessitent de constants aller-retours entre plusieurs champs disciplinaires. Les observations faites à différents niveaux, que ce soit au niveau d'un bouton synaptique [Bi et Poo, 1998; Bi et Poo, 2001] ou du temps de réponse d'un sujet [Thorpe *et al.*, 1996], peuvent changer la vision que nous avons de certains mécanismes et nous amener à développer des modèles radicalement différents. L'approche qui consiste à proposer des modèles pour reproduire les données observées est similaire à celle du *reverse engineering*. Les modèles mis au point selon ce principe mettent à la disposition de la communauté scientifique des outils permettant la compréhension fine des mécanismes mis en jeu. Une fois ceux-ci compris, il est possible de concevoir des modèles *a priori* efficaces, tirant parti des mécanismes étudiés.

Les disciplines intervenant dans la mise au point d'un modèle de neurone sont nombreuses et une liste non exhaustive inclurait la psychologie cognitive, les neurosciences computationnelles et l'informatique. L'interdisciplinarité semble indispensable pour faire progresser la compréhension des mécanismes cérébraux et faciliter leur transfert pour diverses applications [Alexandre, 1997; Hérault, 1999; Cuperlier *et al.*, 2007]. La vision artificielle bio-inspirée, qui constitue le fil directeur de cette thèse, a bénéficié des apports de ce type d'approches interdisciplinaires [Beaudot *et al.*, 1993; Alleysson *et al.*, 2005].

2.1.4 Des modèles à taux de décharge

Une lecture historique des différents modèles de neurones montre que les premiers modèles se sont inspirés de la biologie pour mieux s'en abstraire ensuite, une fois que les propriétés importantes ont été identifiées et qu'il a été possible de construire des

modèles *a priori* qui soient efficaces. [McCulloch et Pitts, 1943] ont proposé un modèle mathématique de neurone, le premier neurone formel, dans lequel le neurone « filtre » les informations en appliquant une fonction non-linéaire sur la somme pondérée de ses entrées.

Dans le cas où la fonction non-linéaire utilisée est une fonction sigmoïde [Hopfield, 1984; Rumelhart et McClelland, 1986], il est possible d’interpréter la valeur résultante comme représentant l’activité moyenne du neurone, c’est-à-dire son taux de décharge. Cette interprétation du codage neuronal de l’information provient d’observations biologiques. Certaines cellules nerveuses sensorielles, comme les corpuscules de Pacini qui sont sensibles à la pression et aux vibrations, encodent l’intensité du stimulus en faisant varier leur taux de décharge. Dans le cortex, un tel codage a été observé par [Hubel et Wiesel, 1962] qui ont montré que les neurones du cortex visuel primaire font varier leur taux de décharge en fonction de l’orientation du stimulus visuel qui est présenté.

Les modèles de neurones et les différents réseaux de neurones artificiels (RNA), qui utilisent ce codage par taux de décharge, ont beaucoup évolué depuis la proposition de [McCulloch et Pitts, 1943]. Ils ont été largement utilisés en modélisation, que ce soit en biologie ou en psychologie. À mesure que les capacités de ces RNA ont été de mieux en mieux comprises, aussi bien expérimentalement que théoriquement, ils sont peu à peu devenus des outils de calcul et de modélisation. L’inspiration biologique s’est progressivement effacée et la communauté qui travaillait sur les RNA a progressivement perdu l’ancrage biologique des premiers modèles. L’étude des RNA sort du cadre de cette thèse et ces réseaux sont l’objet de très bons ouvrages, comme par exemple [Freeman et Skapura, 1991; Haykin, 1994; Bishop, 1995; Arbib, 1998; Dreyfus *et al.*, 2002].

Une approche alternative est de modéliser un réseau comme un “champ neuronal” : les neurones sont vus comme un champ continu de potentiels, dont l’activité en un point dépend de l’activité du voisinage [Amari, 1977]. L’évolution du potentiel de membrane u observé à la position spatiale $\mathbf{r} \in M$ est décrite par l’équation différentielle suivante :

$$\begin{aligned} \tau \frac{\partial u(\mathbf{r}, t)}{\partial t} = & -u(\mathbf{r}, t) + \int_M w_M(\mathbf{r} - \mathbf{r}') f[u(\mathbf{r}', t)] d\mathbf{r}' \\ & + \int_{M'} s(\mathbf{r}, \mathbf{q}) I(\mathbf{q}, t) d\mathbf{q} + h \end{aligned} \quad (2.1)$$

où τ est le taux de relaxation du réseau et h une constante négative. L’influence des connexions latérales est décrite par la première intégrale, où f est une fonction d’activation qui permet d’obtenir le taux de décharge en fonction du potentiel d’une position spatiale donnée, w_M décrit le poids des connexions latérales en fonction de la distance $|\mathbf{r} - \mathbf{r}'|$. Les stimuli externes sont pris en compte dans le second terme intégral, avec $I(\mathbf{q}, t)$

la stimulation à la position spatiale $\mathbf{q} \in M'$ et s la fonction de poids de connexion. L'activité du réseau en un point peut être interprétée comme reflétant le taux de décharge. Ce type de réseau permet de modéliser les dynamiques plus ou moins complexes qui ont lieu au sein d'une population de neurones, comme les compétitions locales dues aux interactions latérales qu'il est possible d'observer dans les aires corticales [Stemmler *et al.*, 1995; Sirosh *et al.*, 1996].

Les RNA traditionnels ont cependant montré leurs limites, en particulier pour expliquer certaines observations expérimentales sur la vitesse de traitement [Thorpe et Imbert, 1989], sur la synchronisation [Gray et Singer, 1989] ou sur la plasticité [Edeline, 1996] des neurones biologiques. Le codage par taux de décharge ne permet pas de capturer certains phénomènes qui requièrent une discrimination temporelle plus précise. Ces phénomènes nécessitant une précision temporelle accrue, comme la synchronie, peuvent apporter des solutions à plusieurs questions classiques dans le domaine de la perception [von der Malsburg, 1981; von der Malsburg, 1995; von der Malsburg, 1999].

2.2 Les modèles de neurones impulsionnels

2.2.1 Caractéristiques des modèles de neurones impulsionnels

Une des caractéristiques essentielles des neurones impulsionnels est qu'ils capturent plus finement que les neurones à taux de décharge le traitement de l'information réalisé par les neurones biologiques. En effet, les neurones biologiques produisent une sortie discrète à partir d'une entrée continue. Bien que l'importance de cette caractéristique ait été comprise depuis longtemps [Abu-Mostafa, 1989; Abu-Mostafa et Schweizer, 1990], les modèles de neurones utilisés dans les RNA ne capturaient pas correctement cet aspect. Les modèles explicatifs du niveau microscopique, comme celui bien connu de Hodgkin-Huxley [Hodgkin *et al.*, 1952; Hodgkin et Huxley, 1952b; Hodgkin et Huxley, 1952d], reproduisent de façon extrêmement précise les mesures de voltage enregistrées³, prenant en compte les différents courants membranaires. Cependant, le neurone de Hodgkin-Huxley est un modèle régi par 4 équations différentielles couplées et il est beaucoup trop complexe et coûteux en calculs pour être mis en œuvre dans un réseau de grande taille.

3. Hodgkin, Huxley et Katz ont pu proposer leur modèle car ils ont travaillé sur des neurones géants du calmar. Les neurones et les axones étaient suffisamment gros pour permettre les enregistrements avec les électrodes disponibles à l'époque. Il a fallu attendre une dizaine d'années après leur travaux pour pouvoir disposer des outils nécessaires à l'enregistrement des neurones de vertébrés.

Les modèles impulsionnels, au niveau microscopique, ne sont pas explicatifs : le but recherché avec ces modèles n'est en aucun cas d'expliquer les mécanismes électrochimiques qui conduisent les neurones à décharger. Le but recherché est de décrire le plus précisément possible les trains de potentiels d'action émis en fonction des stimulus. Les potentiels d'action doivent évidemment être décrits avec leurs étiquettes temporelles, correspondant à leurs instants d'émission, c'est pourquoi ces modèles sont communément appelés *spiking neurons*.

Les modèles impulsionnels sont caractérisés par les propriétés suivantes :

- une prise en compte explicite du temps ;
- une règle de décision spécifiant quand le neurone émet un potentiel d'action ;
- une explication des changements déclenchés par l'émission d'un potentiel d'action (généralement le retour du potentiel à un niveau proche du potentiel de repos).

2.2.2 Le modèle d'intégrateur à fuite

Un des modèles les plus simples vérifiant ces conditions est le modèle « intègre-et-décharge à fuite » [Abbott, 1999; Gerstner et Kistler, 2002], aussi appelé intégrateur à fuite ou *leaky integrate-and-fire* (LIF). La figure 2.2 montre le circuit électrique qui est utilisé pour modéliser la membrane d'un neurone intègre-et-décharge⁴. Il est possible de décrire les variations du potentiel de membrane à l'aide de l'équation différentielle suivante :

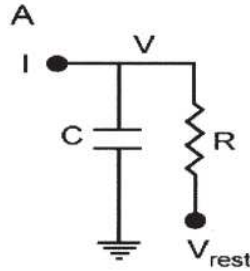


FIGURE 2.2 Circuit électrique permettant de modéliser les variations du potentiel de membrane V d'un neurone intègre-et-décharge soumis à un courant I , où C est la capacité membranaire, R la résistance membranaire et V_{rest} le potentiel de repos. Extrait de [Abbott, 1999].

$$\begin{cases} C \frac{dV}{dt} = g_l(V - E_l) + I(t), \text{ si } V \leq \vartheta \\ \text{sinon émission d'un potentiel d'action et } V \text{ revient à } V_{reset} \end{cases} \quad (2.2)$$

4. Ce circuit décrit l'évolution du potentiel de membrane comme étant équivalent à $\tau \frac{dV}{dt} = -V + RI$, avec $\tau = RC$.

dans laquelle V est le potentiel de membrane, C est la capacité de la membrane, g_l est la conductance de fuite (ou *leak* en anglais), E_l est le potentiel de repos et $I(t)$ est le courant injecté dans la membrane à l'instant t . Cette équation décrit les variations subliminaires du potentiel de membrane, c'est-à-dire tant que le neurone n'émet pas de potentiel d'action. Ces variations sont déterminées par deux termes. Le premier terme décrit le courant de fuite, qui tend à ramener le potentiel du neurone à sa valeur de repos, avec une « force » inversement proportionnelle à la différence entre la valeur courante du potentiel et le potentiel de repos. Le second terme décrit le courant injecté dans le neurone : si le potentiel de membrane dépasse un seuil fixé ϑ , alors le neurone émet un potentiel d'action et son potentiel est réinitialisé à une valeur de potentiel arbitraire V_{reset} , généralement proche de E_l .

2.2.3 Les variantes des modèles intègre-et-décharge

Il existe beaucoup de variantes de ces modèles intègre-et-décharge, ces variantes tendent à modéliser plus exactement les trains de potentiels d'action générés en fonction des entrées. Parmi ces variantes, les intègre-et-décharge quadratiques ([Gerstner et Kistler, 2002, chapitre 4.1.2] et [Ermentrout et Kopell, 1986; Ermentrout, 1996]) sont beaucoup plus fiables pour reproduire la dynamique d'un réseau de neurones de type Hodgkin-Huxley au niveau mésoscopique [Latham *et al.*, 2000; Feng, 2001].

$$\begin{cases} C \frac{dV}{dt} = g_l(V - E_l)(V - \vartheta) + I(t), \text{ si } V \leq \vartheta \\ \text{sinon émission d'un potentiel d'action et } V \text{ revient à } V_{\text{reset}} \end{cases} \quad (2.3)$$

Pour le niveau microscopique, dans [Fourcaud-Trocmé, 2003; Fourcaud-Trocmé *et al.*, 2003], les auteurs proposent un modèle d'intègre-et-décharge exponentiel qui prédit bien les temps d'émission de potentiel d'action quand le neurone est soumis à des variations rapides de ses entrées :

$$\begin{cases} C \frac{dV}{dt} = g_l(V - E_l)e^{(V - E_l)} + I(t), \text{ si } V \leq \vartheta \\ \text{sinon émission d'un potentiel d'action et } V \text{ revient à } V_{\text{reset}} \end{cases} \quad (2.4)$$

Un autre modèle, le neurone d'Izhikevich, propose une approche alternative. En effet, son auteur est parti d'observations biologiques des motifs d'émission de potentiels d'action d'un neurone. Après les avoir séparés en 20 catégories différentes (voir la figure 2.3), il propose un modèle impulsif [Izhikevich, 2003] qui permet de reproduire tous ces motifs différents [Izhikevich, 2004]. Ce modèle décrit l'évolution du potentiel

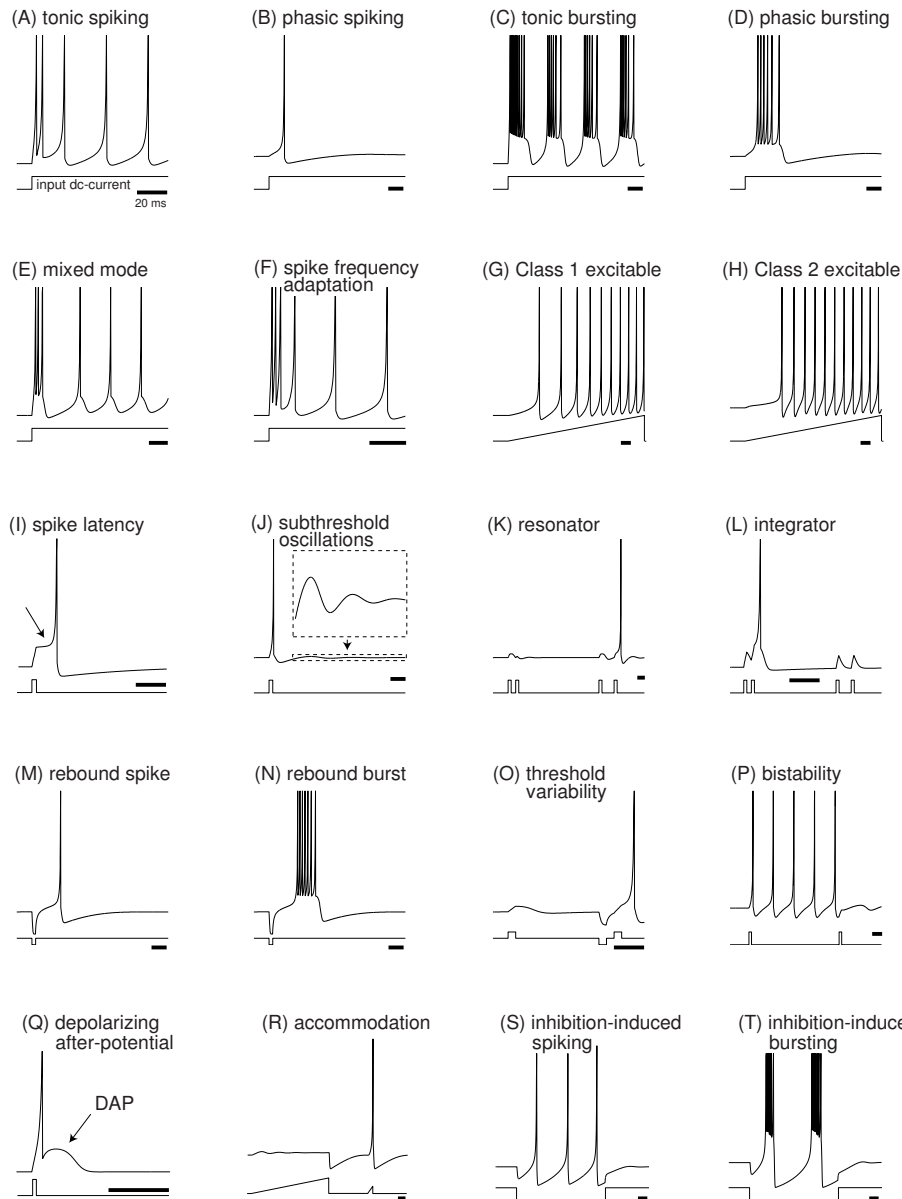


FIGURE 2.3 Classification des différents types de motif de décharge des neurones corticaux. Les barres horizontales représentent un intervalle de 20 ms. [Source : [Izhikevich, 2004] – www.izhikevich.com]

de membrane v par un système de deux équations différentielles couplées :

$$\begin{cases} v' = 0.04v^2 + 5v + 140 - u + I \\ u' = a(bv - u) \end{cases} \quad (2.5)$$

et si v dépasse le seuil alors

$$\begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \quad (2.6)$$

où u modélise la période réfractaire qui suit l'émission d'un potentiel d'action. En faisant varier les valeurs des quatre paramètres a , b , c et d , il est possible d'obtenir les différents modèles de décharge présentés sur la figure 2.3.

Ces modèles, qui sont des généralisations du neurone intègre-et-décharge [Gerstner et Kistler, 2002; Fourcaud-Trocmé *et al.*, 2003], permettent de modéliser plus précisément les trains de potentiels d'action en sortie d'un neurone, mais au prix d'une complexité plus importante ce qui induit un surcoût computationnel non négligeable lors des simulations.

2.2.4 Le modèle SRM

Un autre type de modèle, le *Spike Response Model* (SRM), permet de modéliser précisément les niveaux micro- et mésoscopique [Gerstner et van Hemmen, 1992; Gerstner et Kistler, 2002]. C'est un modèle phénoménologique, en ce sens qu'il permet de décrire différents phénomènes avec un formalisme unifié. En particulier, il n'impose pas de description spécifique des entrées synaptiques ou de l'émission de potentiels d'action.

Comme les intègre-et-décharge quadratique et exponentiel, ce modèle est une généralisation de l'intégrateur à fuite mais avec quelques différences. Tout d'abord le SRM ne dépend pas du voltage mais du temps d'émission du dernier potentiel d'action. Ceci permet de formuler les équations d'une façon différente, c'est-à-dire qu'au lieu d'utiliser une ou plusieurs équations différentielles, le modèle SRM exprime le potentiel de membrane sous une forme événementielle. Si un neurone i a émis un potentiel d'action à l'instant \hat{t}_i , l'évolution de son potentiel de membrane V_i peut s'écrire sous la forme :

$$V_i(t) = \eta(t - \hat{t}_i) + \sum_j w_{ij} \sum_f \epsilon_{ij}(t - \hat{t}_i, t - t_j^f) + \int_0^\infty \kappa(t - \hat{t}_i, s) I^{ext}(t - s) ds \quad (2.7)$$

la fonction η décrit la forme du potentiel d'action quand V_i dépasse le seuil ϑ , c'est-à-dire l'évolution du potentiel de membrane entre l'instant où il est supérieur à ϑ et le

moment où il est revenu à V_{reset} . Le second terme décrit les perturbations induites par les potentiels d'action entrants, avec t_j^f les instants d'émission des potentiels d'action d'un neurone présynaptique j et w_{ij} le poids synaptique de j vers i . Le dernier terme décrit l'influence d'un courant entrant I^{ext} .

Les noyaux η , ϵ et κ représentent la réponse du potentiel V_i respectivement à l'émission d'un potentiel d'action, aux stimulations en provenance des synapses et à celle des courants d'entrée. Ce formalisme permet de prendre aisément en compte des réponses plus ou moins complexes aux perturbations qui viennent modifier le potentiel de membrane. Ces noyaux, qui décrivent les réponses du modèle, sont à l'origine du nom *Spike Response Model*.

Le modèle SRM utilise un seuil dynamique pour décrire la période réfractaire. Le seuil est ainsi exprimé comme une fonction du temps écoulé depuis le dernier potentiel d'action émis : $\vartheta(t - \hat{t}_i)$. La période réfractaire absolue Δ_{abs} est décrite en fixant une valeur de ϑ inatteignable. Le modèle SRM permet de prendre en compte deux aspects de la période réfractaire relative, qui sont illustrés sur la figure 2.4. D'une part le neurone doit être soumis à un plus fort courant pour décharger, pour ceci le seuil suit une décroissance exponentielle entre la valeur du seuil à δ_{abs} et sa valeur de repos. D'autre part, les différents noyaux de réponse sont fonction de \hat{t}_i et peuvent donc simuler le fait que le neurone est plus ou moins sensible aux stimulations.

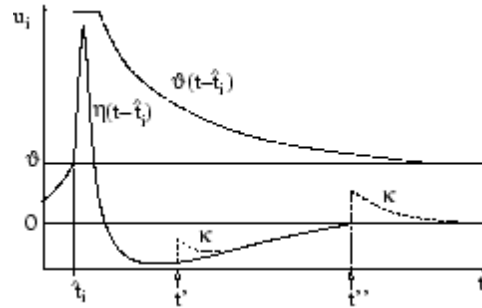


FIGURE 2.4 Illustration du déroulement temporel du potentiel de membrane, ici u_i , dans le modèle SRM. Le neurone i a émis un potentiel d'action au temps \hat{t}_i , le noyau $\eta(t - \hat{t}_i)$ décrit la réponse du potentiel de membrane à l'émission de ce potentiel d'action. À t' , le neurone reçoit un courant d'entrée qui provoque une dépolarisation faible car le neurone est encore dans sa période réfractaire, prise en compte par le noyau κ . À t'' , la période réfractaire est terminée et un courant d'entrée provoque une dépolarisation normale. Le seuil dynamique $\vartheta(t - \hat{t}_i)$ est représenté sur cette figure. [Source : [Gerstner et Kistler, 2002] – Spiking Neuron Models]

Ce modèle permet donc de décrire différents types de neurones avec un formalisme unifié, mais il n'est pas toujours nécessaire de détailler autant les différents noyaux de réponse, en particulier dans les cas où une implémentation peu coûteuse en temps de

calcul est recherchée. Il existe donc une version simplifiée de ce modèle [Gerstner et van Hemmen, 1992; Gerstner, 2000; Gerstner et Kistler, 2002] qui ne prend en compte que les temps de décharge des neurones présynaptiques :

$$V_i(t) = \eta(t - \hat{t}_i) + \sum_j w_{ij} \sum_f \epsilon_0(t - t_j^f) + \int_0^\infty \kappa_0(s) I^{ext}(t - s) ds \quad (2.8)$$

Ce modèle est appelé SRM₀ et la figure 2.5 illustre le type de réponse donnée par ce modèle quand il est soumis à différents courants d'entrée. Le choix de noyaux ϵ et κ simples permet d'implémenter des neurones impulsionnels en minimisant les temps de calculs nécessaires pour mettre à jour le potentiel de membrane. La partie suivante détaille différentes possibilités pour décrire l'influence des courants synaptiques et donc des fonctions utilisables pour décrire le noyau ϵ .

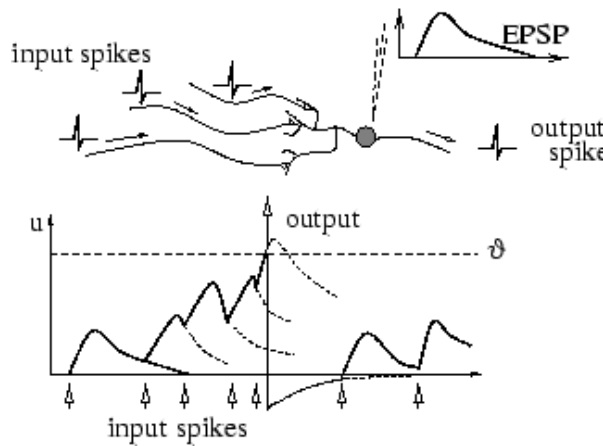


FIGURE 2.5 Illustration du modèle SRM₀. Le neurone reçoit plusieurs potentiels postsynaptiques excitateurs, aux instants indiqués par des flèches, et tous ces potentiels postsynaptiques sont ajoutés au potentiel de membrane, ici u_i . Une fois que le potentiel atteint le seuil, le neurone décharge et l'ajout d'un courant négatif η réinitialise le potentiel de membrane. [Source : [Gerstner et Kistler, 2002]] – Spiking Neuron Models

2.3 Les modèles de synapses

Les modèles de neurones impulsionnels permettent de décrire les variations du potentiel de membrane. Il existe cependant un autre paramètre important dans le choix d'un modèle impulsif, celui de la modélisation des courants synaptiques. La synapse constitue en fait un système dynamique d'une complexité équivalente à celle du neurone [Tsodyks *et al.*, 1998]. Les courants synaptiques constituent la quasi-totalité des entrées d'un neurone et sont le plus souvent la cause de l'émission de potentiels d'action.

2.3.1 Les courants synaptiques

Comme nous l'avons vu dans la partie 1.1.6, les synapses biologiques permettent à un potentiel d'action émis par le neurone présynaptique de causer au niveau du neurone postsynaptique un changement de potentiel qui va se propager jusqu'au corps cellulaire : le potentiel postsynaptique (PPS). Selon la nature de la synapse, ce potentiel peut être excitateur (PPSE) ou inhibiteur (PPSI).

Le modèle d'interaction synaptique le plus simple consiste à considérer que le PPS induit un changement ponctuel instantané sur le potentiel de membrane du neurone postsynaptique i . L'amplitude de ce saut de potentiel est fonction du poids synaptique w_{ij} . Ainsi, pour un PPS reçu à t_j^f , le potentiel V_i est mis à jour avec :

$$V_i(t_j^f) \leftarrow V_i(t_j^f) + w_{ij} \quad (2.9)$$

Il est cependant possible d'utiliser des modèles biologiquement plus plausibles pour décrire l'influence d'un PPS sur le potentiel de membrane. Un modèle largement utilisé est appelé fonction α [Rall, 1967] et décrit la transformation réalisée par la synapse, qui va convertir une entrée discrète – le potentiel d'action présynaptique – en sortie continue – le PPS. En un certain sens ce modèle réalise l'opération inverse du neurone, qui convertit une entrée continue en sortie discrète. Cette fonction α , pour exprimer l'influence d'un potentiel d'action sur la synapse j , s'écrit le plus souvent sous la forme :

$$PPS_j(t) = \frac{t}{\tau_{syn}} e^{-\frac{t}{\tau_{syn}}} \quad (2.10)$$

Le courant PPS_j peut par exemple être utilisé avec l'équation 2.2 de l'intégrateur à fuite pour donner :

$$C \frac{dV_i}{dt} = g_l(V_i - E_l) + \sum_j w_{ij} \sum_f PPS_j(t - t_j^f) \quad (2.11)$$

Une autre façon de modéliser les courants synaptiques consiste à décrire les mouvements ioniques en termes de conductances, hypothèse que nous avons vue dans la partie 1.1.5. Les modèles intègre-et-décharge qui utilisent ce formalisme sont couramment appelés modèles à conductance, ou *conductance-based models* [Destexhe et al., 1994; Troyer et Miller, 1997; Rudolph et Destexhe, 2006]. Il est ainsi possible de modéliser

une conductance synaptique exponentielle avec les équations suivantes :

$$\begin{cases} \tau \frac{dV}{dt} = -(V - E_l) - \sum_{j,f} g(t - t_j^f)(V - E_{syn}) \\ \tau_s \frac{dg}{dt} = -g \\ g \leftarrow g + w_j, \text{ quand } t = t_j^f \end{cases} \quad (2.12)$$

avec E_{syn} le potentiel d'équilibre de la synapse. Quand un potentiel d'action arrive sur la synapse j au temps t_j^f , il déclenche un changement instantané de la conductance synaptique $g \leftarrow g + w_j$.

Le principal problème de la modélisation des courants synaptiques concerne le coût en termes de temps de calcul. En particulier, dès que l'on cherche à implémenter plusieurs types de synapses, qui diffèrent par leur constantes E_{syn} ou τ_s , il est nécessaire d'utiliser des méthodes d'approximation coûteuses, du type Newton ou Runge-Kutta, pour intégrer les équations différentielles. Même s'il est possible de faire appel à ces méthodes d'intégration seulement dans les cas où un neurone est susceptible de décharger [Brette, 2006], la modélisation des conductances synaptiques est encore difficilement compatible avec des contraintes sur le temps de traitement par une machine, surtout à l'échelle d'un réseau de neurones de grande taille.

2.3.2 La plasticité synaptique

Les mécanismes de modification des poids synaptiques sont de plus en plus largement utilisés. Ces modifications sont de deux types, selon qu'elles concernent la modélisation de la fatigue synaptique ou bien l'adaptation du poids synaptique en fonction de la différence temporelle entre les temps de décharge des neurones pré- et postsynaptique.

La fatigue synaptique permet de rendre compte de la disponibilité des vésicules synaptiques au moment où un potentiel d'action atteint la synapse. En effet si un neurone émet plusieurs potentiels d'action en peu de temps, ses synapses peuvent venir à manquer de vésicules de neurotransmetteurs et donc propager des PPS d'une intensité plus faible. [Markram *et al.*, 1998; Perrinet *et al.*, 2001; Perrinet, 2003].

Le second type de modification des poids est connu sous le nom de plasticité synaptique et constitue une réinterprétation de la règle de Hebb [Hebb, 1949]. Les idées proposées par Hebb prennent en particulier un sens nouveau dès lors que l'on considère un codage temporel. Ainsi, si un neurone présynaptique participe de façon répétée ou persistante à faire décharger un neurone postsynaptique, Hebb postule que l'efficacité de la synapse devrait augmenter. Le lien qu'il établit entre croissance synaptique et causalité s'adapte particulièrement bien aux neurones impulsionnels et cette règle est

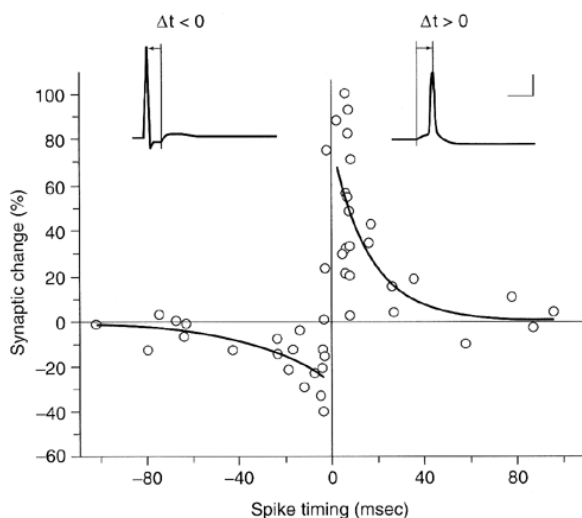


FIGURE 2.6 Fenêtre de STDP pour la modification des poids synaptiques. La différence $\Delta t = t_{pre} - t_{post}$ indique l'écart temporel qui sépare l'émission d'un potentiel d'action entre les neurones pré- et postsynaptique. Si $\Delta t < 0$ la valeur du poids synaptique est dépréciée et si $\Delta t > 0$ le poids est renforcé, car le neurone présynaptique a participé à faire décharger le neurone postsynaptique. [Source : [Bi et Poo, 1998]]

couramment appelée règle de Hebb temporelle.

Les mécanismes de modification de poids synaptiques qui utilisent cette règle hebbienne temporelle peuvent prendre des formes différentes [Abbott et Nelson, 2000] : *synaptic scaling*, *synaptic redistribution* [Markram et Tsodyks, 1996] ou *Spike Timing-Dependent Plasticity* (STDP) [Song et al., 2000; Song et Abbott, 2001]. Ce dernier mécanisme, la STDP, permet de réaliser localement une adaptation des poids en fonction de la coïncidence de l'émission des potentiels d'action entre deux neurones [Bi, 2002; Dan et Poo, 2004]. Les règles de STDP utilisent la différence temporelle $\delta t = t_{pre} - t_{post}$, avec t_{pre} l'instant où le neurone présynaptique émet un potentiel d'action et t_{post} l'instant d'émission du neurone postsynaptique, et une fenêtre temporelle [Song et Abbott, 2001; Meunier et Paugam-Moisy, 2005] (voir figure 2.6) pour expliciter la modification du poids de la synapse. La STDP est largement utilisée pour réaliser des apprentissages avec des réseaux de neurones impulsionnels ; voir [Paugam-Moisy, 2006] pour une revue des problématiques d'apprentissage en lien avec les neurones impulsionnels.

2.4 Synthèse

Dans cette thèse nous avons choisi un point de vue intermédiaire, où nous nous sommes inspirés des connaissances sur le fonctionnement des réseaux biologiques sans pour autant chercher à produire des modèles explicatifs pour les neurosciences, c'est-à-dire que nous n'avons pas essayé de reproduire exactement les observations biologiques.

Les modèles de neurones impulsionnels décrivent l'évolution du potentiel membranaire en fonction des entrées et ils permettent ainsi d'explicitier les temps précis d'émission des potentiels d'action. Il est ainsi possible, en utilisant ces modèles, de recoder les informations sous une forme temporelle et discrète. Les neurones impulsionnels peuvent capturer des dynamiques temporelles fines et ne sont pas aussi coûteux en terme computationnel que les modèles qui tentent de reproduire le plus fidèlement possible les observations biologiques. Ils représentent en cela un compromis intéressant en capturant une partie des propriétés des neurones biologiques, ce qui permet de reproduire des dynamiques riches et variées, tout en proposant des modèles relativement peu complexes, qui peuvent être implémentés pour simuler des réseaux de grande taille.

Chapitre 3

Les implémentations de mécanismes d'attention visuelle

Sommaire

3.1	Quels intérêts pour la vision artificielle ?	54
3.2	Les modèles attentionnels	55
3.2.1	Les modèles neuropsychologiques	56
3.2.2	Les systèmes de traitement d'images	57
3.2.3	Les approches bio-inspirées	59
3.3	Une architecture de la vision préattentive avec des neu-	
	rones impulsionnels	64
3.3.1	Intérêts d'une implémentation impulsionnelle	65
3.3.2	Un système <i>anytime</i>	68
3.3.3	Méthode d'implémentation du système	70

Nous avons vu dans le chapitre 1 les structures neuronales impliquées dans la perception visuelle et les différentes théories concernant l'attention et la préattention visuelle. Dans ce chapitre, nous allons détailler les différentes implémentations existantes de mécanismes d'attention visuelle avant de présenter ce qui constitue la problématique de cette thèse : que peuvent apporter les neurones impulsionnels à l'implémentation d'une architecture de la préattention ? Nous commencerons par une brève discussion sur les intérêts des processus attentionnels pour les systèmes biologiques et pour les systèmes de vision artificielle (partie 3.1), avant de voir des implémentations qui reproduisent les observations expérimentales ou qui s'inspirent du processus préattentionnel pour proposer des systèmes artificiels performants (partie 3.2). Nous terminerons ce chapitre en développant les intérêts que présentent les neurones impulsionnels pour implémenter une architecture de vision préattentionnelle (partie 3.3).

3.1 Quels intérêts pour la vision artificielle ?

Il apparaît expérimentalement que la focalisation attentionnelle est nécessaire à la perception, mais cela ne permet pas de savoir pourquoi le système visuel utilise un tel processus, qui traite séquentiellement différentes régions spatiales du champ visuel. Quels sont les intérêts de ce mécanisme de sélection attentionnelle et sont-ils pertinents pour les systèmes de vision artificielle ? Cette dernière question est abordée d'un point de vue formel dans [Tsotsos, 1989; Tsotsos, 1990]. L'auteur définit deux cas théoriques de recherche visuelle et présente les preuves de leur complexité algorithmique. Le premier cas est celui d'une recherche visuelle non contrainte (*unbounded visual search*), qui modélise une tâche de reconnaissance dans laquelle aucun indice ne permet d'optimiser la recherche. C'est un processus complètement ascendant, qui correspond au type de processus mis en jeu dans le système de vision artificielle décrit dans [Marr, 1982]. Le deuxième cas modélise une recherche visuelle contrainte (*bounded visual search*), dans laquelle il est possible d'optimiser la reconnaissance car on connaît d'avance la cible et la tâche. L'auteur démontre que la complexité du cas de recherche visuelle contrainte est linéaire en fonction du nombre de pixels [Tsotsos, 1989]. Dans le cas de la recherche visuelle non contrainte, c'est un problème NP-complet [Tsotsos, 1990; Tsotsos, 1991], dont la complexité temporelle est en $O(N2^{PM})$, avec N le nombre de prototypes, P la taille de l'image en pixels et M le nombre de caractéristiques calculées pour chaque pixel. Cette analyse lui permet de proposer différentes optimisations et approximations pour réduire cette complexité [Tsotsos, 1992; Tsotsos et al., 1995].

Cette analyse permet de passer outre la justification récurrente selon laquelle le cerveau ne serait pas assez puissant pour traiter toutes les informations en parallèle et que, étant donné ses capacités limitées, la sélection attentionnelle serait un « goulot d'étranglement » incontournable pour le traitement des informations visuelles. Ce qui peut sembler très curieux, surtout quand on considère le fait que la structure du cerveau est massivement parallèle. L'analyse formelle de Tsotsos permet de replacer dans une perspective évolutionniste l'apport de ce mécanisme de sélection de l'information. Le fait de sélectionner une partie de la scène visuelle avant de la traiter plus complètement permet la mise en place de différentes heuristiques pour trouver rapidement une information recherchée, ce qui est impossible dans le cas d'une chaîne de traitements complètement parallèle. Un traitement séquentiel munit le système d'un mécanisme d'exploration, appelé par Tsotsos *hypothesize-and-test* [Tsotsos, 1992], qui peut ainsi être l'objet du processus de sélection naturelle. Ceci apporte une certaine souplesse au système perceptif, qui serait inexistante si le processus était complètement parallèle, et qui lui permet de s'adapter au mieux à son environnement.

Il semble donc important de doter les systèmes de vision artificielle d'un processus de sélection attentionnelle, surtout si la finalité de ceux-ci est d'être utilisés dans un cadre de robotique autonome. Nous allons voir dans la partie suivante les différentes implémentations de l'attention dans des systèmes artificiels.

3.2 Les modèles attentionnels

Parmi les théories de l'attention visuelle présentées dans la partie 1.3.3, certaines combinent les traits caractéristiques sur une carte de saillances [Koch et Ullman, 1985], appelée *map of locations* par [Treisman et Souther, 1985] et *activation map* par [Wolfe, 1994]. Bien que la notion de saillance soit largement répandue et utilisée [Itti, 2007], il n'existe pas de consensus clair pour sa définition : il est largement admis que c'est le résultat des activations générées par la combinaison de traits caractéristiques et que les saillances indiquent les régions vers lesquelles l'attention pourrait potentiellement se déplacer. Elles sont généralement considérées comme étant le reflet de contrastes locaux et comme étant le résultat d'un processus ascendant n'impliquant pas d'influences descendantes.

Au niveau biologique, l'existence et la localisation de cette carte de saillances sont des questions encore largement débattues. Différentes structures cérébrales ont été proposées comme pouvant potentiellement accueillir cette carte : le CGL [Koch et Ullman, 1985], le pulvinar [Robinson et Petersen, 1992; Shipp, 2004], le colliculus supérieur [Kustov et Robinson, 1996], l'aire visuelle primaire V1 [Koch et Ullman, 1985; Li, 2002], V4 [Mazer et Gallant, 2003] ou encore LIP¹ [Deco *et al.*, 2002; Gottlieb, 2007]. L'idée qu'une carte de saillances soit observable dans une aire donnée du SNC, comme c'est possible pour d'autres cartes topiques dans le cerveau², n'est pas soutenue par des preuves expérimentales claires. Il est tout à fait possible qu'elle soit fonctionnellement distribuée sur plusieurs aires cérébrales [Desimone et Duncan, 1995], comme le réseau oculomoteur³ par exemple [Fecteau et Munoz, 2006], ou bien même qu'il existe plusieurs cartes de saillances [Niebur, 2007].

Les implémentations de modèles reposant sur les saillances sont nombreuses dans la littérature scientifique. Il est toutefois possible de les regrouper en trois grandes catégories : les modèles neuropsychologiques que nous verrons dans la partie suivante,

1. *Lateral Intraparietal Area* (singe) correspond à l'aire *Parietal Eye Field* chez l'humain [Shipp, 2004]

2. Comme par exemple l'organisation rétinotopique de l'aire visuelle primaire V1 ou tonotopique (organisation spatiale reflétant la sensibilité des neurones à différentes fréquences) de la cochlée et de l'olive supérieure latérale.

3. LIP, *Frontal Eye Field*, colliculus supérieur, pulvinar et formation réticulée

les systèmes de traitement d'images qui seront présentés dans la partie 3.2.2 et les approches bio-inspirées, présentées dans la partie 3.2.3.

3.2.1 Les modèles neuropsychologiques

Les implémentations de modèles psychologiques sont élaborées dans le cadre d'une démarche explicative. Le but premier de ces implémentations est de valider la cohérence du modèle et de vérifier que le modèle proposé, par le biais de son implémentation, permet de retrouver des résultats similaires à ceux observés dans des conditions expérimentales. À un autre niveau, ces implémentations offrent une opportunité pour mieux comprendre certains troubles neurologiques ou les conséquences de certaines lésions cérébrales.

Les buts de ces implémentations sont relativement éloignés de ceux qui sont développés dans cette thèse, en particulier pour les contraintes de temps de traitements et d'adéquation avec un système de vision artificielle. Il existe plusieurs modèles purement qualitatifs, visant à reproduire les comportements observés expérimentalement, comme le *Selective Attention Model* [Phaf *et al.*, 1990], le modèle *Guided-Search* [Wolfe, 1994], le modèle *Search via Recursive Rejection* [Humphreys et Müller, 1993], le *FeatureGate model* [Cave, 1999], le modèle *Multiple Object Recognition and Attentional Selection* [Mozer, 1991; Mozer et Sitton, 1998] ou le *Selective Attention for Identification Model* (SAIM) [Heinke et Humphreys, 2003]. Des explications détaillées et une comparaison de ces différentes implémentations sont présentées dans [Heinke et Humphreys, 2005].

Parmi ces implémentations de modèles neuropsychologiques, certaines emploient une approche neuronale pour proposer une explication du phénomène attentionnel à un niveau mésoscopique et macroscopique (voir 2.1.2). Nous pourrions citer en particulier le modèle neuronal de [Olshausen *et al.*, 1993], dans lequel un réseau de neurones sélectionne dynamiquement une région de l'espace visuel à transmettre aux aires corticales chargées de la reconnaissance d'objets. Ce modèle utilise des *control neurons* pour modifier dynamiquement l'efficacité synaptique des connexions intracorticales. C'est en quelque sorte une implémentation biologiquement plausible du faisceau attentionnel, proposant un mécanisme neuronal pour former des représentations invariantes, en position et en échelle, des objets du monde visuel.

Dans [Deco et Zihl, 2001; Deco *et al.*, 2002], les auteurs proposent un modèle neuronal macroscopique qui s'inspire du modèle *Guided-Search* [Wolfe, 1994]. Leur implémentation utilise une approximation du champ moyen [Wilson et Cowan, 1972; Amit et Tsodyks, 1991; Kepler *et al.*, 1992] pour décrire l'activité des cartes caractéristiques et de la carte de saillances. Leur modèle permet de retrouver des vitesses de

temps de traitements des stimuli en adéquation avec les observations expérimentales. Le modèle de [Taylor et Rogers, 2002] est très semblable et propose un modèle de contrôle du déplacement de l'attention pour expliquer les temps de réponse des sujets. Cependant ces modèles sont limités à une utilisation sur des scènes visuelles très simples, comportant peu de traits caractéristiques.

Les implémentations de modèles psychologiques aident à construire une théorie de la perception en offrant des explications aux observations expérimentales et suggèrent en retour différentes expériences à réaliser. Cependant, ils ne peuvent faire partie d'un système de perception artificielle et ne sont pas adaptés pour traiter des images naturelles. Nous allons voir dans la partie suivante des systèmes de traitement d'images utilisant des saillances, qui sont plus adéquats pour une utilisation sur des images naturelles.

3.2.2 Les systèmes de traitement d'images

Les régions saillantes peuvent être décrites comme les régions les plus informatives, qui sont porteuses de sens, dans une scène visuelle. Cette formulation a amené la communauté de traitement d'images à proposer plusieurs solutions pour extraire les régions saillantes d'une image. L'idée sous-jacente est que les régions saillantes peuvent fournir un levier utile pour réduire l'espace de recherche des algorithmes de reconnaissance d'objet dans une scène visuelle. Les systèmes de traitement d'images construisent ainsi une représentation de l'image qui peut prendre la forme d'un graphe ou d'une représentation spatiale de l'information portée par chaque pixel ou groupe de pixels. Une approche alternative consiste à utiliser une démarche bayésienne pour biaiser la recherche d'objets en fonction du contexte [Oliva et Torralba, 2001; Torralba et Oliva, 2003; Torralba, 2003b; Torralba, 2003a; Torralba, 2005]. Les scènes visuelles sont représentées avec des descripteurs globaux, ce qui permet, une fois le contexte reconnu, de prédire la probabilité de présence d'un objet dans la scène, ainsi que sa position et son échelle.

Les descripteurs d'images utilisés dans ces systèmes peuvent ressembler à certains traits caractéristiques, comme le *Harris corner detector* [Harris et Stephens, 1988; Harris, 1993] qui détecte les intersections ou les coins. Ces descripteurs sont utilisés pour des tâches de comparaison d'images mais sont limités à une utilisation sur des images relativement petites et lorsque les différences entre les images restent faibles. Dans [Zhang *et al.*, 1995], les auteurs proposent une méthode pour utiliser ces détecteurs de Harris sur des images plus grandes, en prenant en compte les informations au voisinage des points détectés.

Cependant, ces descripteurs s'appuient sur les propriétés géométriques locales des objets présents dans l'image : ils ne sont pas adaptés pour décrire des objets « non-

géométriques » (comme certaines textures par exemple) et ne peuvent discriminer différents objets s'ils sont trop proches. Il a été proposé par [Schmid et Mohr, 1997] d'utiliser des détecteurs géométriques pour sélectionner un ensemble de points dans l'image puis d'appliquer des traitements permettant d'extraire une signature (un invariant local) pour chaque point. Cette approche est tout à fait comparable à celle réalisée par les modèles neuropsychologiques décrits dans la partie précédente, en ce sens que les images sont d'abord réduites à une collection de régions saillantes qui vont ensuite guider la tâche à réaliser (comme la reconnaissance d'objets) en réduisant l'espace de recherche.

De nombreux descripteurs ont été introduits dans la littérature. Ils peuvent être répartis en trois catégories [Schmid *et al.*, 2000] selon qu'ils s'appuient sur des contours, des intensités (ou niveaux de gris) ou des modèles paramétriques. Les méthodes d'extraction de contours recherchent les points de courbure maximale, les points d'inflexion ou les points d'intersection de contours dans l'image. Le second type de méthode utilise directement l'intensité des pixels pour trouver des descripteurs, en mesurant les variations de niveaux de gris aux alentours de chaque point de l'image. La dernière méthode recherche des descripteurs en adaptant un modèle paramétrique de l'intensité au signal mesuré. Un état de l'art et une évaluation de ces descripteurs sont proposés dans [Schmid *et al.*, 2000] et [Mikolajczyk et Schmid, 2005]. Nous pourrions aussi citer les travaux de [Lowe, 1999; Lowe, 2004] qui proposent un descripteur, appelé *Scale Invariant Feature Transform* (SIFT) (voir figure 3.1), qui est de plus en plus largement utilisé.

Les modèles bio-inspirés que nous allons présenter dans la partie suivante ont plusieurs avantages par rapport aux descripteurs que nous venons de voir. Les systèmes bio-inspirés ne sont pas restreints à un seul trait caractéristique, comme par exemple la détection de coins ou de lignes car la détection de saillances repose sur l'extraction de plusieurs traits caractéristiques, ce qui les rend plus robustes et utilisables dans des environnements variés. De plus, la saillance étant définie comme un contraste local [Koch et Ullman, 1985; Itti *et al.*, 1998], ceci permet de fusionner les différentes modalités visuelles – c'est-à-dire les traits caractéristiques – d'une façon intéressante qui met en avant les traits caractéristiques les plus discriminants par rapport à leur voisinage. Le dernier avantage concerne le nombre de régions saillantes extraites par les systèmes bio-inspirés, qui se comptent généralement en dizaines, là où les systèmes de traitement d'images trouvent plusieurs centaines (voire milliers) de descripteurs pour des images naturelles.



FIGURE 3.1 Exemple d'utilisation des descripteurs SIFT [Lowe, 1999; Lowe, 2004] dans une tâche de reconnaissance d'objets. Les images utilisées pour l'apprentissage sont présentées à gauche, et les objets peuvent être reconnus dans une image naturelle, au centre, même s'ils sont partiellement cachés. Le résultat de la reconnaissance est présenté à droite. Un parallélogramme encadre chaque objet reconnu et indique les bords de l'image apprise, déformés par la transformation affine retenue par l'algorithme de reconnaissance. Les petits carrés indiquent les descripteurs utilisés pour la reconnaissance. [Source : [Lowe, 2004]]

3.2.3 Les approches bio-inspirées

Les approches bio-inspirées visent à proposer des solutions efficaces, aussi bien en termes de vitesse d'exécution que de qualité des résultats, en adaptant des mécanismes utilisés par les organismes biologiques. Le but d'une telle démarche est de trouver un compromis entre les modèles réalistes et les systèmes artificiels. En effet, les modèles biologiques ou neuropsychologiques réalistes tendent à reproduire le plus fidèlement possible les observations expérimentales. Les systèmes de traitement d'images proposent des solutions optimales et bien adaptées à des situations spécifiques. L'approche bio-inspirée consiste à rechercher un compromis, à abstraire les propriétés essentielles des systèmes biologiques pour proposer des systèmes robustes et capables de fonctionner dans la plupart des situations. Cette approche est très bien adaptée à la conception de systèmes robotiques autonomes, qui doivent faire face à des environnements variés et changeants, souvent inconnus.

Les recherches sur les modèles bio-inspirés en vision artificielle ont connu un tournant majeur lors de la publication des travaux de [Itti *et al.*, 1998]. Les auteurs proposent une architecture de la préattention, qui reprend un modèle initialement proposé par [Koch et Ullman, 1985], comme illustré sur la figure 3.2. Cette architecture implémente un modèle d'attention visuelle ascendant et *feedforward* organisé autour d'une carte de saillances qui permet de décrire les déplacements du focus d'attention. L'image d'entrée est tout d'abord décomposée selon différentes modalités visuelles (couleurs, orientations

et contours dans [Itti, 2000]) à plusieurs échelles spatiales. Pour les orientations par exemple, l'image est convoluée avec des ondelettes de Gabor de différentes orientations (0° , 45° , 90° et 135°) et ce pour 9 échelles spatiales, obtenues par construction d'une pyramide gaussienne [Burt et Adelson, 1983], pour des facteurs de réduction d'image allant de 1 : 1 (niveau 0, image originale) à 1 : 256 (niveau 8). Toutes les images résultantes sont combinées linéairement pour former ce qu'ils appellent une carte de traits caractéristiques (*feature map*). Les cartes de traits caractéristiques des orientations, des couleurs et des contours sont ensuite combinées à nouveau pour former une carte de saillances. Les contributions relatives de chaque carte de traits caractéristiques à la carte de saillances sont modulées pour former une carte de saillances qui permet de reproduire les saccades oculaires de sujets humains [Carmi et Itti, 2006]. La sélection attentionnelle est implémentée avec un algorithme de *winner-take-all*, qui permet de trouver la saillance la plus importante de la carte, ou focus d'attention [Itti et Koch, 2000]. Pour permettre à l'attention d'explorer la scène visuelle et donc au focus d'attention de se déplacer sur la carte de saillances, les auteurs utilisent un mécanisme d'inhibition de retour. Celui-ci inhibe la saillance la plus importante de la carte et une autre passe de l'algorithme de *winner-take-all* permet de choisir le focus d'attention suivant. Cette architecture reproduit au niveau fonctionnel le comportement des modèles préattentionnels que nous avons vus dans la partie 1.3.3 pour l'*overt attention* et la *covert attention* [Itti et Koch, 2000].

Les filtres utilisés dans [Itti, 2000] sont inspirés par les transformations spatiales réalisées par la rétine. Les auteurs utilisent des différences de gaussiennes pour réaliser un traitement analogue à celui des cellules ganglionnaires *on-off* de la rétine [Enroth-Cugell et Robson, 1966] (voir figure 1.8 dans la partie 1.2.2). Les informations locales d'orientations sont extraites à l'aide d'ondelettes de Gabor⁴, qui reflètent bien le filtrage réalisé par les champs récepteurs des cellules simples de l'aire visuelle primaire [Jones et Palmer, 1987; Jones *et al.*, 1987; Daugman et Downing, 1998] (voir la partie 1.2.4 et la figure 3.3 ci-dessous) et dont l'équation complexe dans le domaine spatial est :

$$\begin{aligned} \text{Gabor}(x, y) = & \exp \left[-\frac{1}{2} \left(\frac{(x - x_0)^2}{\sigma_x^2} + \frac{(y - y_0)^2}{\sigma_y^2} \right) \right] \times \\ & \exp [2\pi i \nu ((x - x_0) \cos \theta + (y - y_0) \sin \theta)] \end{aligned} \quad (3.1)$$

où ν représente la fréquence spatiale du filtre, θ son orientation et σ_x et σ_y sont des paramètres d'échelle définissant la taille de l'enveloppe gaussienne. Il est intéressant de noter que la décomposition d'une image en ondelettes de Gabor permet d'obtenir une

4. Ces ondelettes de Gabor sont définies comme le produit d'une sinusoïde complexe et d'une enveloppe gaussienne.

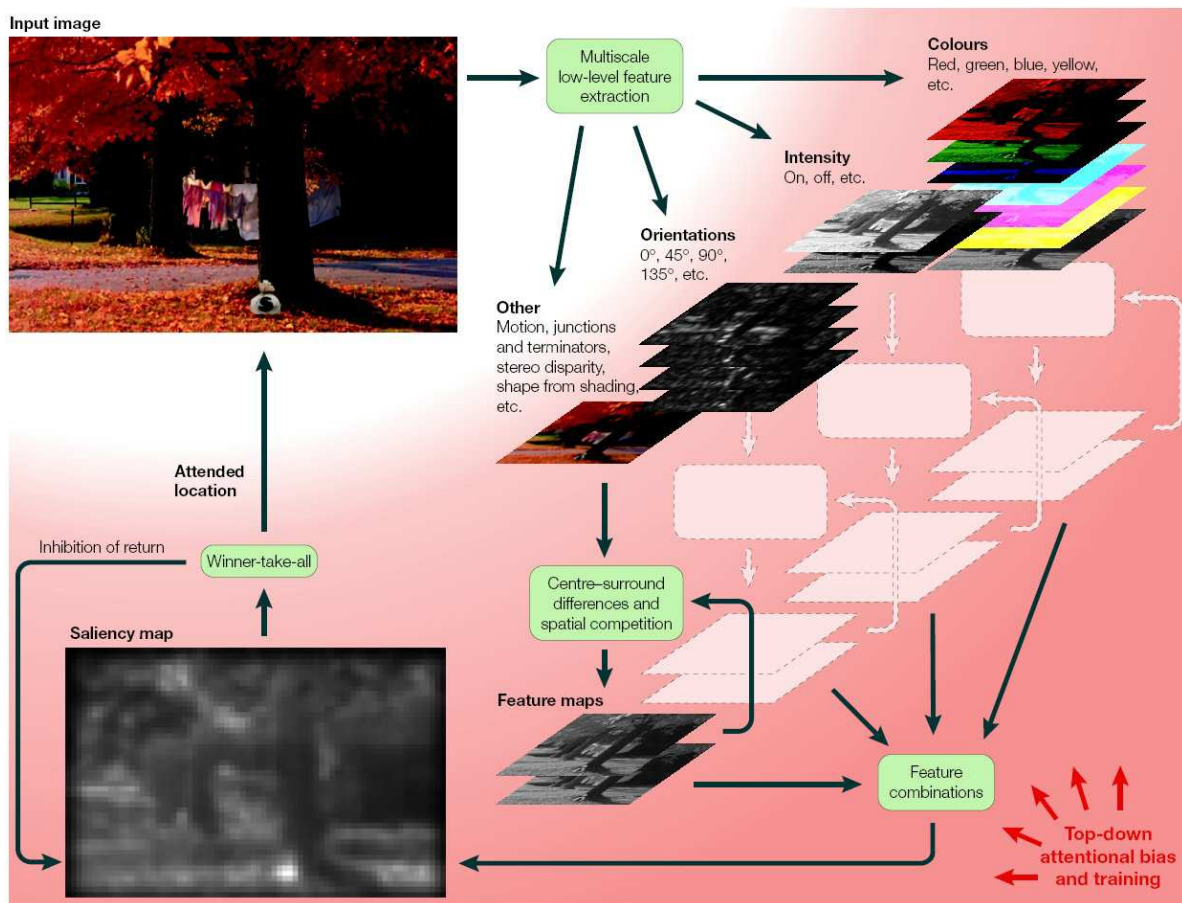


FIGURE 3.2 Architecture de vision artificielle bio-inspirée de [Itti *et al.*, 1998]. L'image d'entrée (en haut à gauche) est décomposée en cartes de traits caractéristiques à différentes échelles spatiales. L'architecture proposée traite différents types de traits caractéristiques : les oppositions de couleurs, les variations de contrastes et les orientations. L'ensemble des informations pour chaque type de traits caractéristiques est regroupé sur une carte (*feature map*). Par exemple pour les oppositions de couleurs, les oppositions rouge/vert et bleu/jaune sont calculées à différentes échelles spatiales puis regroupées sur une seule carte d'opposition de couleurs. Les informations des différentes *feature maps* sont ensuite combinées sur une carte de saillances. Un algorithme de *winner-take-all* sélectionne la région la plus saillante. Pour trouver la seconde région la plus saillante, la région la plus saillante est inhibée sur la carte de saillances avant de relancer le *winner-take-all*. [Source : [Itti et Koch, 2001]]

représentation stable et robuste des objets visuels, comme l'illustrent les travaux réalisés par [Krüger *et al.*, 2000; Krüger et Sommer, 2002a; Krüger et Sommer, 2002b].

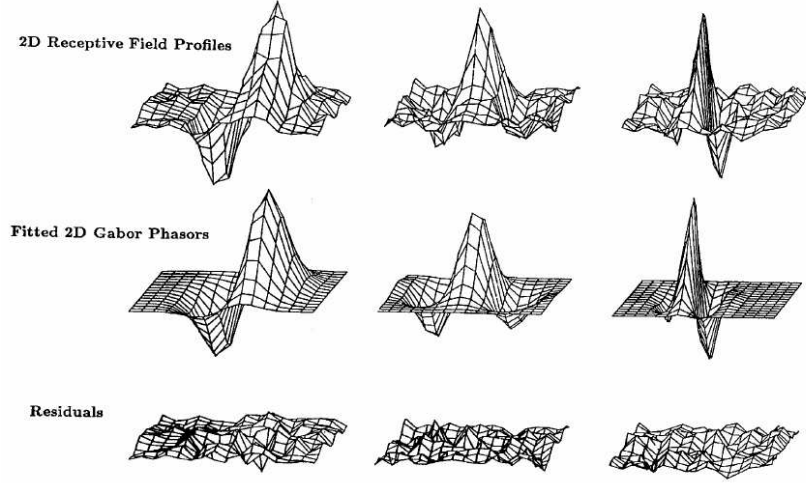


FIGURE 3.3 Comparaison du profil de réponse spatial des cellules simples du cortex visuel primaire (en haut) et du profil de réponse pour le même stimulus obtenu avec une ondelette de Gabor (au milieu) et des résidus non capturés par cette approximation (en bas). [Source : [Jones et Palmer, 1987]]

Un autre type de filtre a été proposé pour décrire les champs récepteurs des cellules simples de V1, il s'agit d'une dérivée première de la fonction gaussienne [Young, 1987] :

$$GD1(x, y, \theta) = -\frac{x \cos \theta + y \sin \theta}{\sigma_x} \cdot \left(\frac{1}{\sqrt{2\pi}\sigma_x} \exp \left[\frac{-(x \cos \theta + y \sin \theta)^2}{2\sigma_x^2} \right] \right) \times \left(\frac{1}{\sqrt{2\pi}\sigma_y} \exp \left[\frac{-(-x \sin \theta + y \cos \theta)^2}{2\sigma_y^2} \right] \right) \quad (3.2)$$

Les dérivées de gaussienne sont utilisées dans des modèles bio-inspirés, comme par exemple [Rao et Ballard, 1995; Rao *et al.*, 2002; Rao et Ballard, 2005], où les auteurs emploient des dérivées premières de gaussiennes orientées pour proposer un modèle d'*overt attention* dans une tâche de recherche visuelle. Dans ces travaux, les auteurs proposent que la saillance soit calculée comme la distance euclidienne entre un vecteur cible de traits caractéristiques (c'est-à-dire une signature à identifier) et des vecteurs de traits caractéristiques extraits à chaque position spatiale de l'entrée visuelle.

Les modèles bio-inspirés sont utilisés dans de nombreuses applications, qui incluent entre autres la détection automatique de cibles (comme des panneaux de signalisation le long des routes ou des véhicules militaires dissimulés dans la végétation [Itti et Koch, 2000]), l'assistance avancée à la conduite automobile (pour la détection d'obstacles ou de piétons) [Michalke *et al.*, 2007; Michalke *et al.*, 2008], la compression d'image et de vidéo pour dégrader sélectivement les zones non saillantes [Maeder *et al.*, 1996; Itti, 2004], le recalage automatique des images sur des écrans de petite taille [Ma et Zhang, 2003;

Le Meur *et al.*, 2006], le traitement d’images médicales [Hong et Brady, 2003; Fouquier *et al.*, 2008] et la robotique, qui utilise les saillances comme amers [Westin *et al.*, 1996; Sela et Levine, 1997; Baluja et Pomerleau, 1997; Frintrop *et al.*, 2006; Siagian et Itti, 2007]).

Bien que le cadre de cette thèse porte sur les modèles de préattention, il nous semble important de mentionner que plusieurs modèles bio-inspirés permettent de prendre en compte des influences descendantes, pour moduler les saillances en fonction d’une tâche en cours par exemple. Le modèle *Selective Tuning* de [Tsotsos *et al.*, 1995] est le premier à avoir été proposé, et des implémentations de ce modèle ont été réalisées pour des tâches de reconnaissance de mouvement [Tsotsos *et al.*, 2005]. Des extensions récentes de l’architecture de [Itti *et al.*, 1998] prennent en compte les influences descendantes pour guider l’attention : le modèle de [Navalpakkam et Itti, 2005] et le modèle VOCUS de [Frintrop, 2005]. Dans [Sun et Fisher, 2003; Sun, 2003], les auteurs proposent une architecture intéressante dans laquelle les saillances reflètent à la fois l’attention spatiale et l’attention centrée sur les objets⁵. Leur modèle est cependant très coûteux en temps de calcul et ne permet pas de résoudre tous les problèmes liés aux traitements de groupes d’objets.

Nous allons maintenant voir les différentes implémentations neuronales de modèles bio-inspirés. Dans [de Brecht et Saiki, 2006], les auteurs proposent une version du modèle de [Itti et Koch, 2000] dans laquelle les cartes de traits caractéristiques et la carte de saillances sont implémentées avec un modèle de neurones prenant en compte la fatigue synaptique. Chaque carte est composée de 64x64 populations de neurones dont l’activité est décrite par une approximation des champs moyens [Tsodyks *et al.*, 1998]. Ce modèle fonctionne bien sur des images artificielles, mais n’est pas adapté pour traiter des images naturelles. D’autres modèles utilisent la théorie des champs neuronaux continus (*Continuum Neural Field Theory*) [Wilson et Cowan, 1973; Feldman et Cowan, 1975; Amari, 1977; Taylor, 1999] pour proposer une implémentation neuronale de la carte de saillances et du mécanisme de *winner-take-all* [Ahrns et Neumann, 1999; Vitay *et al.*, 2005; Maillard, 2007; Fix, 2008]. Ces modèles tirent parti de la compétition locale qui a lieu au sein des champs neuronaux pour laisser émerger le focus d’attention. Ils utilisent des algorithmes de traitement d’images pour extraire des traits caractéristiques qui sont ensuite injectés dans le réseau.

Nous allons maintenant présenter la problématique de travail : proposer une ar-

5. Cette hypothèse postule que l’attention serait dirigée sélectivement vers les objets ou groupes d’objets et non vers une région spatiale donnée. Un état de l’art est proposé par [Scholl, 2001; Heinke et Humphreys, 2005]. Des observations expérimentales récentes tendent à prouver que ces deux types d’attentions ne seraient pas mutuellement exclusifs [Humphreys, 1998] et pourraient partager des mécanismes communs dans les aires pariétales du cerveau [Fink *et al.*, 1997]

chitecture bio-inspirée de la vision préattentionnelle implémentée avec des neurones impulsionnels.

3.3 Une architecture de la vision préattentive avec des neurones impulsionnels

Comme nous l'avons expliqué dans la partie 3.1, la focalisation attentionnelle et l'exploration séquentielle de scènes visuelles confèrent aux systèmes de vision artificielle un avantage, en ce sens qu'elles rendent possible une optimisation des traitements visuels, qui serait impossible avec des algorithmes traitant la scène visuelle dans son ensemble. Les théories de la préattention présentées dans la partie 1.3 définissent un cadre intéressant pour développer des implémentations efficaces de modèles attentionnels.

Nous avons choisi d'inscrire la problématique de cette thèse dans un cadre bio-inspiré, notion que nous avons définie au début de la partie précédente. Nous avons préféré une approche biologiquement inspirée à une approche de type traitement d'images, car cette dernière est plutôt orientée vers la recherche des meilleurs descripteurs d'une image, ceux qui permettent de la caractériser au mieux, de définir une signature unique pour chaque image. Les systèmes de traitement d'images semblent plus adaptés, par exemple, pour des tâches d'indexation ou de reconnaissance d'objets ; or nous souhaitons proposer une architecture qui puisse, à terme, être intégrée sur une plateforme robotique mettant en œuvre des principes bio-inspirés pour la vision [Beaudot, 1994; Benoit, 2007] et qui se positionne dans une approche "incarnée" de la perception [Pylyshyn, 2001; Gaussier *et al.*, 2003; Revel et Gaussier, 2004; Guillaume, 2009].

La conception d'une architecture qui extrait les saillances d'une image nécessite de définir la notion de saillance et comme nous l'avons constaté, il n'existe pas encore de consensus clair. Pour le cadre de ce travail, nous proposons la définition suivante des saillances : ce sont des contrastes locaux, globalement rares, qui portent des informations dans plusieurs modalités visuelles (contours, orientations et couleurs par exemple) et sur différentes échelles spatiales. Cette définition s'inspire des observations de [Masland, 2005], qui souligne la multimodalité des informations traitées par les aires visuelles corticales. Ainsi, cette définition met en avant le fait que les régions saillantes sont les plus susceptibles de contenir des informations.

La structure de l'architecture que nous allons détailler dans la suite de ce manuscrit s'inspire de celle de [Itti et Koch, 2000] (schéma (b) sur la figure 3.4) : une image d'entrée est décomposée selon différents traits caractéristiques qui sont ensuite combinés pour

former la carte de saillances. Un mécanisme de sélection détermine la saillance la plus importante sur cette carte pour positionner le focus d'attention.

Par rapport aux différentes implémentations préattentionnelles présentées dans la partie 3.2, nous proposons une prise en compte originale de l'influence des différentes échelles spatiales, qui s'inspire de celle des voies parvo- et magnocellulaires (voir partie 1.2.3). La séparation des traits caractéristiques réalisée dans les basses fréquences spatiales préactive la carte de saillances, ce qui va permettre de guider l'attention vers ces zones. Les saillances ainsi extraites reflètent les informations des hautes fréquences spatiales sur les régions préactivées par les basses fréquences, ce qui permet d'obtenir un "paysage" de saillances spatialement très précis.

L'implémentation de l'architecture a été conçue pour permettre la prise en compte de l'influence de mécanismes descendants sans que cela nécessite de modifications importantes. Il est ainsi possible d'ajouter des retours proprioceptifs ou des modulations *top-down* ; c'est pourquoi cette architecture se rapproche plutôt du modèle *Guided Search* de [Wolfe, 1994] (schéma (d) de la figure 3.4).

Nous pensons que les modèles de neurones impulsionnels, que nous avons vus dans le chapitre 2, apportent des propriétés intéressantes pour implémenter un modèle de préattention. L'objet de cette thèse est donc de proposer une approche nouvelle des systèmes préattentionnels qui tire parti de ces possibilités.

3.3.1 Intérêts d'une implémentation impulsionnelle

Les neurones impulsionnels opèrent un changement de représentation de l'information, du fait qu'ils émettent des potentiels d'action échelonnés dans le temps pour propager l'information. Ces potentiels d'action sont décrits informatiquement par des événements temporels fournissant une représentation unique de l'information. Ceci nous conduit à proposer une architecture modulaire et flexible, composée de cartes neuronales 2D, ce qui permet d'ajouter facilement de nouvelles fonctionnalités ou de prendre en compte de nouvelles modalités visuelles. De plus, la représentation des informations sous une forme unique, les potentiels d'action, permet de fusionner des informations de natures différentes sans avoir recours à des pré-traitements ou à des conversions. La fusion des différentes modalités visuelles sur la carte de saillances en est ainsi grandement facilitée.

Les neurones impulsionnels ouvrent des perspectives intéressantes pour les systèmes de perception artificielle, qui sont inhérentes à leur mode de calcul. Ils permettent de traiter les informations de façon sélective : seules les régions de l'image dont l'information est pertinente vont induire une dépolarisation suffisante pour permettre aux

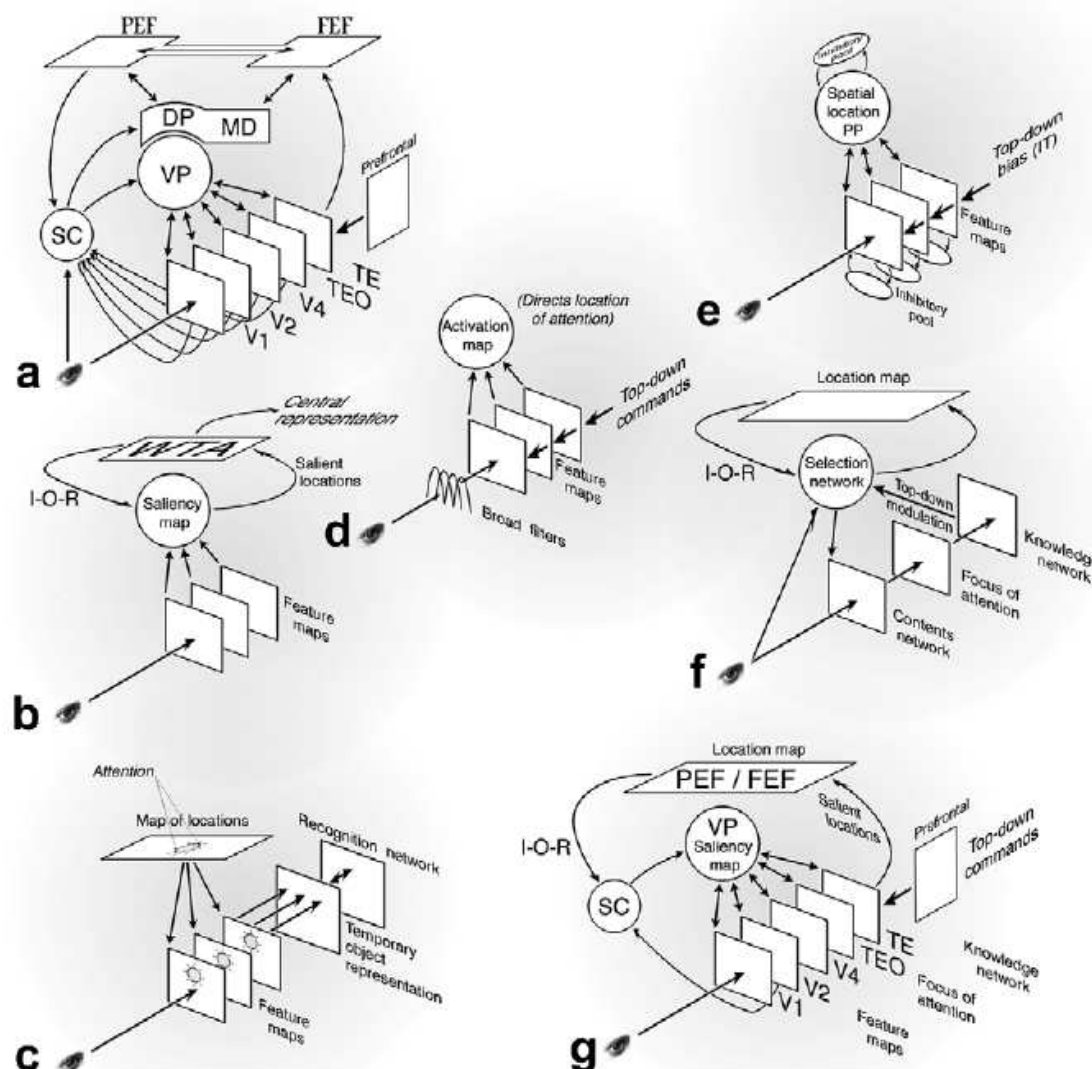


FIGURE 3.4 Schéma présentant les relations entre différents modèles attentionnels et les corrélats neuronaux de l'attention visuelle : (a) Les structures cérébrales impliquées dans l'attention [Shipp, 2004]. (b) Modèle de [Koch et Ullman, 1985], premier à avoir postulé l'existence d'une carte de saillances (voir 3.2.3). (c) *Feature integration theory* de [Treisman et Gelade, 1980; Treisman, 1998]. (d) *Guided search theory* [Wolfe, 1994]. (e) Modèle de [Deco et al., 2002] (voir 3.2.1). (f) Modèle SAIM [Heinke et Humphreys, 2003] (voir 3.2.1). (g) Modèle de l'attention de [Shipp, 2004], version réduite de (a) avec des indications fonctionnelles sur les liens entre les différentes structures cérébrales. [Source : [Shipp, 2004]]

neurones de décharger, les autres neurones resteront silencieux et n'engendreront pas de coût computationnel supplémentaire. Ainsi pendant l'étape de séparation des traits caractéristiques, les informations transmises sont déjà réduites et permettent un gain de temps en ne traitant que les informations pertinentes. Nous pouvons ainsi dire que les neurones impulsionnels implémentent de façon intrinsèque un mécanisme de seuillage et que les informations non propagées ne susciteront aucun calcul supplémentaire.

Les modèles impulsionnels peuvent exhiber de nombreux types de comportements en fonction de la façon dont ils sont paramétrés : en particulier, ils peuvent se comporter comme des intégrateurs temporels ou comme des détecteurs de coïncidences [König *et al.*, 1996]. Les neurones intégrateurs accumulent les informations entrantes et déchargent lorsqu'ils ont reçu suffisamment d'excitations (partie gauche de la figure 3.5). Au contraire, les neurones détecteurs de coïncidences n'émettent des potentiels d'action que lorsqu'ils reçoivent des informations corrélées temporellement (figure 3.5 à droite). Nous avons utilisé ce dernier type de comportement pour réaliser, sur la carte de saillances, la fusion des informations issues des différentes fréquences spatiales. La préactivation par les basses fréquences est réalisée en introduisant un délai temporel entre les informations des basses fréquences – qui arrivent quelques millisecondes plus tôt sur la carte de saillances – et celles des hautes fréquences. Ce processus est rendu possible par la nature temporelle des informations échangées entre les neurones impulsionnels. Il serait difficile d'implémenter une opération de fusion similaire avec un algorithme de traitement d'images classique.

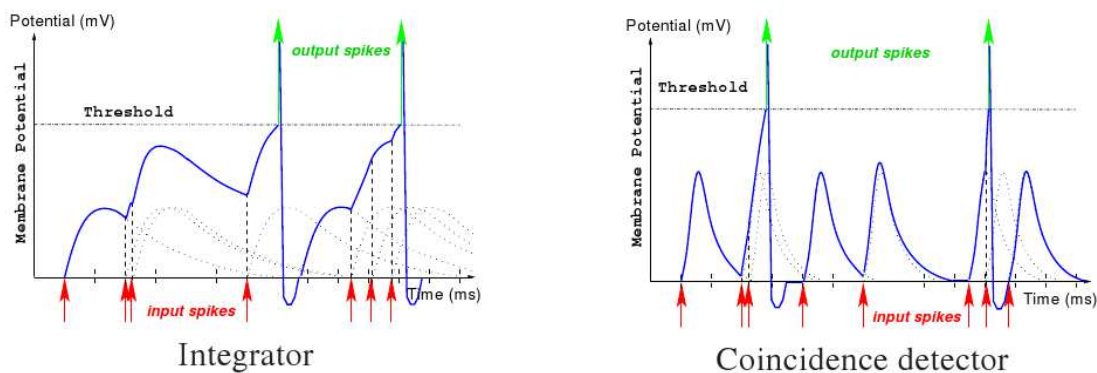


FIGURE 3.5 En fonction des paramètres choisis (seuil et constante d'intégration membranaire), un neurone impulsif peut se comporter comme un intégrateur (à gauche) ou comme un détecteur de coïncidences (à droite) [König *et al.*, 1996]. Ainsi pour un même train de potentiels d'action (*input spikes* indiqués par les flèches en rouge) arrivant sur un neurone impulsif, celui-ci peut décharger à des instants différents (*output spikes* indiqués par des flèches vertes). [Source : [Paugam-Moisy, 2006]]

Concernant les approches neuronales en vision artificielle, nous avons choisi d'utiliser

une façon de convertir les entrées qui est différente de celles existantes. C'est-à-dire que la valeur d'un pixel est transformée en un train de potentiels d'action, plutôt qu'en un unique potentiel d'action comme c'est le cas avec le codage par rang [VanRullen *et al.*, 1998; VanRullen et Thorpe, 2001; Perrinet, 2003]. Les approches qui utilisent le codage par rang introduisent une réinitialisation du système (un *reset*) après chaque présentation d'image [Delorme *et al.*, 1999; Perrinet, 2003] et on parle alors d'images flashées. La compréhension de la perception comme un processus situé nous a amené à considérer un système qui reçoive des séquences d'images, plutôt que des images flashées, et qui est réinitialisé après chaque saccade. Ce type de système est plus proche des observations biologiques, pour lesquelles un tel *reset* peut être observé après chaque saccade [Burr *et al.*, 1994; Thiele *et al.*, 2002] mais pas pendant une fixation. Cette prise en compte des séquences d'images nous permet de proposer un mécanisme de suivi de région saillante, qui fonctionne grâce à l'arrivée en continu de potentiels d'action.

Comme nous l'avons vu dans la partie 3.2.3, plusieurs implémentations neuronales de la vision préattentive ont été proposées [Ahrns et Neumann, 1999; Maillard *et al.*, 2005; Fix *et al.*, 2007; de Brecht et Saiki, 2006]. Ces implémentations emploient des techniques classiques de filtrage pour extraire des traits caractéristiques d'une image d'entrée puis utilisent le résultat de ce filtrage comme entrée de leur réseau. Nous proposons une méthode qui permet de réaliser les opérations de filtrage avec les neurones impulsionnels. Ce filtrage neuronal exploite la représentation temporelle des informations pour implémenter une méthode qualitativement différente du filtrage par convolution. La partie suivante présente les intérêts de la méthode neuronale.

3.3.2 Un système *anytime*

Nous proposons d'utiliser une méthode de filtrage qui utilise le codage temporel des neurones impulsionnels. Cette méthode permet d'obtenir le résultat du filtrage par approximations successives. Il est ainsi possible d'obtenir un résultat proche de celui du filtrage par convolution, mais qui n'est pas parfaitement identique. Une telle méthode, qui permet d'échanger de la précision du résultat contre du temps de calcul, appartient à la classe des algorithmes *anytime*. Cette classe d'algorithme a été d'abord proposée par [Horvitz, 1987] (sous le nom de *flexible computations*) et par [Dean et Boddy, 1988; Boddy et Dean, 1994]. Les caractéristiques les plus importantes d'un algorithme *anytime* sont :

- de pouvoir être arrêté à n'importe quel moment et de retourner une réponse ;
- la qualité générale de la réponse donnée s'améliore en fonction du temps de calcul alloué.

Pour illustrer la seconde propriété, on peut considérer une fonction $\mu(x) = y$ qui évalue, en fonction du temps processeur utilisé x par un algorithme, la qualité de la réponse y de cet algorithme. On pose, par définition, que $\mu(x) = 0$ si l'algorithme ne peut pas fournir de réponse dans le temps imparti. Cette fonction permet de mesurer l'intérêt d'avoir alloué un temps processeur x à un algorithme donné. En utilisant cette fonction, il est possible de distinguer deux classes d'algorithmes, qui sont représentées sur la figure 3.6. La première, représentée à gauche, caractérise les algorithmes dont l'amélioration de la qualité des résultats est discontinue : $\mu(x) = 0$ quand $0 \leq x < \tau$ et $\mu(x) = \kappa$ quand $\tau \leq x < \infty$. La seconde classe regroupe les algorithmes *anytime*, dont la qualité de la réponse est croissante et s'améliore lorsque l'algorithme dispose de plus de temps de calcul. La partie droite de la figure 3.6 illustre un type de réponse possible pour un algorithme *anytime* pour lequel la qualité de la réponse évolue en $1 - e^{-\lambda x}$.

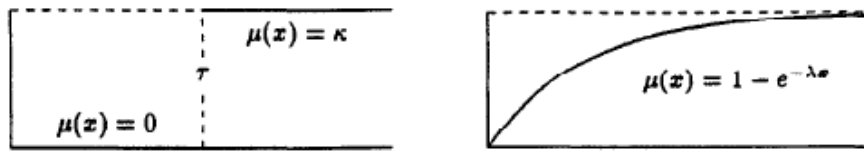


FIGURE 3.6 Évolution de la mesure de la qualité de la réponse $\mu(\cdot)$ en fonction du temps processeur pour un algorithme standard (à gauche) et pour un algorithme *anytime* (à droite). Les abscisses représentent le temps processeur δ alloué à l'algorithme, les ordonnées représentent la qualité de la réponse. [Source : [Dean et Boddy, 1988]]

Il existe de nombreuses familles d'algorithmes appartenant à la classe des algorithmes *anytime*, une liste non exhaustive pourrait inclure des méthodes de calculs numériques, la méthode itérative des éléments finis, les méthodes de Monte-Carlo ou bien encore certaines méthodes pour approcher l'évaluation de réseaux bayésiens. Ces algorithmes sont donc utilisés dans diverses applications comme la sélection de *features* [Last et al., 2001], le diagnostic temps réel [Zilberstein et Russell, 1996; Verberne et al., 2000] ou l'optimisation de contraintes [Loudni et Boizumault, 2003].

L'adaptation de la qualité de la réponse en fonction des ressources disponibles a été mise à profit pour développer des systèmes robotiques capables d'interagir de façon plus souple avec leur environnement [Dean et Boddy, 1988; Boddy et Dean, 1989; Pineau et al., 2003]. Lorsqu'il faut réagir rapidement, les décisions peuvent être prises en peu de temps à partir de très peu d'informations et lorsqu'il est moins urgent de fournir une réponse comportementale, le processus de décision peut collecter plus d'informations pour mieux affiner les critères qui lui permettront de choisir le comportement le plus adapté.

La méthode de filtrage neuronal mise en place nous permet de proposer une approche qui fournit les saillances graduellement, en fonction des ressources disponibles, et qui permet d'obtenir une approximation des saillances les plus importantes dès l'obtention des premiers résultats. Pour cela nous utilisons les informations relatives à la latence du premier potentiel d'action émis et à la fréquence de décharge moyenne de chaque neurone : ces codages de l'information semblent très proches de ceux mis en œuvre par certaines cellules ganglionnaires de la rétine [Gollisch et Meister, 2008b; Gollisch et Meister, 2008a].

3.3.3 Méthode d'implémentation du système

Les neurones impulsionnels utilisent un mode de calcul qui diffère de celui des neurones formels et ils requièrent aussi d'utiliser des approches différentes pour leur programmation. Un des choix les plus importants pour développer un simulateur de neurones impulsionnels concerne la stratégie retenue pour implémenter l'algorithme de simulation des réseaux de neurones, qui peut être synchrone (ou *clock-driven*), asynchrone (ou *event-driven*) ou bien hybride.

Les algorithmes synchrones reposent sur l'utilisation d'une horloge interne, qui fixe les pas de temps Δt auxquels les variables d'état des neurones doivent être mises à jour. Ces variables d'état sont généralement obtenues avec un algorithme d'approximation pour intégrer les équations différentielles (par exemple les méthodes d'Euler ou de Runge-Kutta). Le choix du pas de temps Δt revêt une importance majeure, puisqu'il influe directement sur le coût computationnel de l'algorithme et sur la précision de l'intégration, l'amélioration de la précision de l'intégration se faisant au dépend du temps physique nécessaire à la simulation du réseau. De nombreux simulateurs, adaptés pour modéliser des réseaux de neurones réalistes et mis à la disposition de la communauté des neurosciences, utilisent cette approche : NEURON [Hines et Carnevale, 1997], GENESIS [Bower et Beeman, 1998], NeoCortical simulator, Circuit Simulator (CSIM), BRIAN [Goodman et Brette, 2008], NEST [Diesmann et Gewaltig, 2003] ou SPLIT [Hammarlund et Ekeberg, 1998] par exemple.

Le principal problème des simulations synchrones est que les neurones sont tous mis à jour en même temps, ce qui fait que les potentiels d'action émis sont tous "alignés" temporellement sur les pas de temps. L'approche synchrone, en discrétisant un système neuronal continu, introduit des erreurs systématiques [Cessac et Samuelides, 2006], même lorsque les équations différentielles sont approchées avec une bonne précision. Ceci peut être particulièrement problématique dans les cas où le modèle neuronal

prend en compte le temps exact d'émission des potentiels d'action, comme pour la modification des poids synaptiques par STDP par exemple [Rudolph et Destexhe, 2007].

L'approche asynchrone évite (en principe) cet écueil en traitant les potentiels d'action comme des événements et en calculant leurs temps d'émission au niveau de précision de la machine. De plus, cette approche utilise les principes de la programmation événementielle pour traiter uniquement les neurones qui reçoivent des potentiels d'action [Watts, 1994]. Les calculs nécessaires à la simulation ne dépendent ainsi plus du nombre de neurones mais de l'activité du réseau⁶, c'est-à-dire des potentiels d'action émis. Cependant, l'utilité de cette approche pour réduire le coût computationnel des simulations n'est pas toujours évidente en pratique [Morrison *et al.*, 2005].

Le principal inconvénient de l'approche asynchrone est qu'elle ne peut être mise en œuvre que pour des modèles neuronaux pour lesquels il existe une solution explicite des équations différentielles⁷, ce qui fait qu'elle reste encore relativement peu utilisée. On pourra noter que quelques simulateurs l'implémentent comme NEURON [Hines et Carnevale, 2004], NEST [Diesmann et Gewaltig, 2003], DAMNED [Mouraud *et al.*, 2006] ou MVAspike [Rochel et Martinez, 2003].

Nous avons choisi d'implémenter notre propre système de neurones impulsionnels. Ce système utilise une approche hybride [Morrison *et al.*, 2005] entre les approches synchrone et asynchrone. Comme nous souhaitons pouvoir intégrer plusieurs types de neurones différents dans un même réseau, sans être trop limité dans le choix des modèles neuronaux, nous avons implémenté un simulateur synchrone. Cependant, seuls les neurones recevant des potentiels d'action sont mis à jour, ce qui permet de réduire le nombre de calculs effectués à chaque pas de temps. Ce système d'environ 10 000 lignes de code a été écrit en langage Ada et fonctionne en multithreads. Il est détaillé dans l'annexe A.

6. Voir l'analyse comparée de la complexité des algorithmes synchrones et asynchrones proposée dans [Brette *et al.*, 2007].

7. Pour certains modèles de neurones impulsionnels, même sans solution explicite, il est possible d'approcher le temps de décharge sans engendrer un coût computationnel trop élevé [Brette, 2006; Brette, 2007].

Chapitre 4

Filtrage neuronal d'image

Sommaire

4.1 Le filtrage neuronal	74
4.1.1 La transduction	74
4.1.2 L'intégration	76
4.2 Les codages impulsionnel et fréquentiel	76
4.2.1 Codage des entrées	77
4.2.2 Caractérisation du codage impulsionnel pour le filtrage	79
4.2.3 Utilisation du codage fréquentiel	83
4.2.4 Une approche <i>anytime</i>	86
4.3 Résultats expérimentaux	87
4.3.1 Méthodes	87
4.3.2 Comparaison des méthodes de filtrage	89
4.3.3 Études de paramètres	101
4.4 Conclusion	107

Ce chapitre propose une méthode originale pour implémenter un filtrage d'image avec des neurones impulsionnels. Cette méthode permet d'obtenir des résultats proches du filtrage par convolution et nous l'avons utilisé dans l'architecture préattentionnelle qui sera présentée dans le chapitre 5. Nous verrons tout d'abord les différentes étapes de ce filtrage, qui implémente un algorithme *anytime* (partie 4.1). Cette approche tire parti du fait que les neurones impulsionnels donnent accès à la fois au codage temporel précis de l'information ainsi qu'au codage fréquentiel. Nous étudierons les propriétés de ces deux types de codage dans la partie 4.2. Nous proposons ensuite d'étudier expérimentalement le filtrage neuronal, au travers d'une comparaison avec le filtrage par convolution et d'une étude de paramètres (partie 4.3). La partie 4.4 conclut ce chapitre.

4.1 Le filtrage neuronal

La méthode de filtrage proposée repose sur l'utilisation d'intégrateurs à fuite (voir la partie 2.2.2) pour lesquels nous pouvons reformuler les variations du potentiel membranaire d'un neurone i comme suit :

$$\begin{cases} \frac{dV_i}{dt} = -\lambda_i V_i(t) + u_i(t), \text{ si } V_i < \vartheta \\ \text{sinon émission d'un potentiel d'action et } V_i \text{ revient à } V_{\text{reset}} \end{cases} \quad (4.1)$$

où λ_i est la constante de relaxation membranaire¹ et la fonction $u_i(t)$ représente une commande. Cette commande décrit les perturbations appliquées au potentiel membranaire et peut prendre différentes formes. Elle permet par exemple de simuler une injection de courant, un bruit additif ou l'influence des potentiels d'action reçus.

La méthode de filtrage neuronal proposée peut être décomposée en deux opérations : d'une part la transduction, pour transformer une image en trains de potentiels d'action et d'autre part l'intégration, où des neurones intègrent ces trains de potentiels d'action et déchargent de façon maximale lorsque leurs entrées sont similaires au filtre qu'ils implémentent.

4.1.1 La transduction

La transduction est la transformation réalisée par les récepteurs assurant la conversion d'un signal externe en un signal interne, utilisable par le système. Il s'agit de transformer les informations de luminance ou de couleurs en un signal temporel. Cette opération est réalisée avec une carte d'entrée neuronale 2D de taille $N \times M$, pour une image d'entrée de $N \times M$ pixels, où chaque neurone reçoit un courant proportionnel à la luminance² du pixel correspondant (voir la figure 4.1). C'est-à-dire que nous utilisons une fonction du type $u_i(t) = K L(\mathbf{r}, t)$ dans l'équation (4.1), avec K un coefficient multiplicateur et $L(\mathbf{r}, t)$ la fonction qui décrit les variations de luminance du pixel i de coordonnées $\mathbf{r} = (x, y), \forall x \in \{1..N\}, \forall y \in \{1..M\}$ dans une séquence d'images. Ainsi, plutôt que de convoluer l'image ligne par ligne et colonne par colonne, l'étape de transduction définit un ordre de traitement des pixels qui dépend des données, en traitant les pixels en fonction de leur luminance.

Il est tout à fait possible d'envisager d'autres types de transduction qui soient plus complexes, comme la sortie du modèle rétinien proposée par [Wohrer, 2008; Wohrer et Kornprobst, 2008], ou plus simples, en ne s'intéressant par exemple qu'au

1. On pourra noter que $\lambda = 1/\tau = \frac{g}{C}$, voir l'équation (2.2) de la partie 2.2.2.

2. Le cas des couleurs est détaillé dans la partie 5.1.1.

temps d'émission du premier potentiel d'action [VanRullen *et al.*, 1998; VanRullen, 2000; Perrinet, 2003]. Ce dernier type de transduction utilise un codage par rang et a donné lieu à un logiciel, SpikeNet, qui permet de simuler de larges réseaux en temps réel [Thorpe *et al.*, 2004]. Cependant avec ce type de codage une séquence d'image est considérée comme une suite d'images "flashées" les unes après les autres : les neurones de la carte d'entrée reviennent à leur état initial après chaque image. Nous avons considéré un mécanisme de transduction différent, qui n'introduit pas de remise à zéro après chaque image et qui prend en compte l'ensemble des potentiels d'action émis comme nous l'avons expliqué dans la partie 3.3. Ceci nous permet d'utiliser les informations temporelles précises (instants de décharge des neurones) et moyennes (fréquences de décharge) pour obtenir le résultat du filtrage.

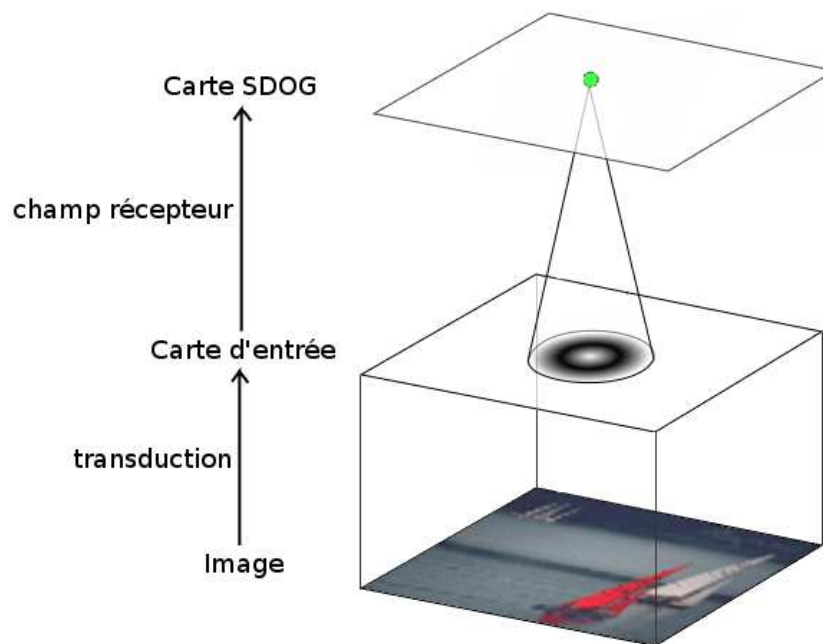


FIGURE 4.1 Le filtrage neuronal est réalisé en deux opérations, la transduction et l'intégration. Pour la transduction, les neurones de la carte d'entrée reçoivent un courant qui dépend des valeurs de luminance des pixels de l'image, avec une correspondance d'un neurone par pixel. L'intégration est réalisée par les neurones de la carte SDOG qui implémente un filtre similaire à un DOG. Chaque neurone reçoit les potentiels d'actions émis par les neurones de la carte d'entrée situés dans son champ récepteur. Un neurone de la carte SDOG est représenté en vert ainsi que son champ récepteur. Les niveaux de gris à l'intérieur du champ récepteur représentent les valeurs des poids de connexion : les valeurs les plus positives sont en blanc et les plus négatives sont en noir.

4.1.2 L'intégration

La seconde opération, l'intégration, est réalisée par les neurones d'une seconde carte neuronale. Nous considérerons ci-après le cas d'une carte réalisant un filtrage similaire à celui obtenu par une convolution avec une différence de gaussiennes (ou DOG), que nous appellerons SDOG (figure 4.1).

Cette carte est reliée à la carte d'entrée au travers d'un masque de connexion [Wolff *et al.*, 1999], qui explicite les poids et les délais des connexions entre un neurone de la carte SDOG et des neurones de la carte d'entrée. Un masque définit un type de champ récepteur et il est le même pour tous les neurones de la carte SDOG. Bien que les poids et les délais soient fixes, ceci peut s'apparenter à du *weight sharing* [Fukushima, 1980; Rumelhart et McClelland, 1986].

Les neurones de la carte SDOG intègrent les potentiels d'actions émis par la carte d'entrée et les temps de décharge des neurones de la carte SDOG indiquent le résultat du filtrage. Plus les entrées sont proches du filtre utilisé, plus le neurone décharge rapidement. Les potentiels d'action émis par les neurones de la carte SDOG sont porteurs d'informations que l'on peut interpréter de deux façons selon que l'on s'intéresse au codage impulsionnel ou au codage fréquentiel. Le codage impulsionnel est défini pour un neurone donné comme la latence du premier potentiel d'action, les potentiels d'actions suivants ne sont pas pris en compte. Le codage fréquentiel reflète la fréquence moyenne de décharge : c'est l'intervalle moyen qui sépare l'émission de deux potentiels d'action (intervalle inter-spike ou ISI) pour un neurone donné.

4.2 Les codages impulsionnel et fréquentiel

Le tableau 4.2 explicite l'utilisation du codage impulsionnel et fréquentiel. La méthode proposée implique que pour filtrer une image les neurones de la carte d'entrée puissent émettre plusieurs potentiels d'action (PA sur le tableau 4.2). Ces potentiels d'action sont intégrés par les neurones de la carte SDOG et le résultat du filtrage est obtenu soit en utilisant uniquement le premier potentiel d'action émis pour chaque neurone de la carte SDOG, soit en calculant la fréquence moyenne de décharge sur plusieurs potentiels d'action. À titre de comparaison, le codage par rang [VanRullen *et al.*, 1998; Perrinet, 2003] peut être utilisé pour représenter le résultat d'un filtrage par convolution à l'aide d'un seul et unique potentiel d'action. Il n'existe pas de codage qui utilise plusieurs potentiels d'action de SDOG à partir d'un seul potentiel d'action émis en entrée.

		Carte d'entrée	
		1er PA	Plusieurs PA
Carte SDOG	1er PA	Codage par rang	Codage impulsionnel
	Plusieurs PA	—	Codage fréquentiel

TABLE 4.1 Tableau explicitant les différents types de codage temporels possibles en fonction du nombre de potentiels d'action (PA) pris en compte sur la carte d'entrée et sur la carte SDOG. Il n'y a pas de codage possible dans la case en bas à gauche.

Avant d'étudier les codages impulsionnel et fréquentiel, il est important de préciser que les neurones de la carte SDOG doivent être paramétrés pour exhiber un comportement d'intégrateur ([König *et al.*, 1996], voir la partie 3.3.1) pour obtenir des résultats approchant ceux d'une convolution. Il est en effet nécessaire de s'assurer que ces neurones prennent en compte les informations issues de différentes valeurs de luminance avant de décharger. Un choix adéquat des valeurs de poids et de la constante de relaxation membranaire permet de s'assurer que le neurone intègre plusieurs potentiels d'action avant de décharger. Une étude de paramètres concernant cet aspect est proposé dans la partie 4.3.3.

4.2.1 Codage des entrées

Nous allons détailler les différences entre les informations portées par le code impulsionnel et le code fréquentiel. La méthode de transduction utilisée induit une première remarque concernant les fréquences de décharge des neurones : des neurones de la carte d'entrées codant pour des valeurs de luminance différentes vont décharger avec des fréquences différentes. Ainsi, selon les paramètres choisis, un neurone codant pour une valeur de luminance élevée peut décharger beaucoup plus rapidement qu'un neurone codant pour une valeur de luminance plus faible (figure 4.2).

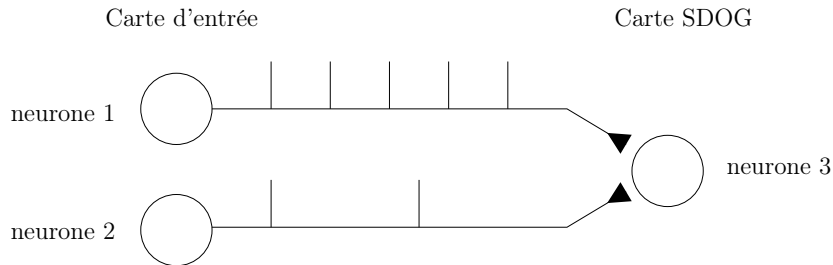


FIGURE 4.2 Potentiels d'action émis par deux neurones de la carte d'entrée recevant des entrées d'intensité différentes : leur fréquence de décharge varie en fonction de la luminance. Ces neurones sont connectés à un neurone de la carte SDOG.

Si l'on considère un neurone i de la carte d'entrée, qui reçoit un courant d'entrée

constant et proportionnel à la luminance L_i d'un pixel d'une image :

$$\begin{cases} \frac{dV_i}{dt} = -\lambda_i V_i(t) + K L_i, \text{ si } V_i < \vartheta \\ \text{sinon émission d'un potentiel d'action et } V_i \text{ revient à } V_{\text{reset}} \end{cases} \quad (4.2)$$

Ce courant peut amener le neurone à décharger si $L_i > \lambda_i \vartheta / K$ et dans ce cas, avec comme condition initiale $V(t_0 = 0) = 0$, le temps d'émission \hat{t}_i du potentiel d'action est :

$$\hat{t}_i = -\frac{1}{\lambda_i} \ln \left(1 - \frac{\lambda_i \vartheta}{K L_i} \right) \quad (4.3)$$

La figure 4.3 illustre les variations de \hat{t} en fonction de la luminance, ainsi que les fréquences de décharge mesurées expérimentalement. On peut noter sur la partie droite de la figure 4.3 que certaines valeurs de τ amènent le neurone à décharger seulement lorsque L_i est élevée³. Les résultats des simulations présentées sur cette figure ont été obtenus en soumettant un neurone à un courant proportionnel une valeur de luminance L_i normalisée entre 0 et 1. La valeur du coefficient multiplicateur K est fixée à 1, le potentiel de repos est $E_l = -70\text{mV}$ ⁴, le seuil $\vartheta = -40\text{mV}$ et le potentiel membranaire est remis à $V_{\text{reset}} = -80\text{mV}$ après chaque émission de potentiel d'action. Les valeurs de τ sont indiquées sur la figure. Les simulations sont réalisées sur 10^4 pas de temps, soit 1 seconde de temps simulé.

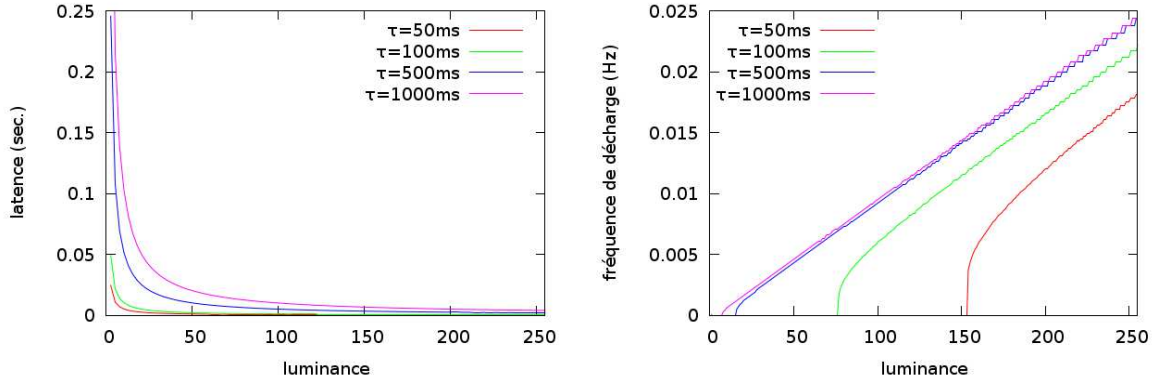


FIGURE 4.3 Gauche : temps d'émission du premier potentiel d'action (ou latence) d'un neurone de la carte d'entrée en fonction de la valeur de la luminance. Les courbes montrent les variations de \hat{t} , voir l'équation (4.3), pour plusieurs valeurs de $\tau \in \{50, 100, 500, 1000 \text{ ms}\}$, avec $\tau = 1/\lambda$. Droite : fréquence de décharge Φ mesurée expérimentalement en fonction de la luminance, pour plusieurs valeurs de τ .

Lorsque la valeur de L_i est suffisante pour susciter un potentiel d'action, la fréquence de décharge du neurone augmente linéairement avec la luminance. En effet, si on consi-

3. ainsi que pour K et ϑ , non représentés sur la figure 4.3

4. La valeur de E_l intervient dans le changement de variable suivant $V_i(t) \leftarrow V_i(t) - E_l$.

dère la fréquence de décharge $\Phi_i = 1/\hat{t}_i$, avec L_i supposée constante dans le temps, on obtient :

$$\Phi_i = -\frac{\lambda_i}{\ln\left(1 - \frac{\lambda_i \vartheta}{KL_i}\right)} \quad (4.4)$$

On considère les valeurs de luminance L_i élevées : on pose $x = \frac{\lambda_i \vartheta}{KL_i}$ et on étudie le développement limité en 0 de $\frac{1}{\ln(1-x)}$. On a, à l'ordre n et au voisinage de 0, $\ln(1-x) = -\sum_{k=1}^n x^k/k + o(x^n)$, soit au premier ordre $\ln(1-x) \approx -x$. Ainsi les variations de Φ_i pour des valeurs de luminance L_i élevées peuvent être approximées par :

$$\Phi_i \approx \frac{K}{\vartheta} L_i \quad (4.5)$$

La fréquence de décharge Φ_i des neurones de la carte d'entrée varie linéairement en fonction de la luminance L_i .

4.2.2 Caractérisation du codage impulsionnel pour le filtrage

Le codage impulsionnel permet d'implémenter un filtrage rapide et approximatif. Les neurones de la carte SDOG déchargent avant d'avoir pris en compte toutes les informations disponibles dans leur champ récepteur. Ils réalisent un filtrage à partir d'un sous-échantillon du signal qui privilégie les valeurs élevées de luminance. Il est possible de décrire ce processus comme une prise de décision : au vu des informations qu'un neurone a reçu, il décharge pour indiquer que ses entrées sont proches de la réponse impulsionnelle du filtre qu'il implémente. Les résultats sont ainsi ordonnés temporellement, ce qui permet de récupérer les coefficients les plus importants en premier.

Cependant, cette décision est une approximation et peut introduire un biais dans le filtrage. La figure 4.4 illustre une situation où cette approximation peut être faussée : si le neurone 4 décharge avant d'avoir pris en compte un potentiel d'action inhibiteur émis par le neurone 3, le résultat du filtrage est faussé. Cette situation peut survenir lorsque la partie inhibitrice du champ récepteur est positionnée sur une partie de l'image où la luminance est plus faible.

Le modèle neuronal employé sur la carte SDOG est un intégrateur à fuite dont le modèle de commande (voir l'équation (4.1)) modélise l'influence de potentiels post-synaptiques instantanés. Le train de potentiels d'action, composé de $N_i(t)$ potentiels d'action⁵ émis aux instants t_i^f , en provenance du neurone i reçus par le neurone j est

5. Pour plus de lisibilité, $N_i(t)$ est noté N_i dans la suite du manuscrit.

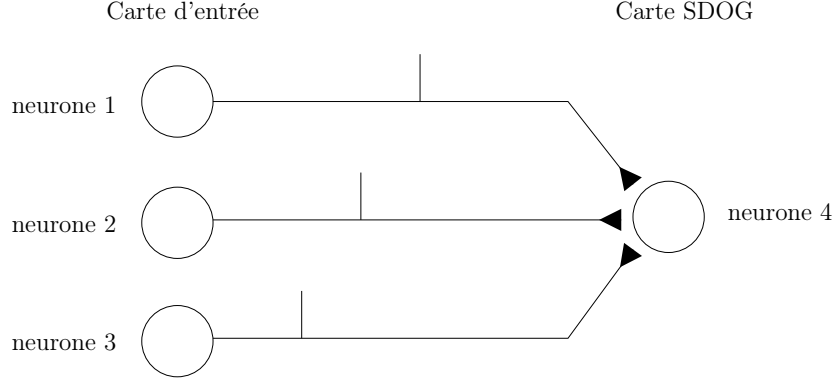


FIGURE 4.4 Potentiels d'action émis par les neurones 1, 2 et 3 de la carte d'entrée, recevant des entrées d'intensités différentes, et qui sont connectés au neurone 4 de la carte SDOG. Le neurone 4 peut décharger après avoir reçu les potentiels d'action émis par 1 et 2, mais avant d'avoir reçu celui émis par 3. Cette situation est problématique, en particulier dans le cas où le neurone 3 est inhibiteur : le neurone 4 n'aurait peut être pas déchargé s'il avait pris en compte ce potentiel d'action.

décrit par :

$$S_i(t) = \sum_{f=1}^{N_i} \delta(t - t_i^f) \quad (4.6)$$

où $\delta(x)$ est une distribution de Dirac, avec $\delta(x) = 0$ pour $x \neq 0$ et $\int_{-\infty}^{\infty} \delta(x) dx = 1$. Ces potentiels d'action provoquent sur le neurone j un changement ponctuel instantané (ou saut de potentiel) dont l'amplitude est fixée par le poids synaptique w_{ij} :

$$\begin{cases} \frac{dV_j}{dt} = -\lambda_j V_j(t) + \sum_{i=1}^{P_j} w_{ij} S_i(t), & \text{si } V_j < \vartheta \\ \text{sinon émission d'un potentiel d'action et } V_j \text{ revient à } V_{\text{reset}} \end{cases} \quad (4.7)$$

Le potentiel membranaire de j est modifié par les trains de potentiels d'actions émis par les P_j neurones présynaptiques.

L'évolution du potentiel membranaire j est décrite par :

$$V_j(t) = \sum_{i=1}^{P_j} w_{ij} \sum_{f=1}^{N_i} e^{-\lambda_j(t-t_i^f)} H(t, t_i^f) \quad (4.8)$$

où $H(t, t_i^f)$ est la fonction échelon, qui vaut 1 quand $t \geq t_i^f$ et 0 sinon.

Nous allons proposer une analyse du codage impulsionnel, en partant de la constatation que les neurone de la carte d'entrée déchargent régulièrement. Si l'on considère l'instant t_i^f , auquel un neurone i de la carte d'entrée émet un f -ième potentiel d'action, on a $t_i^f = f\hat{t}_i$, où \hat{t}_i est l'intervalle inter-spike. Nous pouvons donc écrire l'équation (4.8)

sous la forme :

$$V_j(t) = \sum_{i=1}^{P_j} w_{ij} \sum_{f=1}^{N_i} e^{-\lambda_j(t-f\hat{t}_i)} H(t, f\hat{t}_i) \quad (4.9)$$

De plus, nous nous intéressons uniquement à tous les $f\hat{t}_i < t$, pour cela nous choisissons N_i de telle sorte que :

$$N_i\hat{t}_i \leq t < (N_i + 1)\hat{t}_i \quad (4.10)$$

Ainsi, nous avons $\forall f \in [1, N_i], H(t, f\hat{t}_i) = 1$.

Plaçons-nous à l'instant T_j , en choisissant T_j de telle sorte que pendant l'intervalle $[t_0, T_j[$ le neurone n'ait pas encore déchargé :

$$V_j(T_j) = \sum_{i=1}^{P_j} w_{ij} \sum_{f=1}^{N_i} e^{-\lambda_j(T_j-f\hat{t}_i)} \quad (4.11)$$

Nous pouvons exprimer T_j sous la forme :

$$\exists \zeta_i \text{ tel que } 0 \leq \zeta_i < 1 \text{ et } T_j = N_i\hat{t}_i + \zeta_i\hat{t}_i \quad (4.12)$$

La figure 4.5 illustre la situation décrite par l'équation (4.12) : à l'instant T_j le neurone j a reçu N_i potentiels d'action du neurone i . Nous utilisons la définition de N_i de l'équation (4.10) pour exprimer T_j en fonction de \hat{t}_i . Par rapport à l'instant $N_i\hat{t}_i$, l'instant T_j intervient au plus tôt à $N_i\hat{t}_i$ et au plus tard juste avant l'instant $(N_i + 1)\hat{t}_i$.

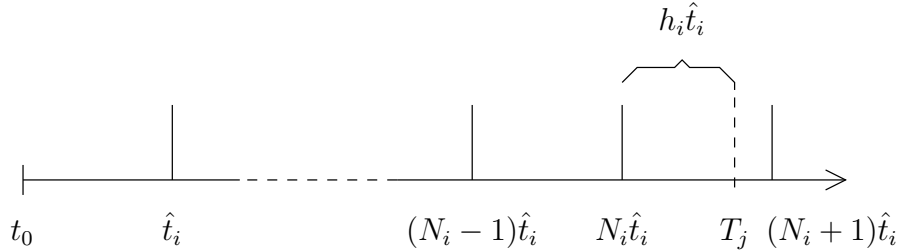


FIGURE 4.5 Un neurone i émet des potentiels d'action aux instants $f\hat{t}_i$. Il est possible d'exprimer un instant T_j donné en fonction de \hat{t}_i . Voir l'équation (4.12).

Nous pouvons donc écrire l'équation (4.11) comme :

$$V_j(T_j) = \sum_{i=1}^{P_j} w_{ij} Z_i \sum_{f=1}^{N_i} e^{-\lambda_j\hat{t}_i(N_i-f)} \quad (4.13)$$

avec $Z_i = e^{-\lambda_j\hat{t}_i\zeta_i}$.

Nous remarquons que :

$$\sum_{f=1}^{N_i} e^{-\lambda_j \hat{t}_i (N_i - f)} = \sum_{k=0}^{N_i-1} \left(e^{-\lambda_j \hat{t}_i} \right)^k \quad (4.14)$$

En posant $x = e^{-\lambda_j \hat{t}_i}$, nous pouvons remarquer que :

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad (4.15)$$

et en déduire :

$$\sum_{k=0}^{\infty} e^{-\lambda_j \hat{t}_i} = \frac{1}{1 - e^{-\lambda_j \hat{t}_i}} \quad (4.16)$$

Ainsi, nous avons :

$$\lim_{t \rightarrow \infty} V_j(t) = \sum_i \frac{Z_i w_{ij}}{1 - e^{-\lambda_j \hat{t}_i}} \quad (4.17)$$

Comme nous l'avons expliqué dans la partie 4.2, nous avons choisi d'utiliser des neurones SDOG qui se comportent comme des intégrateurs, c'est-à-dire pour lesquels $\lambda_j \ll 1$ et $\lambda_j \ll \hat{t}_i$. Ainsi, nous avons $\lambda_j \hat{t}_i \approx 0$ et comme $e^{-\lambda_j \hat{t}_i} < Z_i \leq 1$, nous avons donc $H_i \approx 1$.

Considérons le développement limité, à l'ordre 1 et en 0, de $e^{\lambda_j \hat{t}_i} \approx 1 - \lambda_j \hat{t}_i$. L'équation (4.17) conduit ainsi à :

$$\lim_{t \rightarrow \infty} V_j(t) \approx \sum_i \frac{w_{ij}}{\lambda_j \hat{t}_i} = \sum_i \frac{w_{ij}}{\lambda_j} \Phi_i \quad (4.18)$$

En utilisant le résultat de l'équation (4.5), on obtient pour des valeurs de L_i élevées :

$$\lim_{t \rightarrow \infty} V_j(t) \approx \sum_i \frac{K}{\lambda_j \vartheta} w_{ij} L_i \quad (4.19)$$

avec $\frac{K}{\lambda_j \vartheta} \approx 1$ car nous avons utilisé des neurones intégrateurs pour lesquels $\lambda_j \ll 1$ et le seuil ϑ est grand par rapport aux poids synaptiques w_{ij} .

Le résultat d'un filtrage par convolution est donné par une réponse du type :

$$R_j = \sum_i w_{ij} L_i \quad (4.20)$$

Nous pouvons donc constater que le codage impulsionnel permet d'approcher le résultat d'un filtrage par convolution. Cette approximation est valable lorsque le neurone j intègre l'ensemble des potentiels d'action émis par les neurones situés dans son champ récepteur avant de décharger. Ce n'est pas le cas en pratique puisque les neurones de

SDOG déchargent avant d'avoir intégré tous les potentiels d'action (voir la figure 4.4).

Pour caractériser une situation, plus réaliste, nous pouvons reprendre l'équation (4.15) et l'exprimer sous la forme :

$$\sum_{k=0}^{\infty} x^k = \sum_{k=0}^{N-1} x^k + x^N \sum_{l=0}^{\infty} x^l \quad (4.21)$$

$$= \sum_{k=0}^{N-1} x^k + x^N \left(\frac{1}{1-x} \right) \quad (4.22)$$

On en déduit :

$$\sum_{k=0}^{N-1} x^k = \frac{1-x^N}{1-x} \quad (4.23)$$

En appliquant ce résultat à l'équation (4.11) avec $x = e^{-\lambda_j \hat{t}_i}$, nous pouvons en conclure que :

$$V_j(T_j) = \sum_{i=1}^{P_j} w_{ij} H_i \frac{1 - e^{-\lambda_j N_i \hat{t}_i}}{1 - e^{-\lambda_j \hat{t}_i}} \quad (4.24)$$

Ce résultat permet d'exprimer la valeur du potentiel membranaire d'un neurone de la carte SDOG à l'instant T_j . Il a alors intégré $N = \sum_{i=1}^{P_j} N_i$ potentiels d'action provenant des neurones de la carte d'entrée situés dans son champ récepteur. On pose $Q = \frac{\lambda_j \vartheta}{K}$ pour exprimer le potentiel V_j à l'instant T_j en fonction de L_i . Dans le cas où les valeurs de luminance sont élevées, c'est-à-dire si $Q \ll L_i$, nous avons :

$$V_j(T_j) \approx \sum_{i=1}^{P_j} w_{ij} \frac{1 - e^{-QN_i/L_i}}{1 - e^{-Q/L_i}} \quad (4.25)$$

4.2.3 Utilisation du codage fréquentiel

Les approximations réalisées par le codage impulsionnel sont progressivement réduites dès lors que l'on considère l'évolution du codage fréquentiel. Comme nous l'avons vu sur la figure 4.2, les neurones des cartes d'entrée déchargent de façon régulière, avec différentes fréquences de décharge. Les neurones de la carte SDOG intègrent ces entrées et émettent des potentiels d'action de façon non régulière, comme illustré sur la figure 4.6. On peut voir sur cette figure que les neurones déchargent de façon irrégulière car les entrées sont différentes dans chaque intervalle interspike (ISI).

L'ISI varie au cours du temps et les observations expérimentales montrent que la valeur moyenne de l'ISI tend vers le résultat du filtrage. Ceci s'explique par le fait que les

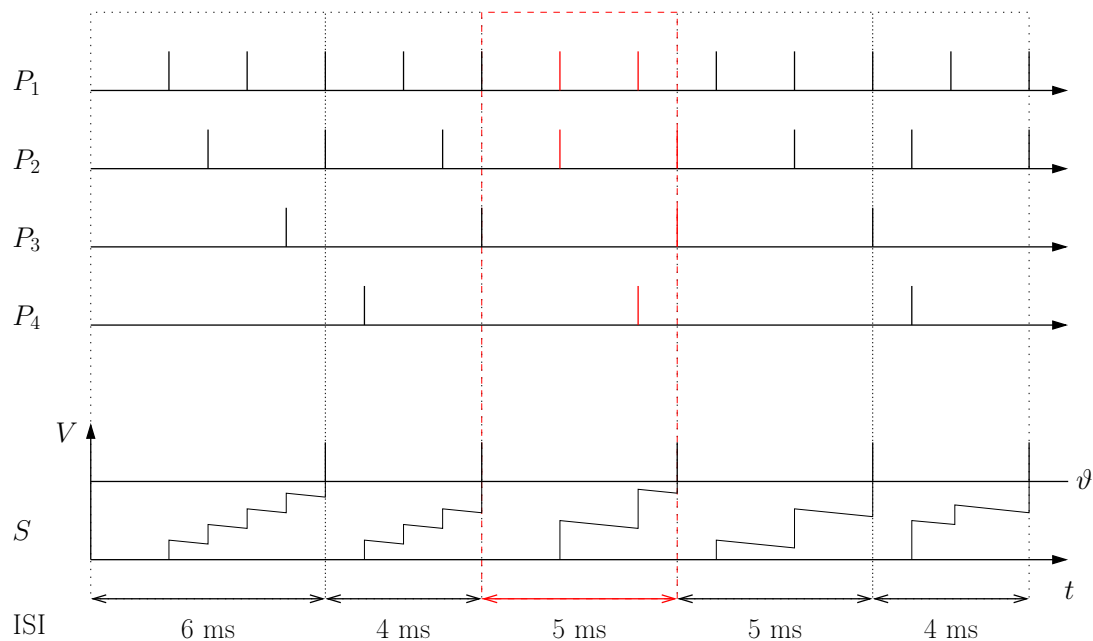


FIGURE 4.6 Illustration de l'évolution de l'intervalle interspike (ISI) d'un neurone S de la carte SDOG, qui reçoit les potentiels d'action des neurones présynaptiques P_1 , P_2 , P_3 et P_4 . Les neurones présynaptiques ont respectivement pour ISI 2 ms, 3 ms, 5 ms et 7 ms. Les potentiels d'action sont figurés par les barres verticales. Les variations du potentiel membranaire V du neurone S sont représentées sur la partie basse de la figure. Chaque ISI est représenté par un cadre en pointillé et sa valeur est indiquée en bas de la figure. À titre d'exemple, les potentiels d'action appartenant au troisième ISI sont indiqués en rouge et ils forment ce qui est appelé dans le texte un sous-échantillon.

neurones réalisent un sous-échantillonnage de leur entrées les trains de potentiels d'action qu'ils reçoivent. Un sous-échantillon correspond à la portion du train de potentiels d'action, émis par les neurones de la carte d'entrée et reçu par un neurone de SDOG, qui contribue à faire décharger un de neurone de SDOG. Par exemple, les potentiels d'action de P_1 , P_2 , P_3 et P_4 indiqués en rouge sur la figure 4.6 forment un sous-échantillon. Nous avons vu dans la partie précédente que ce sous-échantillonnage permet d'approcher le résultat du filtrage et que la qualité de l'approximation dépend du nombre d'entrées N prises en compte. Le codage fréquentiel prend en compte les résultats obtenus sur différents sous-échantillons pour calculer une valeur moyenne. Cette valeur moyenne, qui est utilisé par le codage fréquentiel, reflète les informations accumulées sur les différents sous-échantillons. D'un certain point de vue, on pourrait assimiler ce type d'approche à une méthode de Monte-Carlo. Le codage fréquentiel permet d'obtenir une approximation du résultat du filtrage dont la qualité s'améliore lorsque l'on considère un nombre plus important de sous-échantillons, c'est-à-dire lorsque le temps simulé dans le réseau est plus long.

Pour que l'approximation du filtrage obtenue avec le codage fréquentiel soit bonne, nous avons fait l'hypothèse que les sous-échantillons diffèrent suffisamment les uns des autres, ce qui semble être vérifié expérimentalement. Pour vérifier que la variabilité des sous-échantillons est suffisante pour produire un résultat de filtrage correct, nous avons étudié dans la partie expérimentale l'influence de l'ajout d'un bruit sur les neurones de la carte d'entrée. Si les sous-échantillons diffèrent assez peu les uns des autres, l'ajout d'un bruit dans les entrées permet d'augmenter la variabilité des sous-échantillons et donc d'améliorer la convergence vers le résultat du filtrage.

Dans la partie 4.3.2, nous allons donc comparer les résultats obtenus pour deux conditions expérimentales. Elles utilisent toutes les deux le codage fréquentiel et pour la seconde, nous avons ajouté un bruit (supposé uniforme) aux neurones de la carte d'entrée :

$$\begin{cases} \frac{dV_i}{dt} = -\lambda_i V_i(t) + K L_i + \eta(t), \text{ si } V_i < \vartheta \\ \text{sinon émission d'un potentiel d'action et } V_i \text{ revient à } V_{\text{reset}} \end{cases} \quad (4.26)$$

avec

$$\tau_n \frac{d\eta}{dt} = -\eta(t) + \rho(t) \quad (4.27)$$

où $\rho(t)$ est une variable aléatoire uniforme et τ_n une constante.

La figure 4.7 montre l'influence d'un tel bruit sur le temps de décharge d'un neurone. On peut constater que les quatre premiers potentiels d'action sont émis de façon précoce par rapport aux potentiels d'action déterministes et que cette tendance s'inverse ensuite

pour amener le neurone stochastique à décharger de façon tardive. Cette figure est obtenue avec un potentiel de repos $E_l = -70\text{mV}$, le seuil $\vartheta = -40\text{mV}$ et le potentiel membranaire est remis à $V_{\text{reset}} = -80\text{mV}$ après chaque émission de potentiel d'action. Les deux neurones reçoivent un courant d'entrée proportionnel à la luminance $L_i = 128$ avec $K = 1$ et pour le neurone stochastique, nous avons utilisé $\tau_n = 100\text{ms}$ et $\rho(t)$, une variable aléatoire supposée uniforme comprise entre -1 et 1

Il apparaît donc que l'ajout de bruit permet d'augmenter la variabilité des trains de potentiels d'action. Ceci nous permettra donc de vérifier si les sous-échantillons, c'est-à-dire le trains de potentiels d'action reçu par un neurone de SDOG avant qu'il émette lui-même des potentiels d'action, présentent une variabilité suffisante pour produire une bonne approximation du résultat du filtrage. Si ce n'est pas le cas, l'ajout de bruit permet d'améliorer la qualité de l'approximation produite.

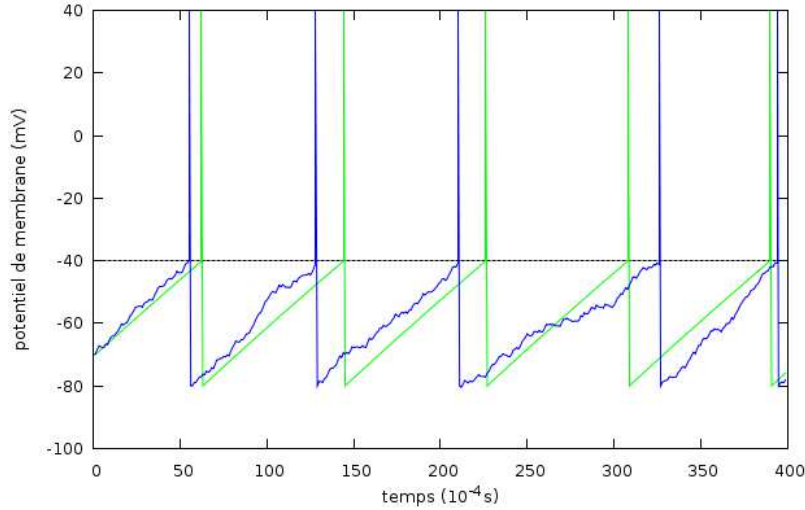


FIGURE 4.7 Évolution du potentiel membranaire de deux neurones, le premier est directement contrôlé par la luminance (en vert, voir l'équation (4.2)) et le second est fonction de la luminance et d'un bruit additif (en bleu, voir l'équation (4.26)).

La partie suivant propose une étude expérimentale du filtrage neuronal qui comporte une comparaison avec le filtrage par convolution et une étude de paramètres.

4.2.4 Une approche *anytime*

Le codage impulsionnel réalise une première approximation, rapide, du filtrage par convolution. Le codage fréquentiel permet d'obtenir une réponse beaucoup plus proche du filtrage par convolution mais comme il nécessite que les neurones aient émis plusieurs potentiels d'action, il n'est disponible qu'après un nombre de calculs plus important.

Les neurones impulsionnels permettent d'implémenter une méthode de filtrage différente du filtrage par convolution. Dans un premier temps, il est possible d'obtenir une première approximation du résultat du filtrage avec le codage impulsionnel. Dès que plusieurs neurones de la carte SDOG ont émis plus d'un potentiel d'action, le codage fréquentiel permet d'obtenir un résultat proche du filtrage par convolution, qui s'améliore au fur et à mesure des calculs. Cette méthode neuronale de filtrage d'image permet d'obtenir un résultat graduellement : plus le temps processeur alloué à l'algorithme est important, plus le résultat est proche d'une convolution. Ce type de résultat est caractéristique des algorithmes *anytime* (voir la partie 3.3.2) et permet d'obtenir des résultats partiels avant d'avoir traité toute l'image.

4.3 Résultats expérimentaux

Nous verrons tout d'abord les méthodes utilisées pour construire les résultats des images filtrées ainsi que la mesure de distance choisie pour la comparaison d'images. Cette comparaison est effectuée sur plusieurs images de test et met en jeu le codage impulsionnel et fréquentiel. L'étude de paramètres concerne les constantes de relaxation membranaires des neurones de la carte d'entrée ($\tau_{\text{entrée}}$) et de la carte SDOG (τ_{SDOG}) ainsi que la valeur maximale des poids des masques de connexion (w_{max}^D).

4.3.1 Méthodes

Construction des images filtrées

Pour construire les images résultantes du filtrage, nous avons utilisé deux méthodes simples qui permettent de représenter le code fréquentiel et impulsionnel obtenu. Pour le premier, la valeur de niveau de gris l d'un pixel représente la moyenne du taux de décharge normalisé, calculée sur l'intégralité de la simulation :

$$l(x, y) = \frac{\text{nbpa}(x, y)}{\text{MAX}_{x, y}(\text{nbpa})} * D \quad (4.28)$$

avec $\text{nbpa}(x, y)$ le nombre de potentiels d'action émis par le neurone occupant la position spatiale (x, y) sur la carte choisie (par exemple SDOG) et D le nombre de niveaux de gris de l'image. Le code impulsionnel est représenté en associant un numéro d'ordre à chaque premier potentiel d'action émis par chacun des neurones de la carte considérée. Ce numéro d'ordre est déterminé en fonction du pas de temps auquel le potentiel d'action a été émis et est incrémenté à chaque nouveau pas de temps contenant au moins un

potentiel d'action. Ainsi, plusieurs potentiels d'action peuvent partager le même numéro d'ordre. Ce numéro d'ordre est ensuite représenté en niveaux de gris avec :

$$l(x, y) = D - D * \frac{\text{Ordre}(x, y)}{\text{Ordre}_{\max}} \quad (4.29)$$

avec $\text{Ordre}(x, y)$ le numéro d'ordre du neurone (x, y) et Ordre_{\max} le numéro d'ordre le plus élevé pour la carte considérée.

Filtres utilisés

Les résultats présentés dans cette partie sont les résultats d'un filtrage avec un DOG. La taille du filtre DOG utilisé est relativement petite mais est adaptée à la taille des images traitées, comme nous le verrons dans le chapitre 5. La figure 4.8 montre les valeurs du filtre, qui sont normalisées entre $-w_{\max}^D$ et w_{\max}^D . Ces valeurs constituent les poids du masque de connexion.

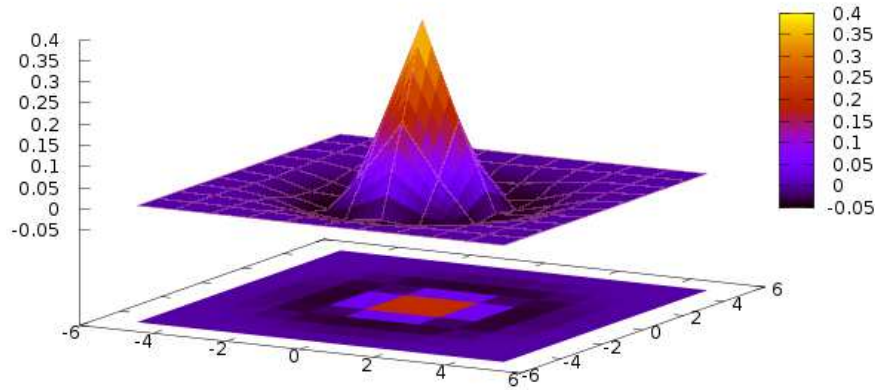


FIGURE 4.8 Filtre différence de gaussiennes (DOG) utilisé, avec $A_1 = 2$, $A_2 = 1$, $\sigma_1 = 1$ et $\sigma_2 = \sqrt{2}$. La taille du filtre est de 11x11 et les valeurs sont normalisées, ici avec $w_{\max}^D = 0.4$.

Mesure d'erreur

Pour caractériser le filtrage neuronal, nous proposons de comparer des images résultantes d'un filtrage neuronal et d'un filtrage par convolution. La comparaison d'image est une tâche complexe, car il n'existe pas de mesure de distance générique satisfaisante ; le choix de la mesure dépend fortement du type de problème à traiter. De plus,

il est difficile de produire une méthode qui permette de traduire la sensation perceptive que l'on a en comparant deux images. Par exemple, la distance entre deux images très semblables “à l'œil nu” peut donner des résultats très différents selon la mesure choisie.

Nous avons choisi d'utiliser l'erreur quadratique moyenne (EQM ou *Mean Squarred Error*) comme mesure de distance.

$$EQM(I_1, I_2) = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (I_1(x, y) - I_2(x, y))^2 \quad (4.30)$$

Cette mesure, sans doute la plus simple, est largement utilisée pour comparer des images. Ceci nous permet de donner des résultats dans une métrique connue et qui facilite les comparaisons, plutôt que de proposer une métrique particulière, conçue pour ce travail. Elle est relativement bien adaptée car les images que nous souhaitons comparer sont alignées : il n'est pas nécessaire de recalcr les deux images avant de mesurer leurs différences. L'EQM possède cependant plusieurs désavantages, car elle est très sensible aux bruits et aux petites perturbations locales. Pour les comparaisons réalisées dans cette partie, les images sont noires sauf pour les régions portant des informations (contrastes d'intensité par exemple). Comme toutes les différences entre images ont la même importance, l'ajout ou la perte d'information (un pixel qui “apparaît” ou “disparaît” d'une image à l'autre) est évalué avec la même métrique qu'une modulation de l'intensité des informations portées (un pixel qui est plus ou moins sombre d'une image à l'autre). Cet aspect peut être problématique car le filtrage neuronal et le filtrage par convolution utilisés ne traitent pas les bords des images de la même façon. Le filtrage par convolution utilisé “étend” l'image autant que nécessaire, en recopiant les pixels qui sont aux bords de l'image, alors que le filtrage neuronal proposé n'utilise aucune méthode particulière pour éviter de produire des artefacts aux bords de l'image. Dans certains cas⁶, nous avons donc omis du calcul de distance les artefacts présents aux bords de l'image.

4.3.2 Comparaison des méthodes de filtrage

Cette partie présente une comparaison expérimentale des méthodes de filtrage neuronal et par convolution. Nous avons réalisé un filtrage par convolution sur différentes images de test, les images résultantes servant de références pour les mesures d'erreur. Nous avons mesuré ces erreurs pour le filtrage neuronal par codage impulsionnel et fréquentiel.

Dans la partie 4.2.3, nous avons formulé l'hypothèse que le codage fréquentiel n'est pas sensible aux approximations réalisées avec le codage impulsionnel. Cette hypothèse

6. Ces cas seront indiqués dans les parties suivantes

sera vérifiée si l'ajout d'un bruit sur les neurones de la carte d'entrée n'améliore pas les résultats du filtrage. Nous présentons donc les résultats obtenus dans cette condition. Nous avons ainsi mesuré pour chaque image de test les différences entre l'image résultant de la convolution par un filtre DOG et trois types de filtrage neuronal : le filtrage utilisant le code impulsionnel, le filtrage utilisant le code fréquentiel et le filtrage utilisant le code fréquentiel avec des entrées bruitées (cette condition est appelée "Stochastique" dans la suite de cette partie). Les paramètres sont détaillés dans les tableaux de la table 4.2 et les simulations ont été réalisées sur 2000 pas de temps, soit 0.2 seconde de temps simulé.

neurones d'entrée		neurones SDOG		masque DOG	
K	1.0	E_l	0	σ_1	1
E_l	-70 mV	V_{reset}	0	σ_2	$\sqrt{2}$
ϑ	-40 mV	ϑ	1	A_1	2
V_{reset}	-80 mV	τ_{SDOG}	200 ms	A_2	1
$\tau_{\text{entrée}}$	200 ms				
τ_n	100 ms				

TABLE 4.2 Paramètres utilisé de la comparaison des méthodes de filtres.

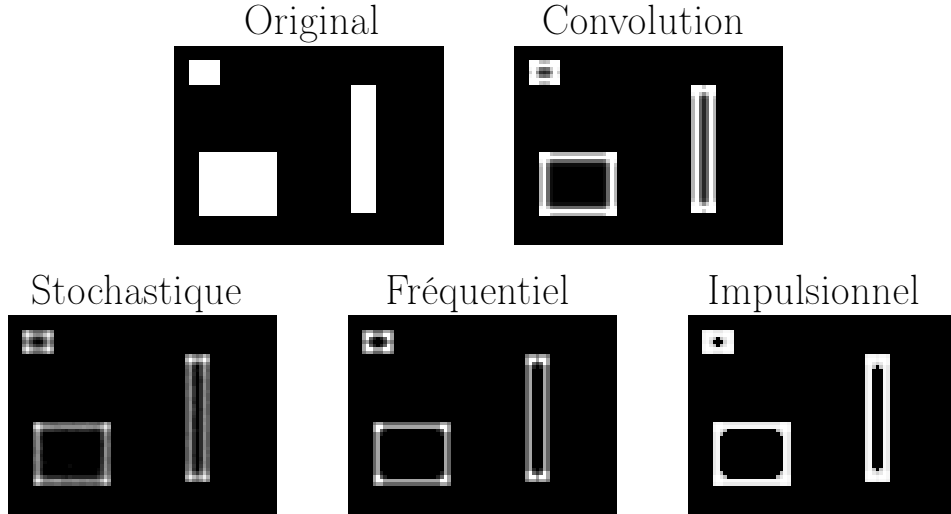


FIGURE 4.9 [Rectangles 1] Le codage fréquentiel (en bas au centre) détecte les contrastes locaux plus finement et donne un résultat plus proche du filtrage par convolution (en haut à droite) que le codage impulsionnel (en bas à droite). L'ajout de bruit dans les entrées pour le codage fréquentiel (la condition "Stochastique" en bas à gauche) influence très peu le résultat du filtrage.

Nous avons utilisé différentes images de test pour les mesures d'erreurs. La première image ("Rectangles 1", figure 4.9) est composée de plusieurs rectangles blancs sur fond noir. La seconde image ("Rectangles 2", figure 4.10) contient trois rectangles blancs et un rectangle noir sur fond gris. L'image "Dégradés" (figure 4.11) est séparée en quatre

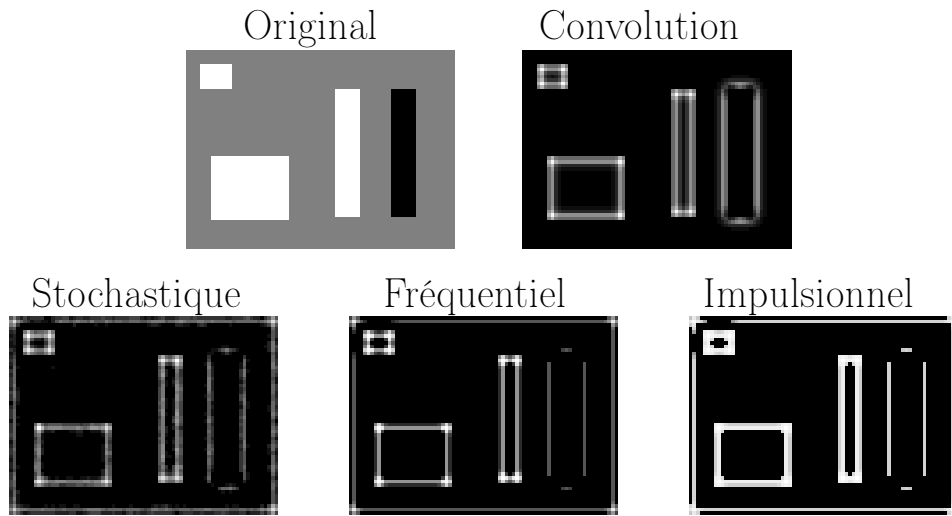


FIGURE 4.10 [Rectangles 2] Par rapport à “Rectangles 1”, on peut constater que le filtrage neuronal accentue l’effet d’atténuation apparaissant sur les coins du rectangle noir. Les résultats obtenus avec le filtrage neuronal sur les rectangles clairs sont similaires à ceux obtenus avec le filtrage par convolution.

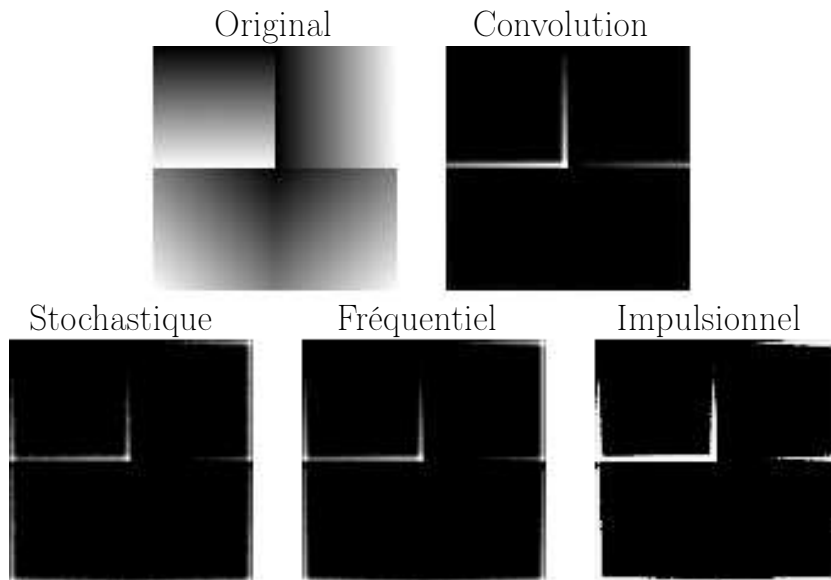


FIGURE 4.11 [Dégradés] Le filtrage neuronal, impulsionnel ou fréquentiel, se comporte comme le filtrage par convolution pour traiter les dégradés.

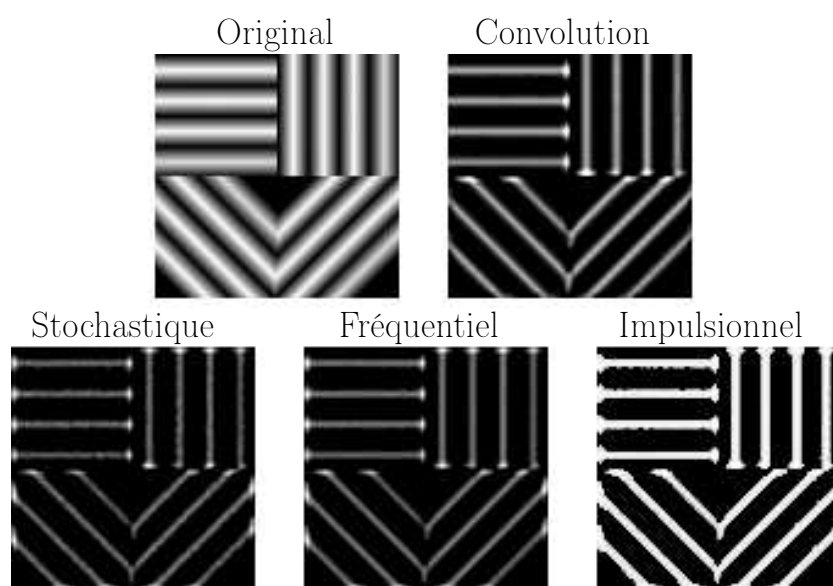


FIGURE 4.12 [Stries dégradées] Le codage impulsionnel offre une approximation du filtrage par convolution. Le codage fréquentiel donne des résultats très proches du filtrage par convolution.

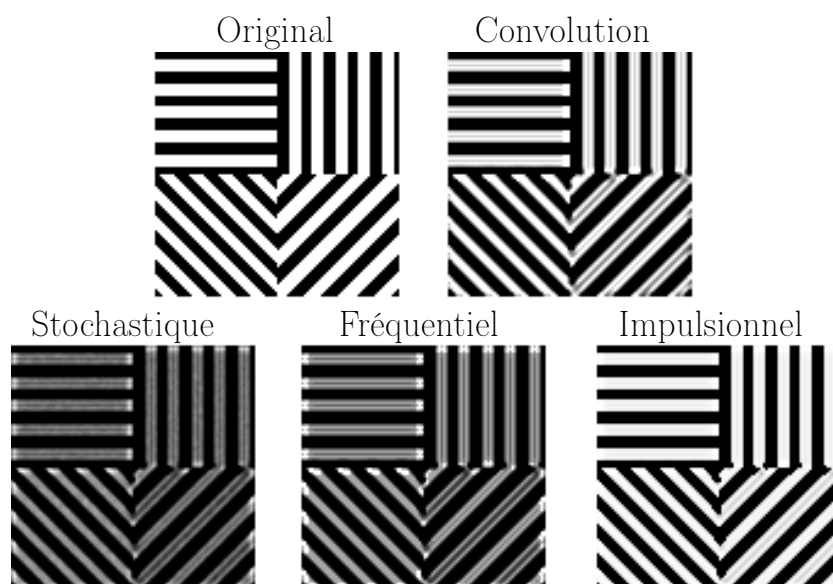


FIGURE 4.13 [Stries 1] Le codage fréquentiel, avec ou sans bruit additif, donne des résultats dont les variations sont semblables à celle observées avec le filtrage par convolution, bien qu'il présente un biais pour les fins de lignes.

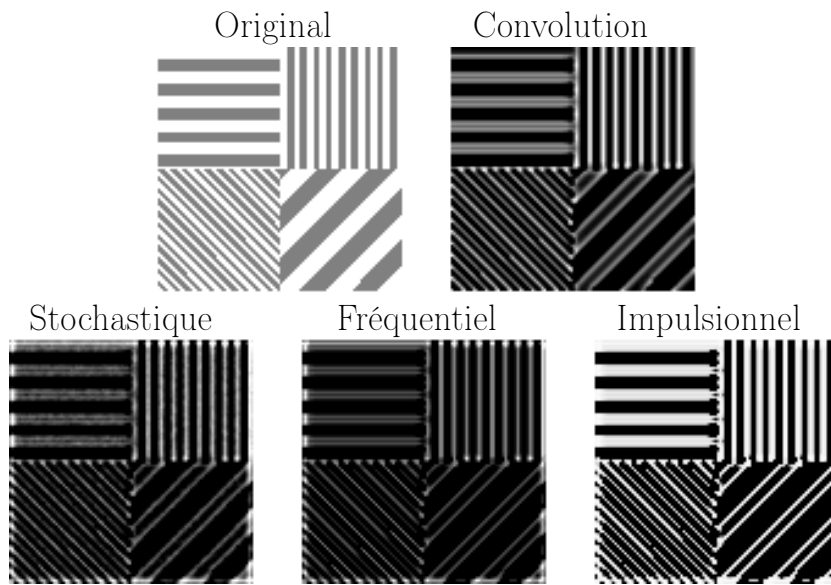


FIGURE 4.14 [Stries 2] Le filtrage neuronal permet d’obtenir un filtrage proche du filtrage par convolution lorsqu’il faut traiter des bandes contrastées de fréquence spatiale différentes.

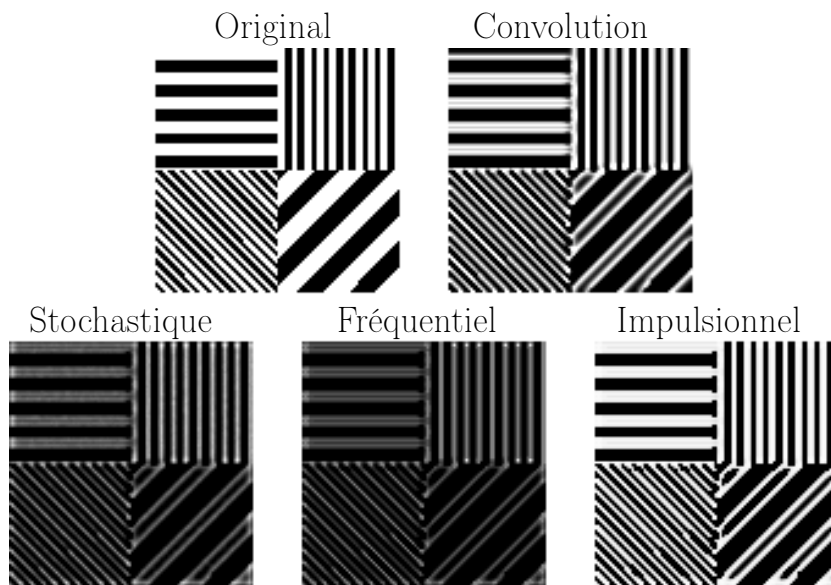


FIGURE 4.15 [Stries 3] Le contraste est plus prononcé que dans la figure “Stries 2” et le filtrage neuronal reproduit les variations fine du filtrage avec un biais pour les terminaisons de lignes. L’ajout de bruit pour le codage fréquentiel ne change pas la qualité des résultats.

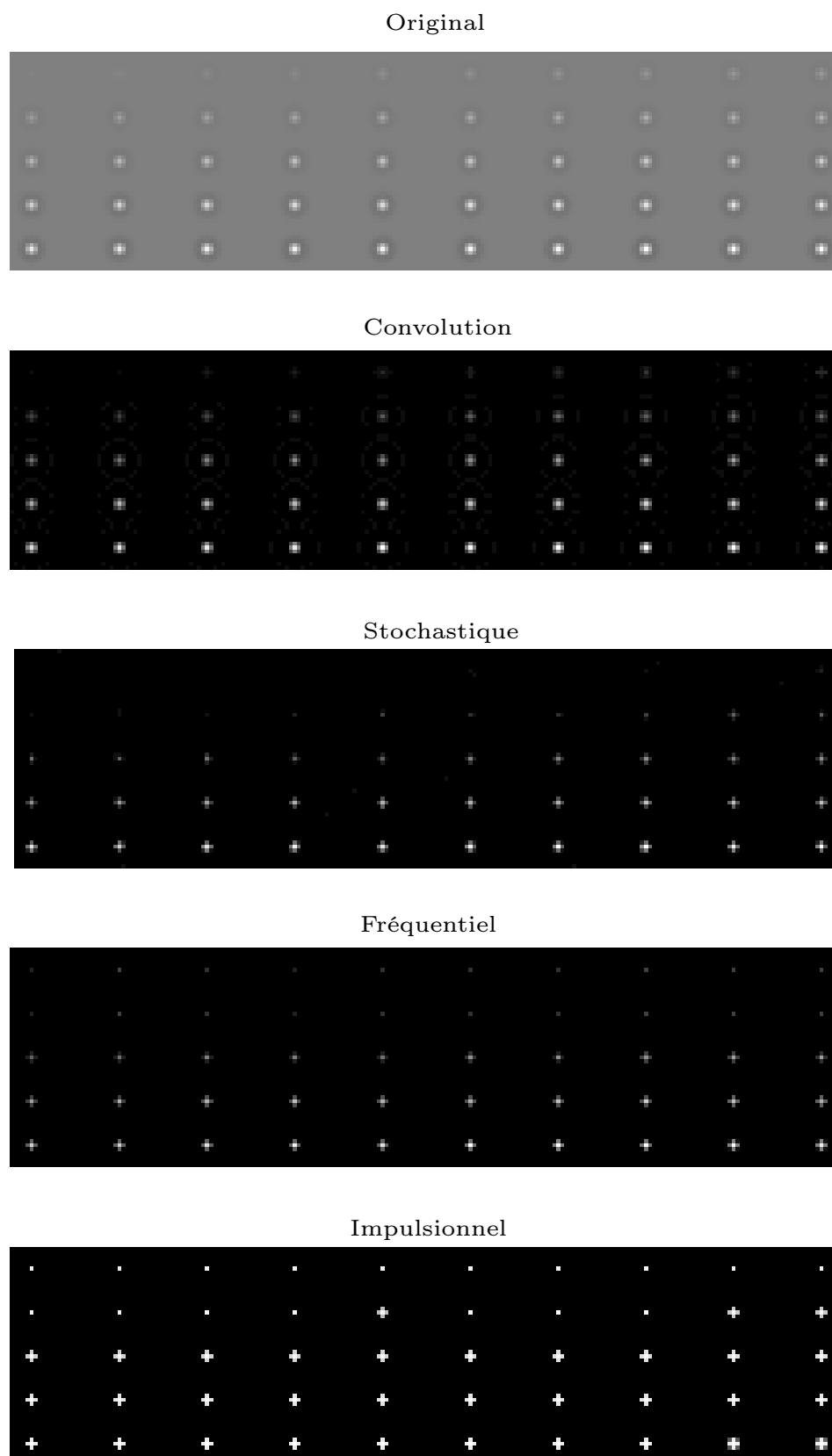


FIGURE 4.16 [Filtres 1] Le codage impulsionnel donne une approximation relativement correcte du filtrage. Les résultats du codage fréquentiel sont très proches des résultats du filtrage par convolution. L'ajout du bruit n'améliore pas les résultats.

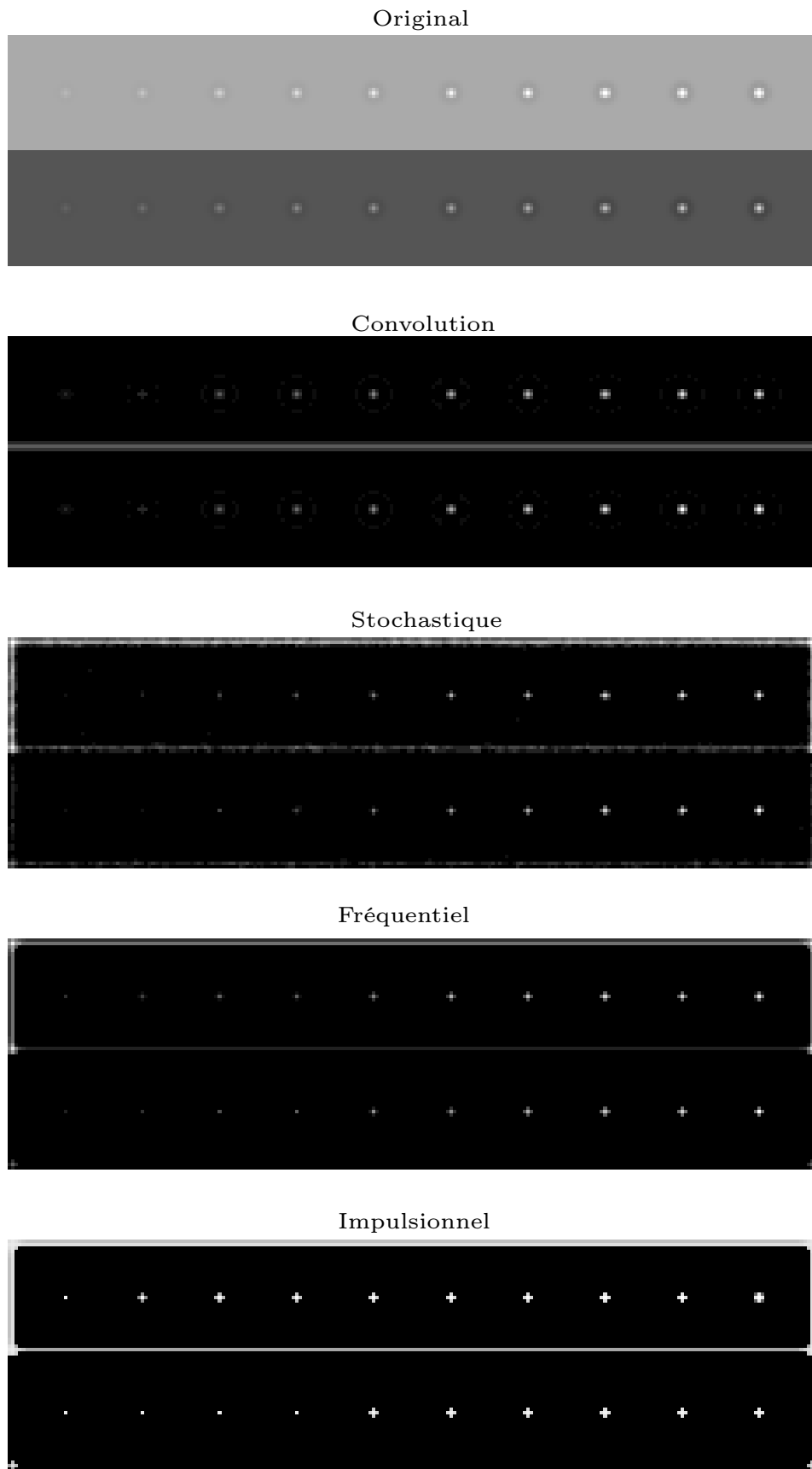


FIGURE 4.17 [Filtres 2] Le codage impulsionnel présente une sensibilité à la luminance puisque les réponses sont différentes entre la partie claire et foncée de l'image. Cet effet affecte beaucoup moins le codage fréquentiel.

parties, chacune contenant un dégradé de gris orienté dans un sens différent. L'image "Stries dégradées" de la figure 4.12 représente des barres orientées, dont les bords sont en dégradé de gris. L'image suivante, "Stries 1" (figure 4.13), est composée de barres orientées blanches sur fond noir. La figure 4.14 montre l'image "Stries 2" qui est composée de barres orientées blanches de différentes largeurs sur fond gris. L'image "Stries 3" est identique mais sur fond noir.

L'image de la figure 4.16, "Filtre 1", est construite en appliquant des patches DOG d'amplitude croissante linéairement de gauche à droite et de haut en bas, sur un fond noir. L'image "Filtres 2" de la figure 4.17 est divisée en deux parties, l'une de fond gris clair et l'autre de fond gris sombre. Sur chaque partie est appliquée la même série de patches différence de gaussiennes d'amplitude croissante linéairement. Les patches de plus grande amplitude (à droite) ne sont pas tronqués, en ce sens que leurs valeurs maximales et minimales ne dépassent pas le nombre de niveau de gris.

Chaque figure montre l'image originale (en haut à gauche), le résultat du filtrage par convolution (en haut à droite) et les résultats du filtrage neuronal en bas, avec la condition stochastique à gauche, le codage fréquentiel au centre et le codage impulsionnel à droite.

Les erreurs mesurées pour chaque image sont récapitulées sur l'histogramme de la figure 4.18 et les valeurs d'EQM sont présentées dans le tableau 4.3. Il est à noter que pour les images "Rectangles 2" et "Stries dégradées", nous n'avons pas pris en compte les bords de l'image dans la mesure d'erreur.

Image	Taille	Code	EQM
Rectangles 1	76x56	Bruit + Fréquentiel	1008.76
		Fréquentiel	937.04
		Impulsionnel	389.417
Rectangles 2	76x56	Bruit + Fréquentiel	255.134
		Fréquentiel	512.701
		Impulsionnel	1185.61
Dégradés	100x100	Bruit + Fréquentiel	725.209
		Fréquentiel	854.605
		Impulsionnel	4266.34
		Bruit + Fréquentiel	995.575
../..			

suite de la page précédente			
Image	Taille	Code	EQM
Stries dégradées	100x100	Fréquentiel	805.85
		Impulsionnel	3070.38
Stries 1	100x100	Bruit + Fréquentiel	5969.08
		Fréquentiel	3820.22
		Impulsionnel	927.216
Stries 2	100x100	Bruit + Fréquentiel	1898.49
		Fréquentiel	2770.63
		Impulsionnel	5722.31
Stries 3	100x100	Bruit + Fréquentiel	6860.99
		Fréquentiel	9503.12
		Impulsionnel	1939.52
Filtres 1	1111x33	Bruit + Fréquentiel	726.323
		Fréquentiel	282.341
		Impulsionnel	2853.72
Filtres 2	231x66	Bruit + Fréquentiel	792.123
		Fréquentiel	445.732
		Impulsionnel	1939.04

TABLE 4.3 Valeur de l'EQM par rapport à un filtrage par convolution calculé pour chaque type de filtrage neuronal.

L'histogramme 4.18 montre que les erreurs sur le codage impulsionnel sont plus importantes que pour les conditions considérées du codage fréquentiel ("Stochastique" et "Fréquentiel" sur la figure), ce qui est le résultat attendu. Cependant pour "Stries 1" et "Stries 3", l'erreur sur le codage impulsionnel est plus faible que pour le codage fréquentiel, ce qui semble principalement imputable à la méthode utilisée pour construire les images résultantes. En effet, les bords des stries ont des réponses fortes, ce qui ne laisse que des valeurs de niveau de gris plus sombre pour représenter le corps des stries. Ainsi pour ces deux images, même si la réponse du filtrage par convolution donne des valeurs globalement élevées qui sont plus proches du filtrage par codage impulsionnel, on peut constater que les variations fines (par exemple la réponse du filtrage pour l'intérieur des stries) sont mieux retranscrites par le codage fréquentiel que par le codage impulsionnel.

Le codage fréquentiel permet donc de mieux approcher le filtrage par convolution que le codage impulsionnel. Ceci est confirmé car l'hypothèse formulée dans la partie 4.2.3 est vérifiée : les deux conditions étudiées pour le codage fréquentiel donnent des résultats

comparables. C'est le résultat attendu et ceci indique que le codage fréquentiel permet de réaliser un filtrage plus proche du filtrage par convolution.

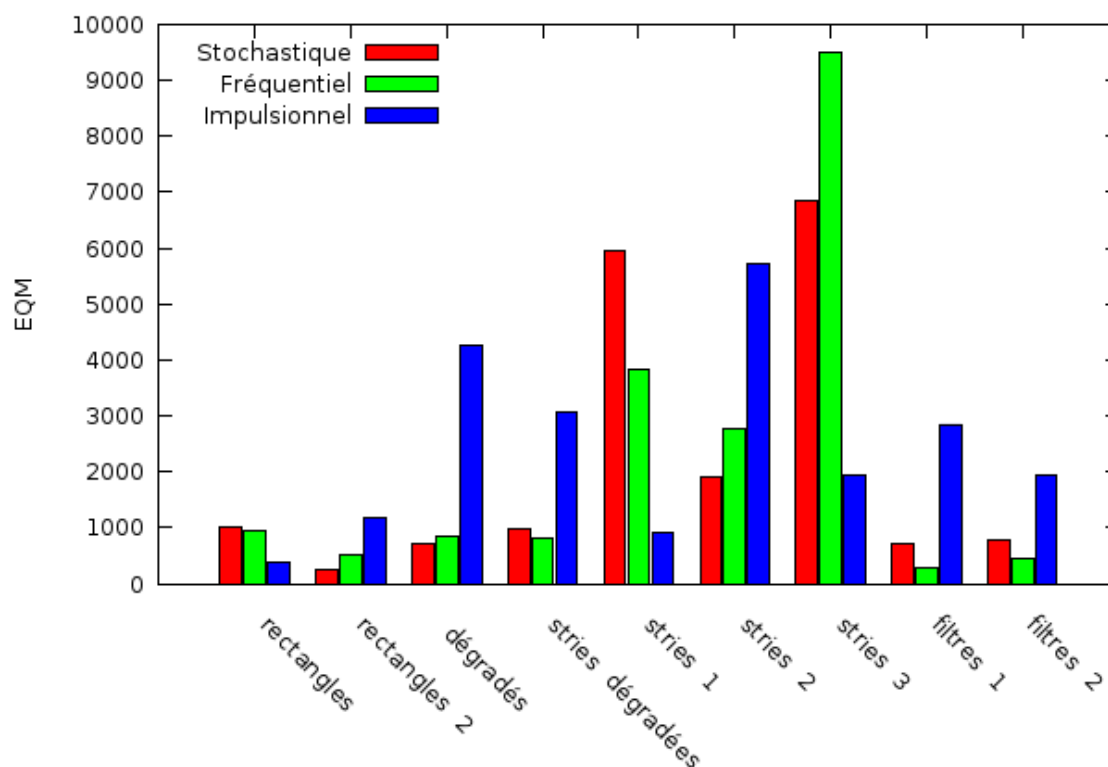


FIGURE 4.18 EQM mesurées entre les images résultantes du filtrage neuronal et du filtrage par convolution. La condition “Stochastique” correspond à la mesure du codage fréquentiel lorsqu’un bruit additif est ajouté aux neurones de la carte d’entrée, le codage “Fréquentiel” consiste à mesurer le nombre moyen de potentiels d’action émis par chaque neurone pour déterminer le résultat du filtrage et le codage “Impulsionnel” réalise un même filtrage en ne mesurant que le temps d’émission du premier potentiel d’action de chaque neurone.

On peut constater que sur les images “Rectangles 1”, “Rectangles 2”, “Stries dégradées”, “Stries 1”, “Stries 2”, “Stries 3”, “Filtres 1” et “Filtres 2” le codage fréquentiel permet d’obtenir une réponse dont les variations locales sont plus fines et reflètent mieux les variations qui apparaissent sur le filtrage par convolution. Les erreurs mesurées pour les codages fréquentsiels de “Stries 1” et “Stries 3” sont très importantes car l’EQM amplifie largement le fait que les valeurs absolues de luminance des pixels diffèrent entre le codage fréquentiel et le filtrage par convolution. Ainsi, même si les variations locales sont mieux exprimées dans le codage fréquentiel, les erreurs mesurées sont très importantes. L’EQM ne permet donc pas de refléter la précision des variations locales du codage fréquentiel.

Concernant les remarques spécifiques à chaque image, pour “Rectangle 2” (figure 4.10), le rectangle noir sur fond blanc est moins bien détecté avec le filtrage neuronal

qu’avec le filtrage par convolution. Ceci peut être dû à un paramétrage des neurones inadéquat qui est amplifié par la sensibilité aux valeurs absolues des luminances du filtrage neuronal. L’image “Dégradés” (figure 4.11) montre que l’opération de filtrage réalisée se comporte bien pour traiter des variations fines de luminance, aussi bien avec le codage impulsionnel qu’avec le codage fréquentiel. Les images “Stries 2” et “Stries 3” (figure 4.14 et 4.15) montrent que le filtrage neuronal reproduit bien les variations fines de la réponse du filtre pour des bandes (ou *gratings*) de fréquences spatiales différentes.

L’image “Filtres 1” (figure 4.16) est composé de patches différence de gaussiennes d’amplitudes linéairement croissantes. Ces patches sont identiques à la réponse impulsionnelle du filtre utilisé. Le filtrage de cette image doit produire des réponses impulsionnelles dont l’intensité est fonction de l’amplitude des patches filtrés. Les patches et les filtres utilisés ne sont pas équilibrés, ceci est en partie dû à la discrétisation réalisée pour traduire ces filtres sous forme matricielle. Le codage fréquentiel permet d’approcher la réponse du filtrage par convolution. Pour le codage impulsionnel, les variations que l’on observe pour des patches de petites amplitudes illustrent le caractère non-linéaire de ce filtrage. L’étude proposée dans la partie 4.2.2 pour caractériser le codage impulsionnel permet d’expliquer la décroissance brutale de la réponse pour les faibles amplitudes. L’image “Filtres 2” permet de vérifier la sensibilité à la luminance du filtrage neuronal, en effet si le codage fréquentiel permet de restituer une réponse proche du filtrage par convolution, le codage impulsionnel est lui beaucoup plus sensible à la luminance moyenne.

Les images “Filtres 1” et “Filtres 2” mettent aussi en évidence un aspect intéressant du codage impulsionnel : les patches dont l’amplitude est la plus grande suscitent une réponse précoce, tandis que les patches de plus faible amplitude apparaissent plus tardivement. Cet effet est mis en évidence sur la figure 4.19. On peut constater que l’approximation réalisée par le codage impulsionnel est biaisée en faveur des valeurs de luminance élevées, car les patches sur fond sombre sont détectés plus tardivement que les patches sur fond clair alors qu’ils ont la même amplitude. Il apparaît néanmoins que les premiers potentiels d’action émis, aux erreurs d’approximations près, codent pour les coefficients les plus importants. Cet effet est particulièrement intéressant lorsque l’on couple plusieurs cartes de neurones, les unes à la suite des autres : comme les coefficients les plus importants sont propagés les premiers, il est ainsi possible d’obtenir de façon précoce le résultat de plusieurs opérations de filtrage successives.

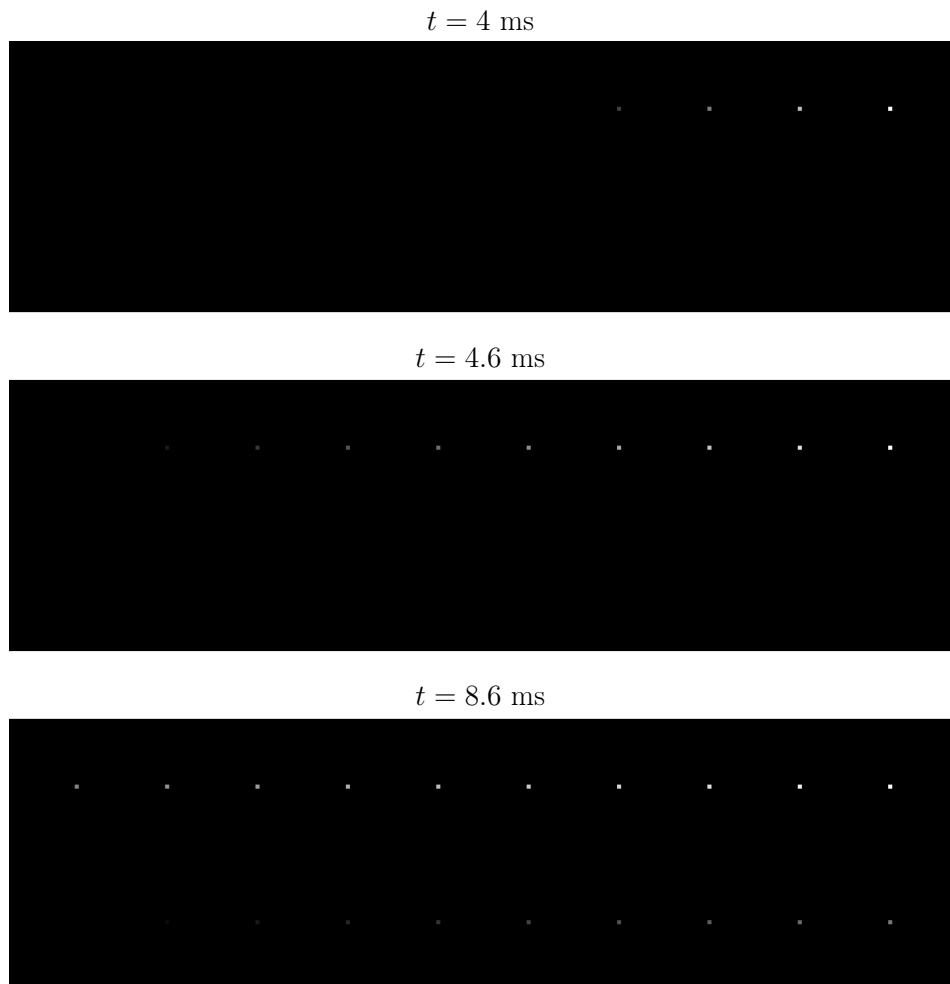


FIGURE 4.19 Mise en évidence de l'ordre d'apparition des résultats du filtrage pour l'image "Filtres 2" avec le codage impulsionnel. Les premiers potentiels d'action émis codent pour les patchs de plus grande amplitude sur fond clair (en haut à gauche). Les patchs sur fond foncé sont détectés plus tardivement.

4.3.3 Études de paramètres

Cette partie propose d’explorer l’influence de différents paramètres du filtrage neuronal : les constantes de relaxation membranaires des neurones de la carte d’entrée $\tau_{\text{entrée}}$ et de la carte SDOG τ_{SDOG} et des valeurs maximum du poids du masque de connexion w_{max}^D . Nous avons choisi d’étudier l’incidence de ces paramètres sur l’image “Filtres 2” (figure 4.17), car elle permet d’une part de mesurer les différences entre le filtrage neuronal et le filtrage par convolution et d’autre part de comparer le filtrage neuronal sur des luminance basses et élevées. Nous distinguons donc dans cette partie deux mesures d’erreur : la première est celle utilisée précédemment, renommée EQM_{ref} . La seconde mesure les différences entre la partie haute, où les patches sont appliqués sur un fond gris clair, et basse de l’image, où les patches sont appliqués sur un fond gris foncé. Ceci permet d’évaluer la sensibilité du filtrage aux valeurs absolues de luminance. Cette mesure est calculée avec :

$$\text{EQM}_{\text{auto}}(I) = \frac{1}{MO} \sum_{x=1}^M \sum_{y=1}^O (I(x, y) - I(x, y + O))^2, \text{ avec } O = N/2 \quad (4.31)$$

Il est à noter que les mesures d’erreur réalisées dans cette partie ne tiennent pas compte des bords de l’image, ce qui permet d’éviter que les artefacts présents aux bords de l’image ne perturbent les mesures.

Nous avons aussi mesuré le nombre moyen N_{PA} de potentiels d’action intégrés par les neurones de la carte SDOG avant qu’ils ne déchargent. Ceci permet d’estimer la quantité d’information sur laquelle le neurone “décide” de décharger. Cette valeur est calculée pour les P neurones d’une carte et pour les S potentiels d’actions émis par chaque neurone avec :

$$N_{\text{PA}} = \frac{1}{PS} \sum_{p \in P} \sum_{k \in S} n_{p,k} \quad (4.32)$$

où $n_{p,k}$ est le nombre de potentiels d’action reçus par le neurone p pendant le k -ème intervalle inter-spike⁷. Les résultats présentés précisent également le nombre moyen de potentiels d’action excitateurs et inhibiteurs reçus, qui sont calculés de façon similaire.

Étude de $\tau_{\text{entrée}}$

Les résultats de la figure 4.20 ont été obtenus avec deux cartes neuronales : une carte d’entrée et une carte SDOG, connectée à la carte d’entrée par un masque SDOG. Les neurones de la carte d’entrée sont soumis à un courant d’entrée propor-

7. intervalle temporel séparant l’émission de deux potentiels d’actions consécutifs

tionnel à la luminance des pixels de l'image, avec $E_l = -70\text{mV}$, $\vartheta = -40.0\text{mV}$, $V_{\text{reset}} = -80\text{mV}$ et $K = 1$. Les simulations présentent les résultats obtenus pour $\tau_{\text{entrée}} \in \{10, 20, 40, 100, 200, 1000\}\text{ms}$. Les poids du masque de connexion correspondent à un filtre différence de gaussiennes de variances $\sigma_1 = 1$ et $\sigma_2 = \sqrt{2}$ et de coefficients $A_1 = 2$ et $A_2 = 1$, de taille 11×11 et dont les valeurs sont normalisées entre -0.4 et 0.4 (voir la figure 4.8). Les neurones de la carte SDOG sont des intégrateurs à fuite, avec $\tau_{\text{SDOG}} = 200\text{ms}$, $E_l = V_{\text{reset}} = 0$ et $\vartheta = 1$, et leurs synapses modélisent des interactions instantanées (voir équation (4.7)). Les simulations sont réalisées sur 2000 pas de temps, soit 0.2s de temps simulé.

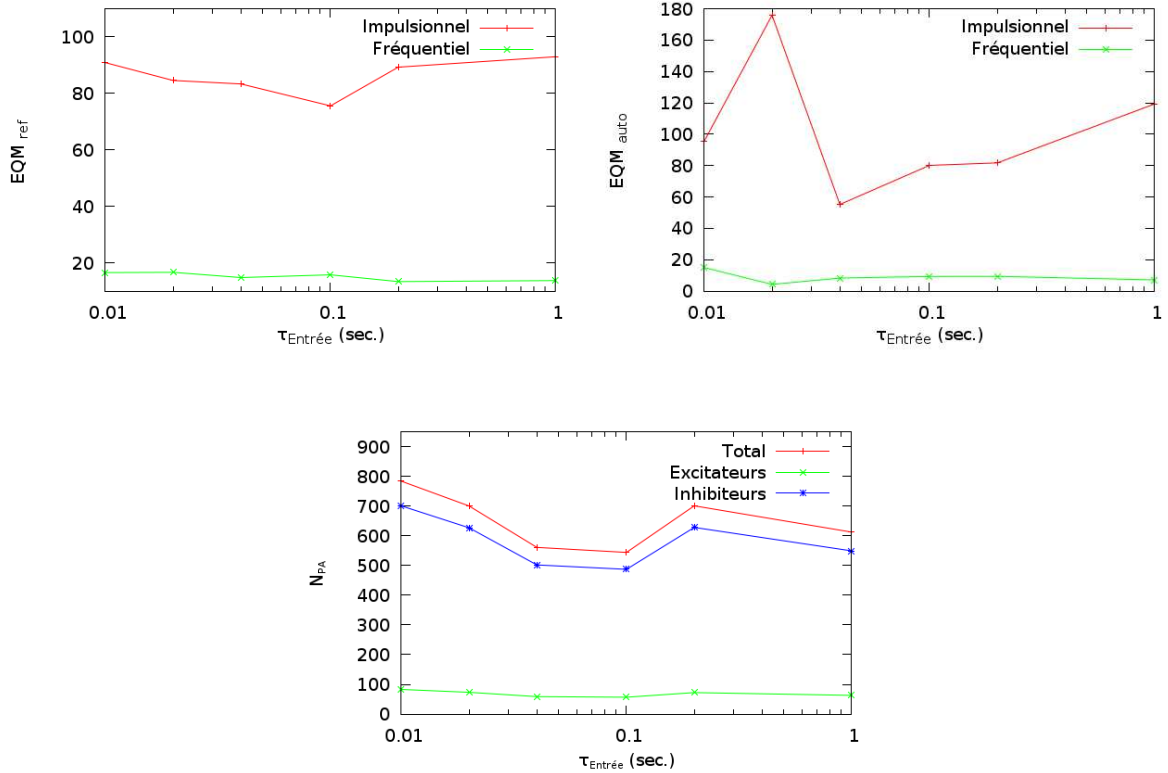


FIGURE 4.20 Influence de $\tau_{\text{entrée}}$ sur le filtrage neuronal, pour le codage impulsionnel (en rouge) et le codage fréquentiel (en vert). À gauche : EQM mesurée entre le filtrage neuronal et le filtrage par convolution pour différentes valeurs de $\tau_{\text{entrée}}$. À droite : EQM mesurée entre la partie supérieure (patches sur fond clair) et inférieure (patches sur fond sombre) de l'image. En bas : nombre moyen de potentiels d'action N_{PA} intégrés avant que les neurones de la carte SDOG ne déchargent.

On peut voir sur la figure 4.20 que le paramètre $\tau_{\text{entrée}}$ a une influence restreinte sur le filtrage : l'EQM_{ref} est relativement constante. Concernant le nombre moyen de potentiels d'action intégrés par les neurones de la carte SDOG, on pourra noter que les potentiels d'action inhibiteurs reçus sont largement plus nombreux que les potentiels d'action excitateurs, car la surface du champ récepteur différence de gaussiennes est

majoritairement négative. Le nombre moyen de potentiels d'action reçus décroît pour $0.01 < \tau_{\text{entrée}} \leq 0.1$, il augmente pour $\tau_{\text{entrée}} = 0.2$ puis décroît à nouveau pour $0.2 < \tau_{\text{entrée}} \leq 1$. L'augmentation de ces valeurs pour $\tau_{\text{entrée}} = 0.2$ est un artefact induit par l'image utilisée. La couleur de fond de la partie basse de l'image d'entrée est un gris sombre dont la valeur de luminance est trop faible pour susciter l'émission de potentiels d'action par les neurones de la carte d'entrée lorsque $\tau_{\text{entrée}} < 0.2$ (voir la figure 4.3). Pour $\tau_{\text{entrée}} \geq 0.2$, Les neurones associés à ces pixels sombres génèrent des potentiels d'action, ce qui augmente significativement le nombre moyen de potentiels d'action intégrés par les neurones de la carte SDOG. Ceci illustre le fait qu'il est possible d'utiliser $\tau_{\text{entrée}}$ pour réaliser un filtre passe-haut sur les luminances, c'est-à-dire que les neurones de la carte d'entrée réalisent un seuillage implicite car les valeurs de luminance basses ne sont pas suffisantes pour les faire décharger.

On peut constater sur la figure 4.21 que le résultat du filtrage varie effectivement peu en fonction de $\tau_{\text{entrée}}$. D'autre part, cette figure illustre les variations d' EQM_{auto} mesurée sur la figure 4.20. Il apparaît que le codage fréquentiel, qui est assez peu sensible à la luminance, n'est pas particulièrement influencé par $\tau_{\text{entrée}}$. Pour le codage impulsionnel, qui donne une réponse biaisée en faveur des luminances élevées, les différences restent peu importantes.

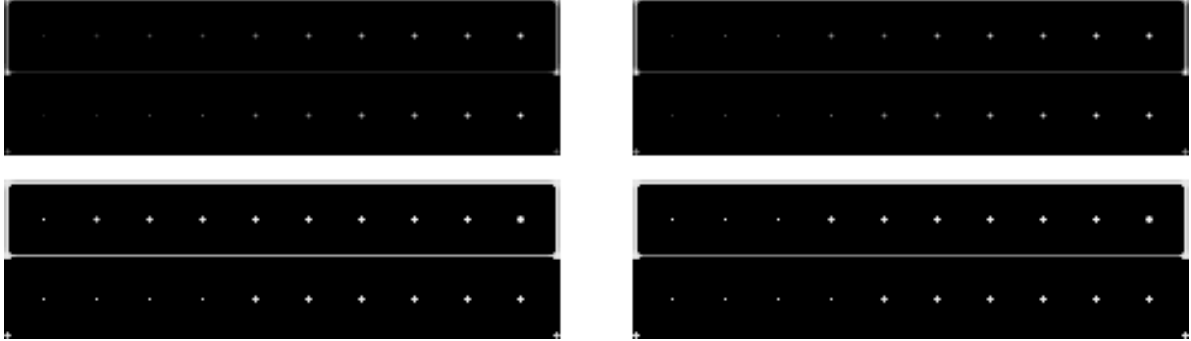


FIGURE 4.21 Résultats du filtrage pour $\tau_{\text{entrée}} = 0.02\text{s}$, à gauche, et pour $\tau_{\text{entrée}} = 0.1\text{s}$, à droite. Les images représentant le codage fréquentiel sont en haut et celles représentant le codage impulsionnel sont en bas.

Étude de τ_{SDOG}

La figure 4.22 montre les résultats obtenus pour différentes valeurs de τ_{SDOG} . Les mesures ont été réalisées pour $\tau_{\text{entrée}} \in \{0.02, 0.1\}$. Ce paramètre a peu d'influence sur l' EQM_{auto} , sauf pour des valeurs de τ_{SDOG} très élevées. Il apparaît que pour les valeurs de τ_{SDOG} inférieures à 0.01, les EQM_{ref} mesurées sont sensiblement identiques pour les codages impulsionnel et fréquentiel. Lorsque $\tau_{\text{SDOG}} > 0.01$, l' EQM_{ref} mesurée pour

le codage impulsionnel croît alors que l' EQM_{ref} mesurée pour le codage fréquentiel diminue.

En effet, lorsque τ_{SDOG} est élevée les neurones se rapprochent alors de simples accumulateurs dont l'évolution du potentiel membranaire⁸ (voir l'équation (4.8)) devient :

$$V_j(t) \approx \sum_i \sum_f w_{ij} H(t, \hat{t}_i^f) \quad (4.33)$$

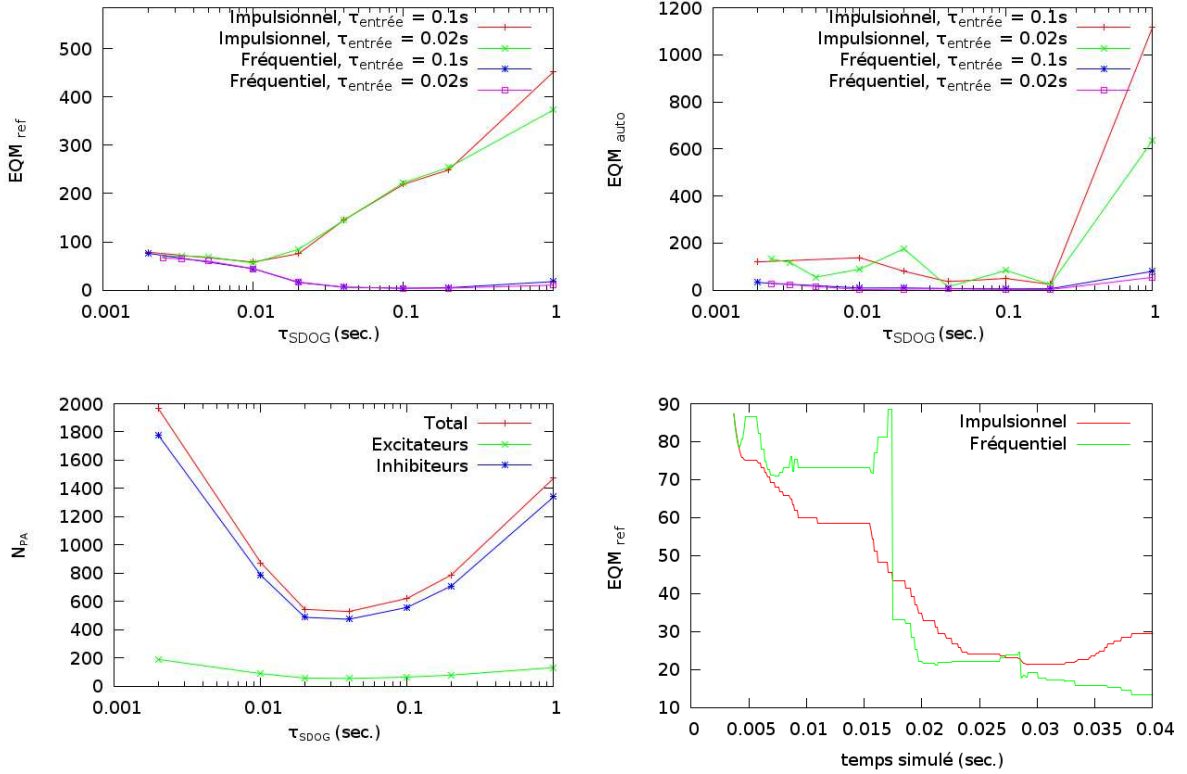


FIGURE 4.22 Influence de τ_{SDOG} sur le filtrage neuronal. En haut à gauche : EQM_{ref} mesurée entre le filtrage neuronal et le filtrage par convolution en fonction de τ_{SDOG} , pour $\tau_{\text{entrée}} \in \{0.02, 0.1\}$. En haut à droite : EQM_{auto} mesurée entre la partie supérieure (patches sur fond clair) et inférieure (patches sur fond sombre) de l'image, pour $\tau_{\text{entrée}} \in \{0.02, 0.1\}$. En bas à gauche : nombre moyen de potentiels d'action N_{PA} intégrés avant que les neurones de la carte SDOG déchargent. En bas à droite : évolution de l' EQM_{ref} pour le codage impulsionnel et fréquentiel en fonction du temps simulé, pour $\tau_{\text{SDOG}} = 0.1\text{s}$.

Les résultats de la figure 4.22 ont été obtenus en utilisant les mêmes valeurs que celles de la partie précédente et en fixant $\tau_{\text{entrée}}$ à 20 et 100ms pour tester les valeurs de $\tau_{\text{SDOG}} \in \{2, 2.5, 3, 5, 10, 20, 40, 100, 200, 1000\}\text{ms}$.

Les résultats présentés en bas à droite de la figure 4.23 montrent une augmentation

8. comme dans la partie 4.2.2, on considère un neurone j de la carte SDOG connecté à des neurones de la carte d'entrée, d'indices i .

de l' EQM_{ref} du codage impulsionnel à partir de 30ms de temps simulé. Ceci vient du fait que les neurones en périphérie des patchs se mettent à décharger car le filtre utilisé n'est pas équilibré (voir la figure 4.22 en bas à droite). Le codage impulsionnel permet cependant d'obtenir rapidement un résultat assez approché du filtrage par convolution. À partir de 17ms de temps simulé, le codage fréquentiel donne un résultat très proche du filtrage par convolution.

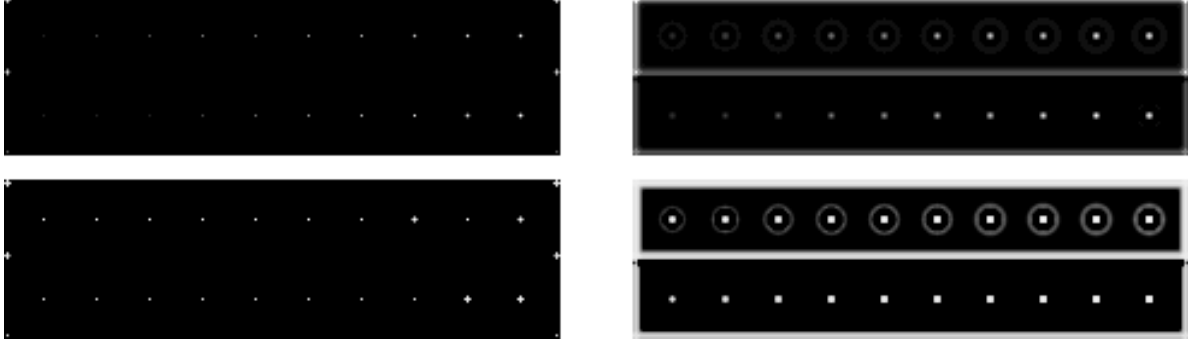


FIGURE 4.23 Résultats du filtrage avec $\tau_{\text{SDOG}} = 0.01\text{s}$ à gauche et $\tau_{\text{SDOG}} = 1\text{s}$ à droite. Les images représentant le codage fréquentiel sont en haut et celles représentant le codage impulsionnel sont en bas. Ces résultats utilisent $\tau_{\text{entrée}} = 0.1\text{s}$.

Étude de w_{max}^D

La figure 4.24 présente les résultats obtenus pour les valeurs de w_{max}^D comprises entre 0.2 et 0.5 et en utilisant les mêmes valeurs de paramètres que précédemment (avec $\tau_{\text{entrée}} = 100\text{ms}$ et $\tau_{\text{SDOG}} = 200\text{ms}$).

On peut constater que les EQM_{ref} sont identiques pour le codage impulsionnel et fréquentiel pour $w_{\text{max}}^D < 0.3$. Pour des valeurs supérieures, les valeurs d' EQM_{ref} divergent : celles du codage impulsionnel se dégradent alors que celles du codage fréquentiel s'améliorent.

L'influence de w_{max}^D est relativement similaire à celle de τ_{SDOG} , puisque des valeurs élevées de τ_{SDOG} tendent à amplifier l'importance du poids des potentiels d'action reçus. L'impact négatif d'un poids élevé sur le codage impulsionnel est dû au même effet que précédemment : les neurones situés sur les patchs déchargent tous dans un laps de temps de plus en plus court (voir la figure 4.25). Le codage fréquentiel est lui affecté positivement par l'augmentation de w_{max}^D , car les neurones déchargent après avoir intégré peu de potentiels d'action et comme l'ordre des potentiels d'action qu'ils reçoivent varie largement, leur fréquence moyenne de décharge tend à approcher correctement un filtrage par convolution.

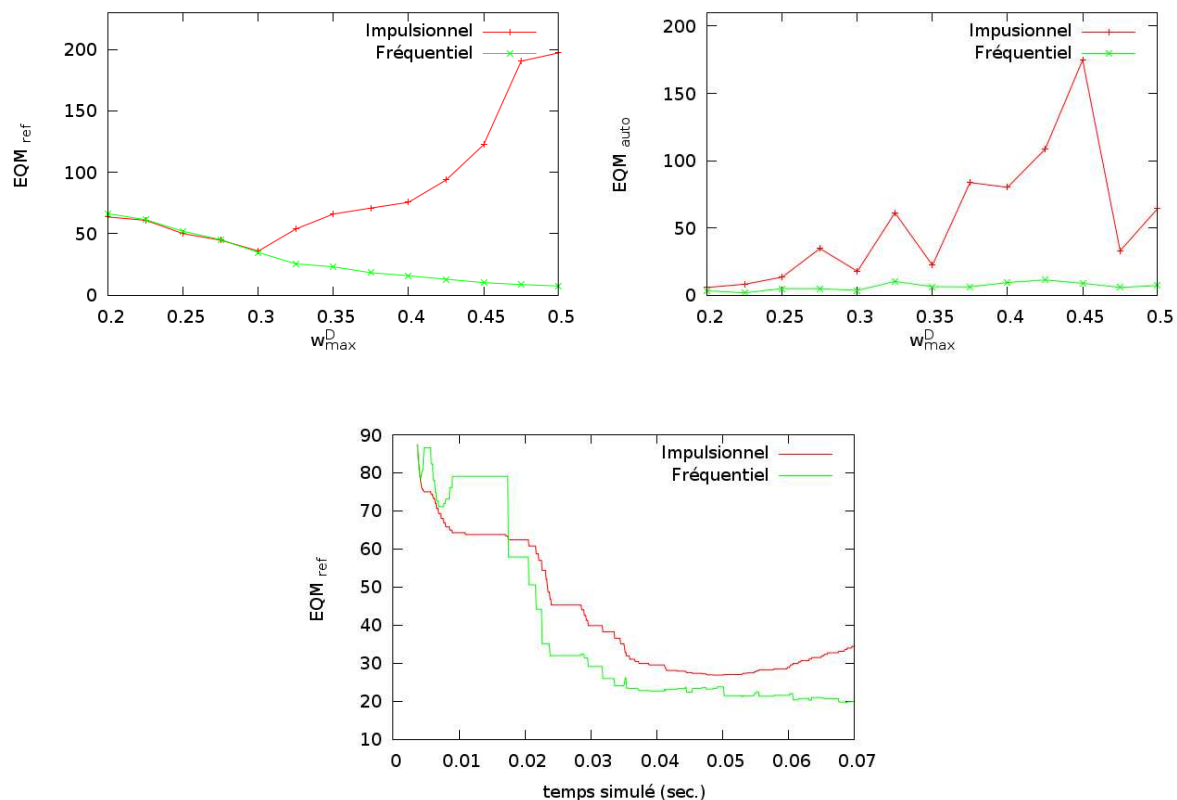


FIGURE 4.24 Influence de w_{max}^D sur le filtrage neuronal, pour le codage impulsionnel (en rouge) et le codage fréquentiel (en vert). En haut à gauche : EQM_{ref} mesurée entre le filtrage neuronal et le filtrage par convolution en fonction de w_{max}^D . En haut à droite : EQM_{auto} mesurée entre la partie supérieure (patchs sur fond clair) et inférieure (patchs sur fond sombre) de l'image. En bas : évolution de l' EQM_{ref} pour le codage impulsionnel et fréquentiel en fonction du temps simulé, pour $w_{max}^D = 0.4$.

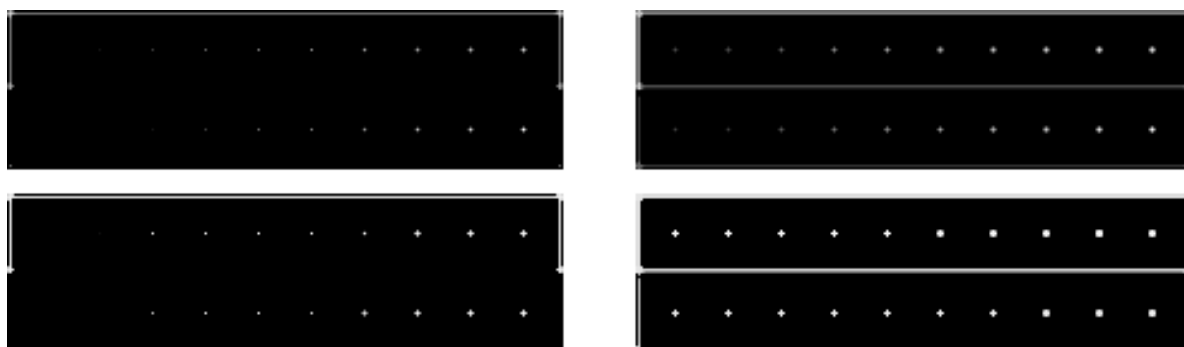


FIGURE 4.25 Résultats du filtrage avec $w_{max}^D = 0.3$ à gauche et $w_{max}^D = 0.5$ à droite. Les images représentant le codage fréquentiel sont en haut et celles représentant le codage impulsionnel sont en bas.

4.4 Conclusion

Dans ce chapitre, nous avons proposé une méthode originale de filtrage neuronal, comportant une étape de transduction et une étape d'intégration. La transduction permet de convertir les valeurs des pixels en code temporel : chaque neurone de la carte d'entrée émet des trains de potentiels d'action dont la fréquence est fonction de la valeur considérée du pixel. L'intégration est réalisée par une seconde carte de neurones, générant des trains de potentiels d'action qui sont interprétés pour obtenir le résultat du filtrage de l'image. L'information portée par ces potentiels d'action peut être décodée selon que l'on s'intéresse aux premiers potentiels d'action de chaque neurone (codage impulsionnel) ou à la fréquence moyenne de décharge des neurones (codage fréquentiel).

Le codage impulsionnel implémente une approximation du filtrage par convolution, qui permet d'obtenir rapidement un résultat partiel. Ce codage est sensible aux luminances et susceptible de produire des résultats inexacts, car le neurone peut décharger avant d'avoir traité toutes les informations disponibles. Le codage fréquentiel permet de passer outre ce problème et il est beaucoup moins dépendant des valeurs absolues des luminances. Cependant, il n'est disponible que plus tardivement car il nécessite un plus grand nombre de calculs. Cette méthode implémente un processus *anytime* : d'une part il est possible d'utiliser les résultats du codage impulsionnel comme première approximation avant d'utiliser le codage fréquentiel pour obtenir des résultats plus précis. D'autre part, le codage fréquentiel offre des résultats d'autant plus proches du filtrage par convolution que le nombre de potentiels d'action considérés pour calculer la fréquence de décharge est important. En d'autres termes, plus le temps de traitement alloué à l'algorithme est important, plus le résultat est proche d'une convolution. Cette approche *anytime* permet ainsi d'obtenir des résultats anticipés avant d'avoir traité l'image complète.

Nous avons proposé une étude d'un cas particulier visant à caractériser le filtrage neuronal par codage impulsionnel, qui indique que le filtrage réalisé est non-linéaire en fonction de la luminance. Nous avons étudié expérimentalement les différences entre le filtrage neuronal et le filtrage par convolution, en comparant la distance entre les réponses de ces filtres pour différentes images de test. Le codage impulsionnel donne des résultats qui approchent assez bien le filtrage par convolution, même si des différences locales peuvent parfois apparaître. Le codage fréquentiel donne des résultats beaucoup plus proches du filtrage par convolution et reproduit les variations fines de la réponse. Une mesure de distance plus complexe pourrait mieux capturer les différences fines entre les différents résultats de filtrage. D'autre part, les méthodes de construction des images résultant du filtrage neuronal pourraient être modifiées pour approcher au

mieux le résultat du filtrage par convolution. L'étude de paramètres proposée en fin de chapitre met en évidence le fonctionnement du filtrage neuronal et montre que la qualité de l'approximation du filtrage obtenue s'améliore lorsque l'algorithme dispose de plus de temps de calcul.

Le chapitre suivant s'appuie sur ce filtrage neuronal pour implémenter une architecture préattentionnelle. Cette architecture, qui permet d'extraire les saillances d'une image, nécessite d'utiliser plusieurs filtres sur plusieurs échelles spatiales. Les propriétés *anytime* du filtrage neuronal sont mises à profit pour obtenir des résultats précoces, avant que l'ensemble des traitements soient terminés.

Chapitre 5

Architecture préattentionnelle

Sommaire

5.1 Description de l'architecture	110
5.1.1 Transduction artificielle	110
5.1.2 Décomposition en traits caractéristiques	114
5.2 Extraction des saillances	116
5.2.1 Mécanismes mis en jeu	116
5.2.2 Résultats de l'extraction de saillances	118
5.3 Focalisation préattentionnelle	127
5.3.1 Sélection et suivi d'une région saillante	128
5.3.2 Évaluation de la robustesse de la focalisation	131
5.3.3 Focalisation sur une séquence d'images naturelles	134
5.4 Conclusion	136

Ce chapitre décrit une architecture de vision artificielle bio-inspirée, qui implémente un processus préattentionnel avec un réseau de neurones impulsionnels. Cette architecture, schématisée sur la figure 5.1, sépare les canaux d'une image d'entrée et en extrait différents traits caractéristiques (oppositions de couleurs, contours et orientations) dans les hautes et basses fréquences spatiales (partie 5.1). Les traits caractéristiques sont obtenus en réalisant un filtrage neuronal, comme expliqué dans le chapitre 4. La carte *PreAc* regroupe les informations obtenues dans les basses fréquences spatiales et préactive la carte de saillances. La carte de saillances combine les informations issues des hautes fréquences spatiales et les régions préactivées par *PreAc* pour extraire des régions saillantes (partie 5.2). Les régions saillantes trouvées sont mises en compétition sur une carte de focus, qui sélectionne et maintient le focus d'attention sur la région la plus saillante. Ce focus d'attention reste stable même lorsque la cible focalisée bouge (par-

tie 5.3). La conclusion sur l'architecture de vision préattentionnelle et sur les résultats obtenus est présentée dans la partie 5.4.

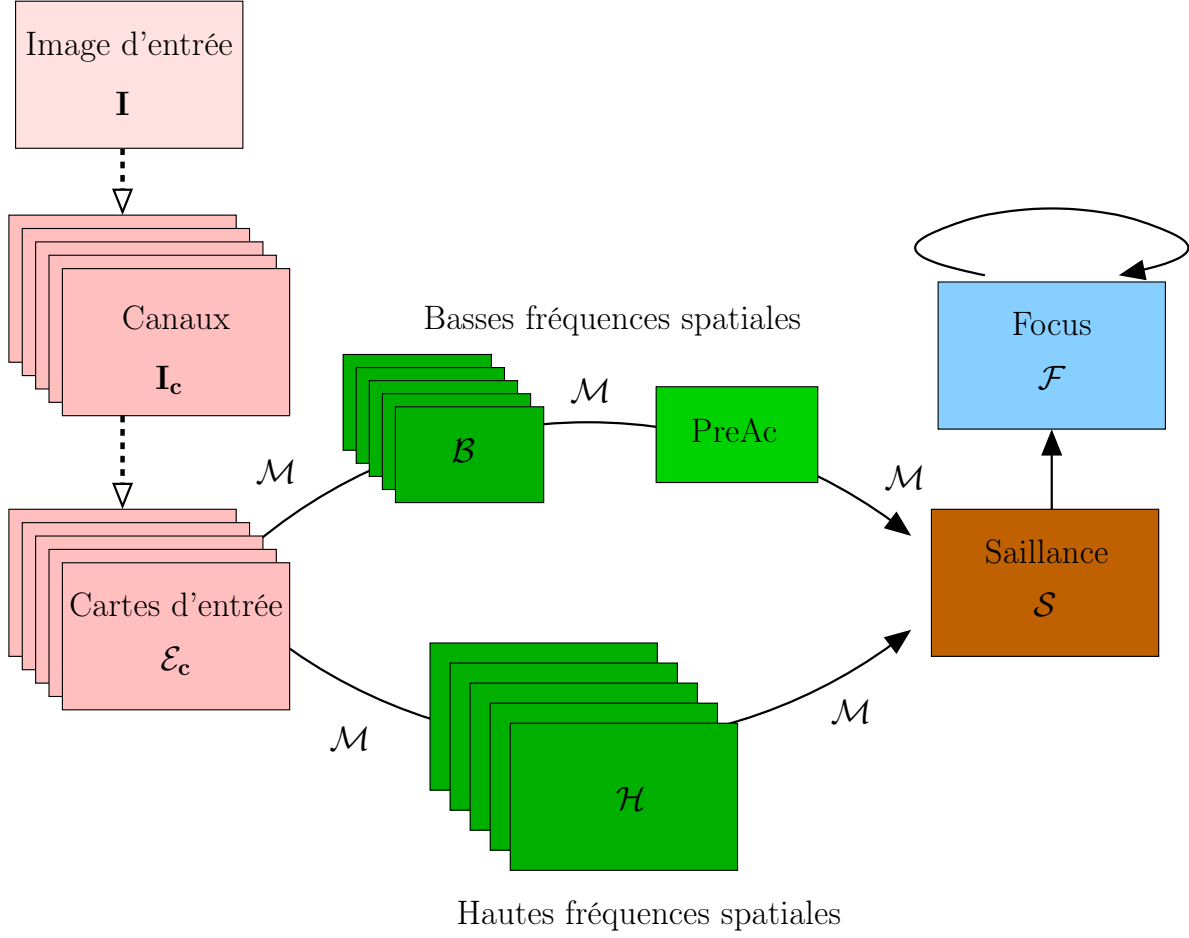


FIGURE 5.1 Schéma présentant l'architecture du modèle de vision attentionnelle : l'image d'entrée I est décomposée selon ses canaux de luminance et de couleurs I_c qui sont mises en entrée du réseau de neurones. Les traits caractéristiques (couleurs, contrastes, orientations) sont extraits pour deux échelles spatiales (B et H) à l'aide de masques de connexion M . Les basses fréquences spatiales sont regroupées sur la carte de $PreAc$ qui préactive la carte de Saillances S . Les informations des hautes fréquences arrivent ensuite, grâce à l'introduction d'un délai dans le réseau de neurones impulsioneels. La carte de Focus F prend en entrée les régions saillantes et sélectionne la plus saillante d'entre elles comme focus d'attention, en la suivant si elle se déplace.

5.1 Description de l'architecture

5.1.1 Transduction artificielle

La transduction, décrite dans la partie 4.1.1, est réalisée par les cartes d'entrées pour différentes informations. À partir des canaux r , g et b (rouge, vert et

bleu) d'une image \mathbf{I} , nous séparons les différentes informations de luminance et de couleurs pour obtenir un ensemble d'images d'entrée $\mathbf{I}_{\mathbf{c}}$ avec $\mathbf{c} \in \mathbf{C}$ et $\mathbf{C} = \{\text{Luminance, Red, Green, Blue, Yellow}\}$. L'image des luminances $\mathbf{I}_{\text{Luminance}}$ est obtenue avec

$$\mathbf{I}_{\text{Luminance}} = \frac{r + g + b}{3} \quad (5.1)$$

Pour séparer les couleurs et constituer les images de couleurs Red, Green, Blue et Yellow, nous avons choisi une méthode similaire à celle utilisée dans [Itti et al., 1998], qui permet de découpler la teinte de la luminance. Les images de couleurs sont obtenues par

$$\mathbf{I}_{\text{Red}} = r - \frac{(g + b)}{2} \quad (5.2)$$

$$\mathbf{I}_{\text{Green}} = g - \frac{(r + b)}{2} \quad (5.3)$$

$$\mathbf{I}_{\text{Blue}} = b - \frac{(r + g)}{2} \quad (5.4)$$

$$\mathbf{I}_{\text{Yellow}} = \frac{r + g}{2} - \frac{|r - g|}{2} - b \quad (5.5)$$

Toutes les valeurs négatives sont rectifiées à 0. Toutes les valeurs de $\mathbf{I}_{\mathbf{c}}$ sont normalisées entre 0 et 1. La figure 5.2 montre les images $\mathbf{I}_{\mathbf{c}}$ résultant de la séparation des différents canaux de l'image d'entrée \mathbf{I} .

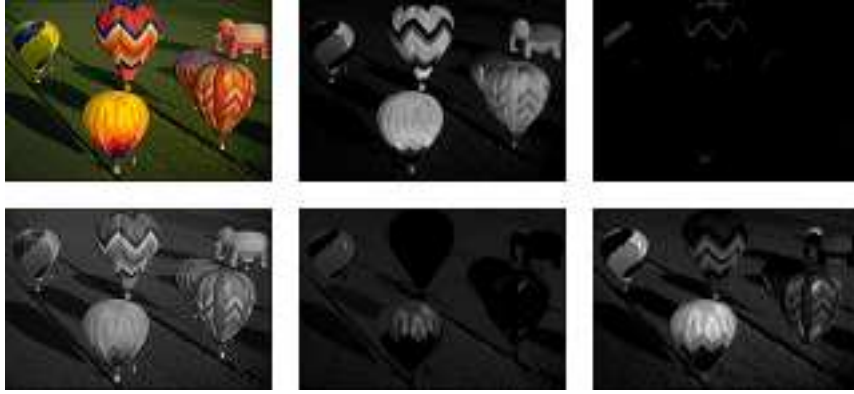


FIGURE 5.2 Illustration de la séparation des informations de luminance et de couleurs. L'image originale \mathbf{I} est en haut à gauche, l'image $\mathbf{I}_{\text{Luminance}}$ est en dessous. Les quatre images de droite montrent le résultat de la séparation des canaux de couleurs, avec de haut en bas et de gauche à droite : \mathbf{I}_{R} , \mathbf{I}_{B} , \mathbf{I}_{G} et \mathbf{I}_{Y} .

Les images $\mathbf{I}_{\mathbf{c}}$ sont utilisées pour déterminer les courants appliqués aux neurones des cartes d'entrées, notée $\mathcal{E}_{\mathbf{c}}$ (voir la figure 5.1). Le potentiel $V_{\mathcal{E}_{\mathbf{c}}, \mathbf{r}}$ du neurone occupant la position spatiale \mathbf{r} de la carte d'entrée $\mathcal{E}_{\mathbf{c}}$ est décrit par :

$$\begin{cases} \frac{dV_{\mathcal{E}_{\mathbf{c}}, \mathbf{r}}}{dt} = -\lambda_{\mathcal{E}} V_{\mathcal{E}_{\mathbf{c}}, \mathbf{r}}(t) + K_{\mathcal{E}_{\mathbf{c}}} I_{\mathbf{c}, \mathbf{r}}, & \text{si } V_{\mathcal{E}_{\mathbf{c}}, \mathbf{r}} < \vartheta \\ \text{sinon émission d'un potentiel d'action et } V_{\mathcal{E}_{\mathbf{c}}, \mathbf{r}} \text{ revient à } V_{\text{reset}} \end{cases} \quad (5.6)$$

avec $I_{c,r}$ la valeur du canal c à la position r . Le filtrage neuronal utilisé permet de traiter des séquences d'images, sans nécessiter une réinitialisation du réseau de neurones. La figure 5.3 montre l'évolution d'un potentiel de membrane $V_{e,r}$ (vert ou gris clair) sous l'influence d'un courant $I_{c,r}(t)$ variable (bleu ou gris foncé). La ligne pointillée indique le seuil ϑ fixé lors de la simulation. Les variations du courant $I_{c,r}(t)$ reflètent les différentes valeurs prises par le pixel $I_{c,r}$ lors de la présentation d'images successives. On pourra noter que le neurone ne décharge pas si $I_{c,r}(t) < \lambda_i \vartheta / K$, comme illustré dans la partie centrale de la figure 5.3.

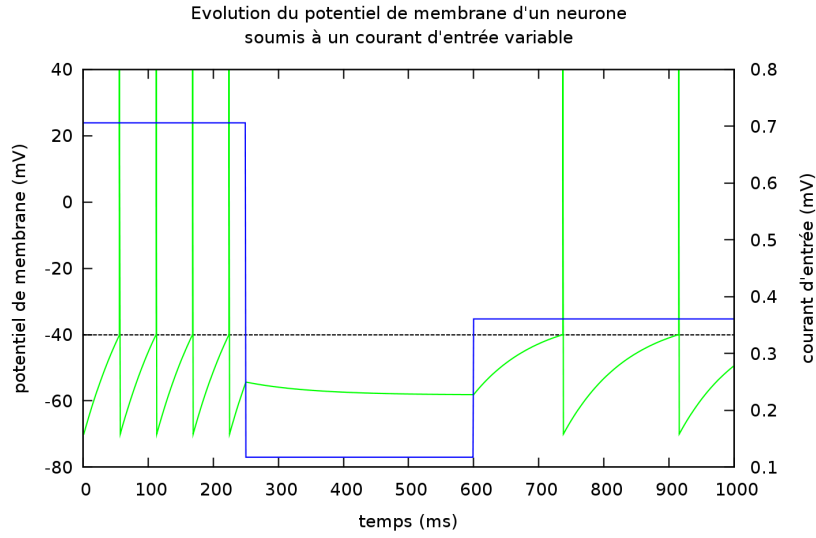


FIGURE 5.3 Évolution du potentiel de membrane $V_{e,r}$ recevant un courant d'entrée variable $I_{c,r}(t)$. De $0 < t \leq 250$ ms le neurone est soumis à un courant d'entrée constant $I_{c,r}^0 = 0.7$ mV, de $250 < t \leq 600$ ms le courant est $I_{c,r}^1 = 0.12$ mV et de $t > 600$ ms le neurone reçoit un courant $I_{c,r}^2 = 0.36$ mV. On pourra noter que, dans la partie centrale de la figure, l'injection du courant $I_{c,r}^1$ est trop faible pour permettre au neurone de décharger.

L'architecture proposée peut donc prendre une image ou une séquence d'images en entrée, les informations de luminance et de couleur de chacune de ces images servent à alimenter 5 cartes neuronales d'entrée (luminance, rouge, vert, bleu et jaune). Ces informations sont transformées en trains de potentiels d'action, en utilisant un codage fréquentiel, et sont envoyés à différentes cartes neuronales qui extraient les traits caractéristiques de l'image d'entrée. Une explication sur l'implémentation des neurones des cartes d'entrée est proposée dans l'annexe A et une amélioration possible de cette implémentation est décrite dans l'annexe C.

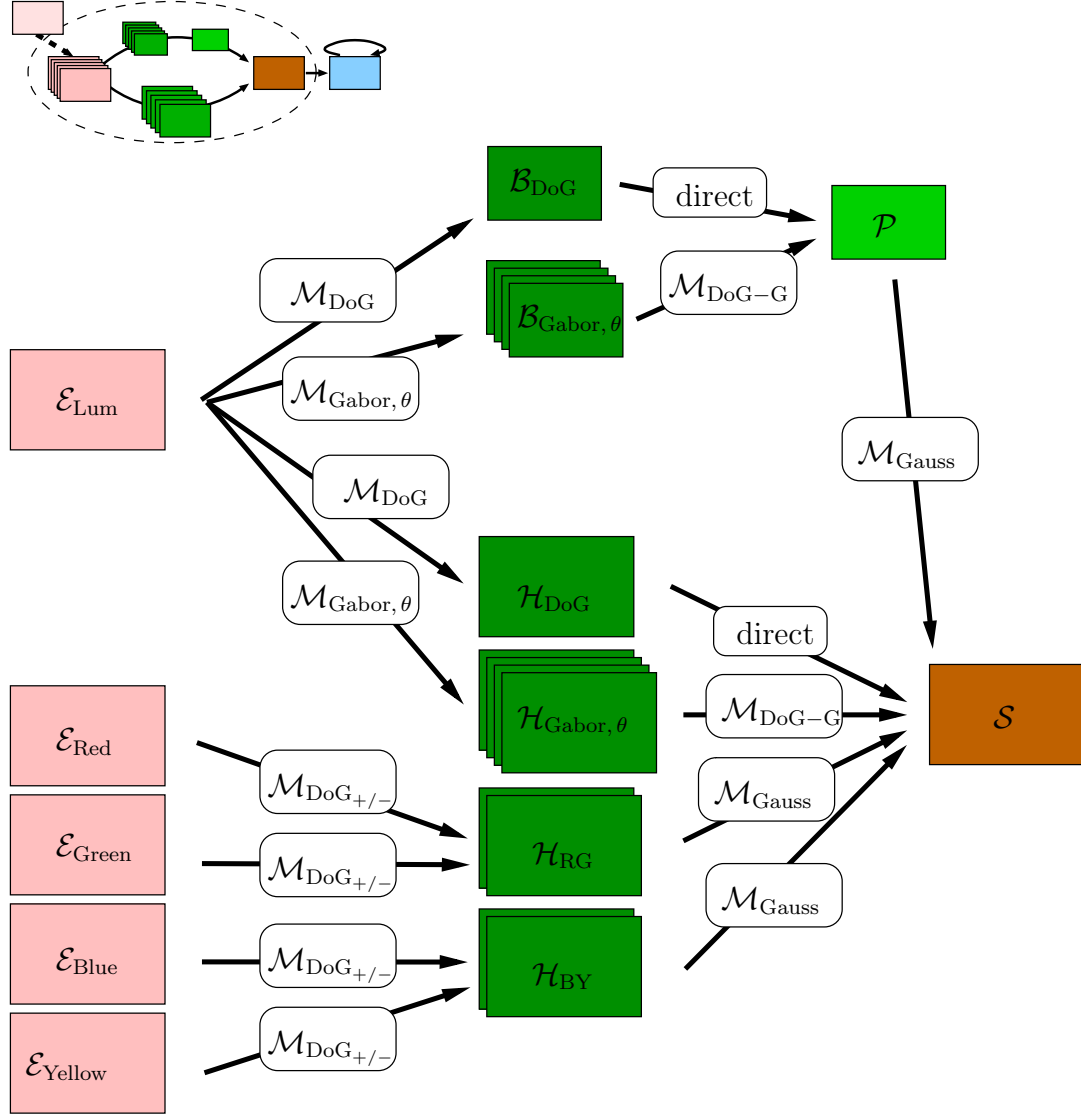


FIGURE 5.4 Schéma précisant la partie de l'architecture dévolue à la décomposition de l'image en traits caractéristiques et l'extraction de saillances. Les boîtes rectangulaires représentent les cartes neuronales et les flèches indiquent les projections, sur lesquelles sont indiqués les types de masque de connexion utilisés.

5.1.2 Décomposition en traits caractéristiques

La décomposition en traits caractéristiques est effectuée par l'intermédiaire de masques de connexion. Un masque de connexion \mathcal{M} explicite les poids et les délais des connexions entre les neurones d'une carte source et ceux d'une carte cible (voir figure 4.1 de la partie 4.1). Ces masques permettent de définir un champ récepteur commun à tous les neurones d'une carte [Wolff *et al.*, 1999]. Les valeurs des matrices de poids sont calculées à partir des réponses impulsionnelles de filtres utilisés en traitement d'images. Les filtres choisis reflètent les propriétés des champs récepteurs des neurones corticaux : les filtres gaussiens, les différences de gaussiennes ou chapeaux mexicains et les ondelettes de Gabor.

Le masque de connexion \mathcal{M}_{DOG} est utilisé pour obtenir, à partir de la carte des luminances $\mathcal{E}_{\text{Luminance}}$, une carte des contrastes locaux \mathcal{H}_{DOG} (voir figure 5.4). Ce masque de connexion implémente une différence centre-pourtour similaire aux traitements réalisés par les cellules ganglionnaires *on-off* de la rétine, comme expliqué dans la partie 3.2.3. Les valeurs des poids utilisés correspondent à la réponse impulsionnelle d'un filtre différence de gaussiennes :

$$\text{DOG}(x, y) = A_1 \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{x^2 + y^2}{2\sigma_1^2}\right) - A_2 \frac{1}{\sigma_2 \sqrt{2\pi}} \exp\left(-\frac{x^2 + y^2}{2\sigma_2^2}\right) \quad (5.7)$$

avec σ_1 et σ_2 les variance des gaussiennes et A_1 et A_2 les coefficients multiplicateurs avec $A_2 > A_1$. Les valeurs des poids utilisé pour le masque \mathcal{M}_{DOG} sont illustrées sur la figure 4.8 de la partie 4.3.1.

Pour obtenir la carte des basses fréquences spatiales, nous avons choisi de réduire la taille de la carte et de conserver le masque aux mêmes dimensions. Ainsi la carte \mathcal{B}_{DOG} , de dimension $\frac{M}{2} \times \frac{N}{2}$, est obtenue à partir du masque de connexion \mathcal{M}_{DOG} . En effet, pour obtenir une décomposition sur plusieurs échelles spatiales, il est possible de faire varier soit la taille du filtre (dans notre cas la matrice de poids) soit la taille des images. Nous avons choisi la deuxième solution qui présente l'intérêt d'utiliser des images plus petites et donc de réduire les temps de calcul.

Nous avons utilisé des ondelettes de Gabor pour extraire les informations d'orientations, ces ondelettes reflètent bien le filtrage réalisé dans V1 (voir la figure 3.3). Elles sont décrites par l'équation (3.1) de la partie 3.2.3. À partir de cette équation complexe,

il est possible de séparer la partie réelle et la partie imaginaire :

$$\text{Gabor}_R(x, y) = \cos(\nu x + \theta y) \exp\left(-\frac{x^2 + y^2}{2\sigma_G^2}\right) \quad (5.8)$$

$$\text{Gabor}_I(x, y) = \sin(\nu x + \theta y) \exp\left(-\frac{x^2 + y^2}{2\sigma_G^2}\right) \quad (5.9)$$

avec comme paramètres la fréquence ν , l'orientation θ , la variance de l'enveloppe gaussienne σ_G . Les valeurs sont normalisées entre $-w_{\max}^G$ et w_{\max}^G (voir la figure 5.5).

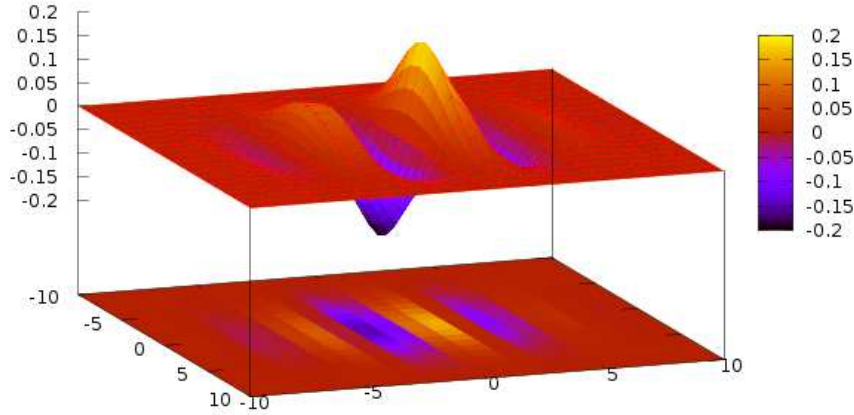


FIGURE 5.5 Partie imaginaire du filtre Gabor orienté à 90° , avec $\sigma_G = 3$, $\nu = 0.2$ et $\theta = 0$, de taille 21×21 et dont les valeurs sont normalisées avec $w_{\max}^G = 0.2$.

Nous avons utilisé quatre orientations (0° , 45° , 90° et 135°), qui nous permettent d'obtenir quatre cartes d'orientations $\mathcal{H}_{\text{Gabor},\theta}$ dans les hautes fréquences (voir figure 5.4). Les quatre cartes $\mathcal{B}_{\text{Gabor},\theta}$ des basses fréquences sont obtenues en utilisant le même principe que pour \mathcal{B}_{DOG} . Les cartes d'orientations $\mathcal{H}_{\text{Gabor},\theta}$ et $\mathcal{B}_{\text{Gabor},\theta}$ projettent vers les cartes suivantes au travers d'un masque différence de gaussiennes $\mathcal{M}_{\text{DOG-G}}$. Ceci permet d'améliorer la robustesse des informations d'orientation en propageant principalement les contrastes locaux d'orientations, comme l'ont proposé [Itti *et al.*, 1998].

Les informations de couleurs ne sont traitées que pour les hautes fréquences spatiales. Nous construisons quatre cartes dites de “double opposition de couleurs” [Itti *et al.*, 1998] : deux cartes d'opposition rouge-vert (\mathcal{H}_{RG1} et \mathcal{H}_{RG2} , voir figure 5.4) et deux cartes d'opposition bleu-jaune (\mathcal{H}_{BY1} et \mathcal{H}_{BY2}). La première carte de double opposition rouge-vert simule une carte dont les neurones ont un champ récepteur qui dans sa partie centrale est excité par le rouge et inhibé par le vert et dans sa partie périphérique est

inhibé par le rouge et excité par le vert. Les neurones de la seconde carte de double opposition de couleurs ont un champ récepteur qui est inhibé par le rouge et excité par le vert dans sa partie centrale et inversement dans sa partie périphérique. Les champs récepteurs utilisés sont illustrés sur la figure 5.6.

Ces cartes de double opposition de couleurs sont connectées aux cartes d'entrée $\mathcal{E}_{\{\text{Red, Green, Blue, Yellow}\}}$ avec des masques $\mathcal{M}_{\text{DOG}_+}$ et $\mathcal{M}_{\text{DOG}_-}$ (voir figure 5.4). Le masque $\mathcal{M}_{\text{DOG}_+}$ est équivalent à \mathcal{M}_{DOG} : les valeurs de sa matrice de poids sont obtenues à partir de la réponse impulsionnelle du filtre DOG définie dans l'équation (5.7). Les valeurs de la matrice de poids de $\mathcal{M}_{\text{DOG}_-}$ sont inversées : elles sont obtenues avec $\text{DOG}_-(x, y) = -\text{DOG}(x, y)$. Ainsi, la première carte d'opposition rouge-vert \mathcal{H}_{RG1} reçoit d'une part les entrées de \mathcal{E}_{Red} au travers d'un masque $\mathcal{M}_{\text{DOG}_+}$ et d'autre part les entrées de $\mathcal{E}_{\text{Green}}$ au travers d'un masque $\mathcal{M}_{\text{DOG}_-}$. La seconde carte d'opposition \mathcal{H}_{RG2} est connectée à \mathcal{E}_{Red} avec un masque $\mathcal{M}_{\text{DOG}_-}$ et à $\mathcal{E}_{\text{Green}}$ avec $\mathcal{M}_{\text{DOG}_+}$.

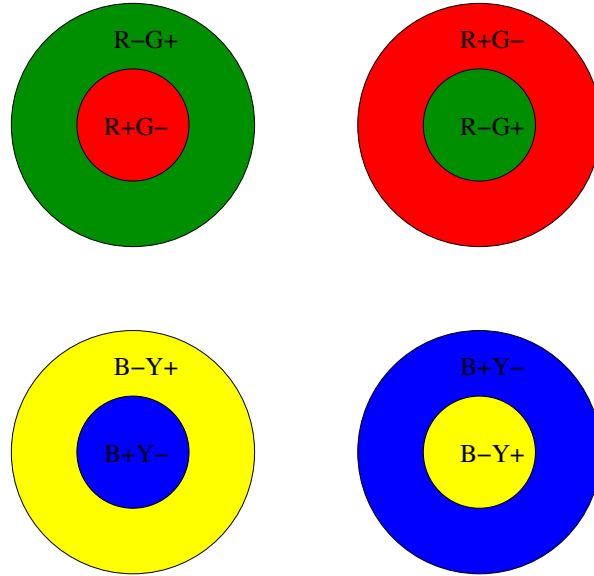


FIGURE 5.6 Illustration des champs récepteurs des neurones de chaque carte de double opposition de couleurs. Les couleurs prises en compte sont le rouge (R), le vert (G), le bleu (B) et le jaune (Y). Dans la partie centrale du champ récepteur, une couleur excite le neurone (notée +) et une couleur l'inhibe (notée -). Ces excitations et inhibitions sont inversées dans la partie périphérique.

5.2 Extraction des saillances

5.2.1 Mécanismes mis en jeu

À partir de la définition de la saillance formulée dans la partie 3.3, nous proposons une implémentation qui repose principalement sur deux mécanismes détaillés dans les

paragraphe suivants. D’une part, les informations issues des basses fréquences spatiales servent à orienter l’attention et d’autre part les neurones de la carte de saillances \mathcal{S} tiennent compte des informations temporelles précises des trains de potentiels d’action qu’ils reçoivent. Ce codage temporel précis est utilisé dans les systèmes biologiques par les cellules ganglionnaires [Meister et Berry, 1999] et dans les aires corticales [Bohte, 2004].

L’orientation de l’attention par les basses fréquences spatiales permet de réduire la recherche de saillances aux régions saillantes dans les basses fréquences. Pour ce faire, la séparation en traits caractéristiques est réalisée au travers de deux voies distinctes, traitant chacune une échelle spatiale. Cette séparation reproduit la ségrégation des informations visuelles réalisée par les voies parvo- et magnocellulaire de la voie rétino-géniculo-striée (voir la partie 1.2.3). La voie magnocellulaire transporte vers le cortex visuel des informations sur les basses fréquences spatiales et ne comporte pas d’informations sur les couleurs. La voie parvocellulaire porte des informations sur les couleurs et d’une plus grande précision spatiale, mais les informations issues de cette voie arrivent dans le cortex visuel avec un léger retard par rapport à celles de la voie magnocellulaire. Ce léger décalage temporel est implémenté en ajoutant un délai fixe dans les masques de connexions entre les cartes \mathcal{H}_{DOG} , $\mathcal{H}_{\text{Gabor}, \theta}$, \mathcal{H}_{RG} , \mathcal{H}_{BY} et la carte de saillances : il permet aux informations des basses fréquences, regroupées sur la carte \mathcal{P} , de préactiver certaines zones de la carte de saillances \mathcal{S} . L’attention est ainsi guidée par des indices globaux et peut s’orienter vers les zones les plus susceptibles d’être intéressantes.

La seconde particularité de l’extraction de saillances réalisée repose sur une spécificité du modèle de neurone utilisé. En effet, les neurones des cartes d’entrées et de traits caractéristiques sont des intégrateurs (voir la partie 4.1), qui accumulent les informations entrantes et déchargent lorsqu’ils ont reçu suffisamment d’excitations. Au contraire, les neurones de la carte de saillances sont des détecteurs de coïncidences et ils n’émettent des potentiels d’action que lorsqu’ils reçoivent des informations corrélées temporellement (voir la figure 3.5 de la partie 3.3.1). Dans l’implémentation proposée, les neurones de la carte de saillances ne déchargent que lorsqu’ils sont préactivés par les informations venant des basses fréquences spatiales et qu’ils reçoivent des potentiels d’action synchrones en provenance de plusieurs modalités des hautes fréquences. Si la connexion entre la carte de préactivation et la carte de saillances est coupée, les neurones de la carte de saillances ne déchargent pas.

Les spécificités du modèle de filtrage neuronal décrit dans le chapitre précédent sont particulièrement intéressantes pour l’extraction de saillance. La détermination des régions saillantes dans une image nécessite de réaliser de nombreuses opérations de filtrage

et une approche utilisant le filtrage par convolution doit nécessairement attendre la fin de la convolution de toutes les images avant de pouvoir déterminer les régions saillantes. Le filtrage neuronal permet d'utiliser le résultat du filtrage dès qu'il est disponible et, comme les calculs sont effectués en parallèle sur toutes les cartes, on peut obtenir les premiers résultats sur la carte de saillances avant que l'image ne soit complètement filtrée pour toutes les modalités et à toutes les échelles. L'architecture neuronale décrite ici implémente bien un processus *anytime*, qui permet d'obtenir les premières saillances dès que possible et donne un résultat d'autant plus complet que l'on poursuit les calculs.

La partie suivante montre les saillances obtenues expérimentalement pour différentes images et met l'accent sur l'aspect *anytime* des résultats ainsi obtenus.

5.2.2 Résultats de l'extraction de saillances

Comme dans la partie 4.3, les saillances obtenues peuvent être représentées sous deux formes selon que l'on considère le codage impulsionnel ou le codage fréquentiel. Le premier s'intéresse au premier potentiel d'action émis par chaque neurone de la carte de saillances et le second à la fréquence moyenne de décharge de chaque neurone. Les images de saillance présentées ici sont construites de la même façon que dans le chapitre précédent.

Il n'existe pas de définition de la saillance qui soit largement acceptée, il est donc difficile de proposer une méthode de comparaison objective qui permette d'évaluer la qualité des saillances extraites par les différents systèmes existants. Quand bien même il existerait une définition unique et une méthode de référence pour extraire les saillances, la définition d'une métrique pour réaliser des comparaisons serait complexe. Nous présentons donc dans cette partie les saillances extraites avec l'architecture neuronale et, pour permettre une comparaison, nous présentons les saillances extraites avec le système de [Itti et al., 1998]. Le laboratoire iLab propose une implémentation, appelée iNVT [Itti, 2002], de la méthode d'extraction de saillance utilisant l'algorithme d'Itti. Cependant, cette méthode est incluse dans un système beaucoup plus large, qui est utilisé pour développer les très nombreux projets de ce laboratoire. Nous avons donc choisi d'utiliser une ré-implémentation de la méthode de [Itti et al., 1998], beaucoup plus facilement paramétrable, proposée par [Fix, 2007].

Il est à noter que l'algorithme d'Itti donne une image résultante d'une taille quatre fois inférieure à l'image d'entrée : pour une image de 256x256 pixels, la carte de saillances est de 64x64 pixels. En effet, les traitements sont réalisés sur une pyramide gaussienne dyadique composée de neuf échelles spatiales qui vont de l'échelle 0 (1 :1) à l'échelle

8 (1 :256). Les résultats des traitements sont combinés sur la carte de saillances, que les auteurs ont choisi de représenter à l'échelle 2 (1 :4). La carte de saillances de l'architecture neuronale est de mêmes dimensions que l'image d'entrée, nous avons donc choisi de mettre en entrée du système des images réduites à l'échelle 1 :4. Ainsi les deux méthodes d'extraction de saillances fournissent en sortie des images de même taille.

La figure 5.7 et les autres figures de cette partie utilisent le même paramétrage. Les neurones de la carte d'entrée des luminances \mathcal{E}_{lum} sont paramétrés avec $\tau = 100\text{ms}$, $E_l = -70\text{mV}$, $\vartheta = -40.0\text{mV}$, $V_{\text{reset}} = -80\text{mV}$ et $K = 1$. Les neurones des cartes de couleurs \mathcal{E}_{Red} , $\mathcal{E}_{\text{Green}}$, $\mathcal{E}_{\text{Blue}}$ et $\mathcal{E}_{\text{Yellow}}$ sont paramétrés comme \mathcal{E}_{lum} , avec $\tau = 1000\text{ms}$ et $K = 5$. Les neurones de la carte de saillances utilisent les mêmes valeurs que \mathcal{E}_{lum} , avec $\tau = 1.8\text{ms}$. Les neurones des cartes de traits caractéristiques sont paramétrés avec $\tau = 200\text{ms}$, $E_l = V_{\text{reset}} = 0$ et $\vartheta = 1$. Les simulations sont toutes réalisées sur 400 pas de temps, soit 40ms de temps simulé. Les paramètres des masques de connexions utilisés sont détaillés dans les tableaux suivants :

\mathcal{M}_{DOG}		$\mathcal{M}_{\text{Gabor},\theta}$	
A_1	2.0	σ_G	3.0
A_2	1.0	ν	$\{0.2, 0.1415, 0.0, -0.1415\}$
σ_1	1.0	θ	$\{0.0, 0.1415, 0.2, 0.1415\}$
σ_2	$\sqrt{2}$	w_{max}	0.2
w_{max}	0.6		

$\mathcal{M}_{\text{DOG}_{+/-}}$		$\mathcal{M}_{\text{Gauss}}$	
idem \mathcal{M}_{DOG}		σ	2.0
w_{max}	0.2	w_{max}	2.0

$\mathcal{M}_{\text{DOG-G}}$		$\mathcal{M}_{\text{direct}}$	
idem \mathcal{M}_{DOG}		w_{max}	2.0
w_{max}	1.5		

TABLE 5.1 Paramètres des masques de connexion utilisés pour l'extraction de saillances.

La figure 5.7 montre les résultats des saillances extraites avec l'approche neuronale et avec l'algorithme d'Itti. Les saillances obtenues ici n'utilisent que les informations de luminance et quatre orientations dans $[0, \pi]$. On peut constater sur cette figure que l'algorithme d'Itti extrait des saillances très diffuses et donne un résultat en "marches d'escalier" ou "gaufrés", dû à l'utilisation de pyramides gaussiennes pour réaliser le filtrage DOG. L'architecture neuronale donne un résultat très différent, beaucoup plus précis dans la localisation des régions saillantes. L'image obtenue avec le codage fréquentiel donne des résultats beaucoup plus fins que celle du codage impulsionnel. On pourra remarquer que l'algorithme d'Itti donne des résultats qui dépendent des valeurs

absolues des luminances : les avions (sombre sur un fond clair) ne sont pas considérés comme saillants. L'architecture neuronale n'est pas du tout affectée par cet effet : les avions sont considérés comme saillant. Ceci illustre le fait que l'approche neuronale permet de détecter des saillances aussi bien sombres que claires, sans avoir besoin de traiter l'image en négatif.

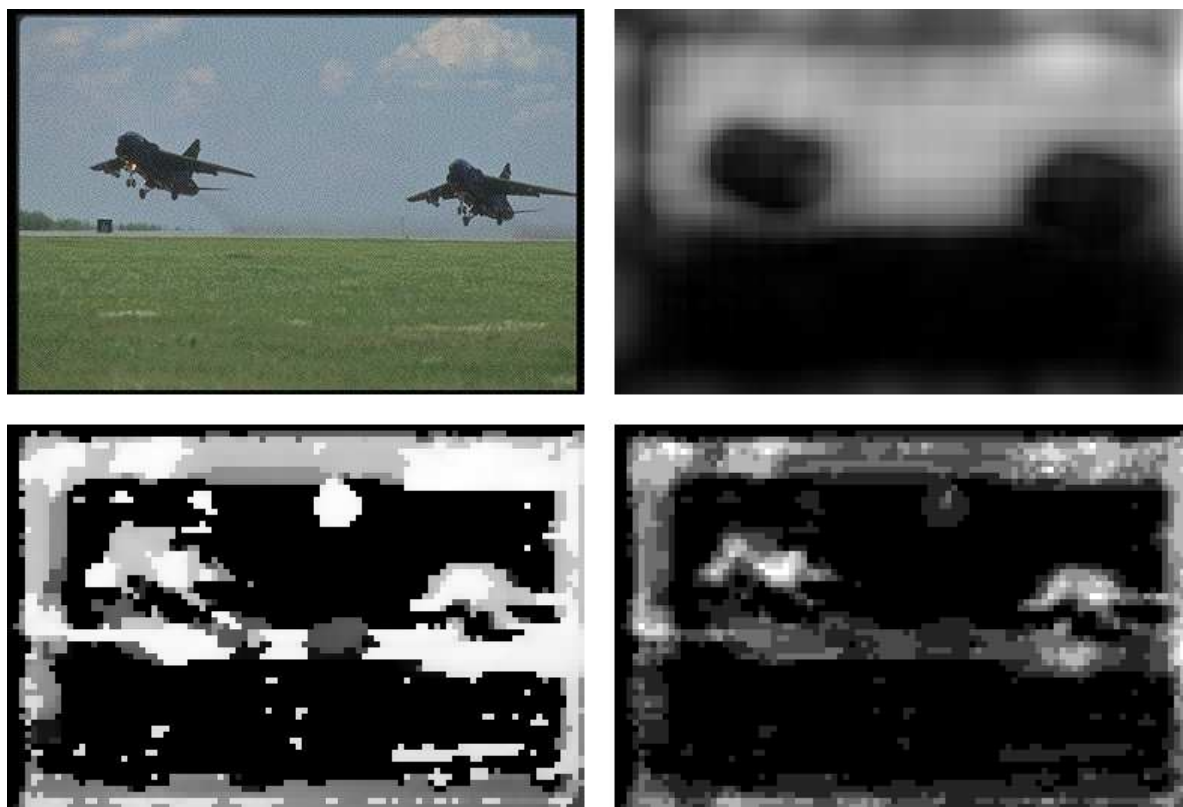


FIGURE 5.7 Comparaison des saillances extraites sur une image pour les luminances. En haut à gauche l'image originale, en haut à droite les saillances obtenues avec [Itti et al., 1998]. En bas, les saillances obtenues avec l'architecture neuronale, à gauche avec le codage impulsionnel et à droite avec le codage fréquentiel.

Pour mettre en évidence le caractère *anytime* des résultats obtenus, nous avons mesuré les différences entre les images résultantes des codages impulsionnel et fréquentiel à chaque pas de temps de calcul et les images résultantes des codages impulsionnel et fréquentiel finaux (au dernier pas de temps). Ces résultats sont présentés sur la figure 5.8, en haut. Les images en bas de la figure 5.8 montrent les résultats obtenus à différents pas de temps. Il apparaît que les résultats obtenus gagnent en précision quand le nombre de pas de temps augmente.

Comme nous l'avons vu dans le chapitre précédent, le codage impulsionnel propage en premier lieu les informations les plus importantes : les coefficients les plus élevés suscitent des potentiels d'action précoces. Les premiers calculs de saillances sont ainsi effectués, aux approximations du filtrage neuronal près, sur les informations les plus

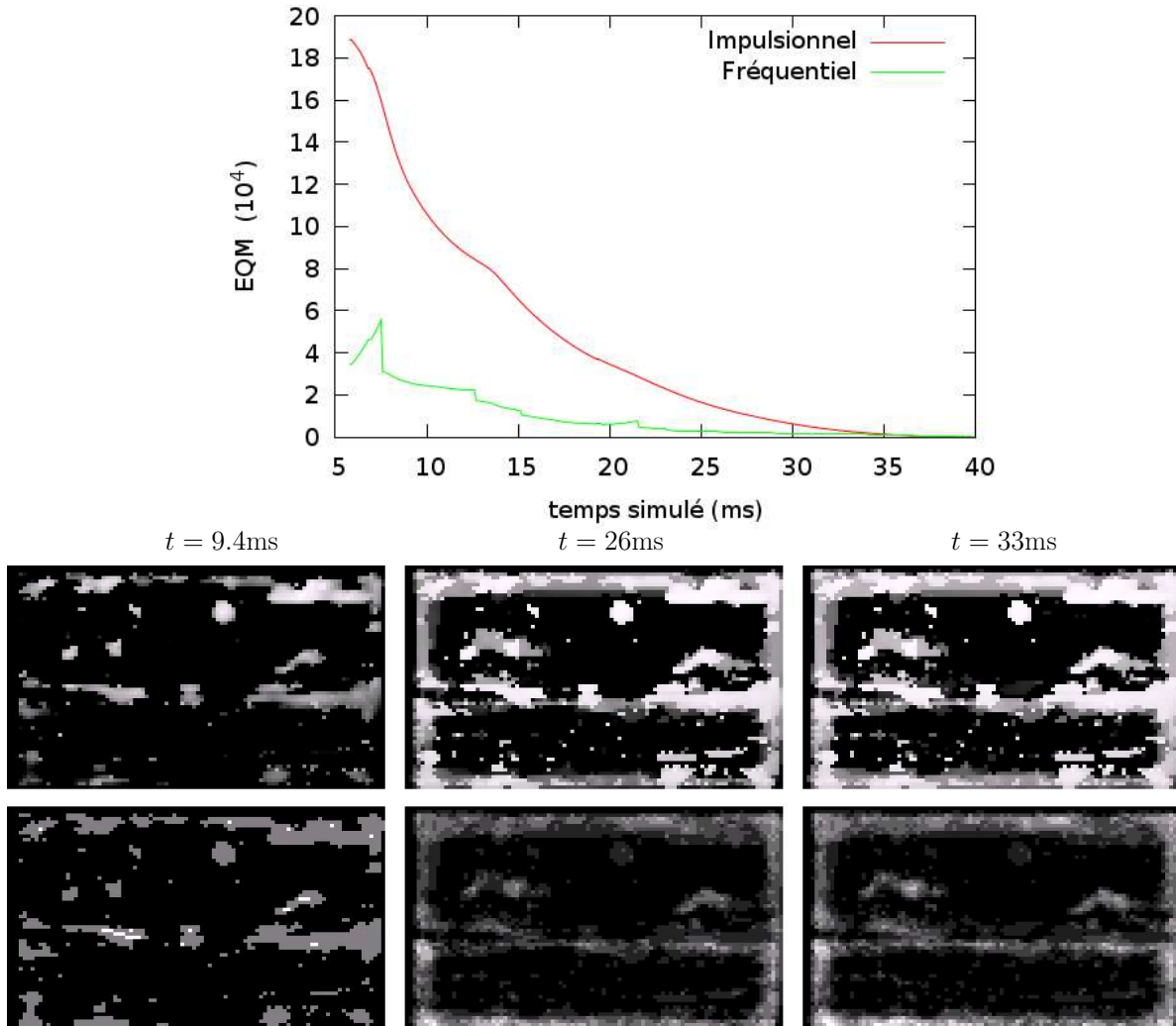


FIGURE 5.8 Mise en évidence du caractère *anytime* de l'extraction de saillance. En haut : l'EQM mesurée entre le résultat de l'extraction de saillance à chaque pas de temps et les saillances obtenues au dernier pas de temps, pour le codage impulsionnel en rouge ou gris foncé et le codage fréquentiel en vert ou gris clair. En bas : résultats de l'extraction de saillance avec le code impulsionnel (en haut) et avec le code fréquentiel (en bas) à différent instants de la simulation : à gauche les images résultantes après 9.4 ms de temps simulé, au centre après 26 ms et à droite après 33ms.

importantes. Ceci permet donc d’obtenir les saillances dans un ordre qui approche celui de leur importance, comme illustré sur la figure 5.9, où les premières régions détectées sont les plus saillantes.



FIGURE 5.9 À partir de l’image d’entrée, à gauche, les saillances obtenus avec le codage impulsif sont représentées en rouge sur l’image de droite. Les nombres indiquent l’ordre d’apparition des saillances.

Le choix de la valeur du paramètre τ des neurones de la carte de saillances influe qualitativement sur les résultats obtenus. Nous avons choisi des valeurs de τ très basses pour que les neurones se comportent comme des détecteurs de coïncidence¹. La figure 5.10 montre les différences obtenues pour différentes valeurs de τ , comprises entre 0.001 et 1 seconde. La courbe en haut de la figure représente l’EQM mesurée entre l’image résultante pour une valeur de τ donnée et l’image résultante avec la valeur de τ la plus basse. Les images en bas de la figure 5.10 illustrent les résultats obtenus pour différentes valeurs de τ . Plus ces valeurs sont élevées, et donc plus le neurone se comporte comme un intégrateur, plus les régions saillantes détectées sont diffuses et homogènes. Lorsque les neurones sont utilisés comme des détecteurs de coïncidence, le nombre de pixels considéré comme saillants diminue et les valeurs de saillances sont plus hétérogènes.

La figure 5.11 montre une comparaison entre les saillances obtenues pour différentes images avec l’architecture neuronale et la méthode de [Itti *et al.*, 1998], en utilisant l’implémentation proposée par [Fix, 2007]. Sur cette figure, seules les informations de luminance ont été considérées, c’est-à-dire que l’architecture utilisée ne comporte pas de cartes \mathcal{E}_{Red} , $\mathcal{E}_{\text{Green}}$, $\mathcal{E}_{\text{Blue}}$, $\mathcal{E}_{\text{Yellow}}$, \mathcal{E}_{RG} et \mathcal{E}_{BY} . Comme nous l’avons remarqué sur la figure 5.7, l’algorithme d’Itti obtient des saillances beaucoup plus diffuses et les valeurs de saillance de chaque pixel varient de façon abrupte. L’architecture neuronale permet d’obtenir des résultats beaucoup plus précis spatialement. Cette précision spatiale est aussi visible sur la figure 5.12, qui montre les saillances obtenues en utilisant les informations de couleurs et de luminance pour extraire les saillances.

L’activité des cartes de traits caractéristiques et de préactivation est illustrée sur les

1. Il est à noter que des effets similaires peuvent être obtenus en modifiant le potentiel de repos E_l ou le seuil ϑ .

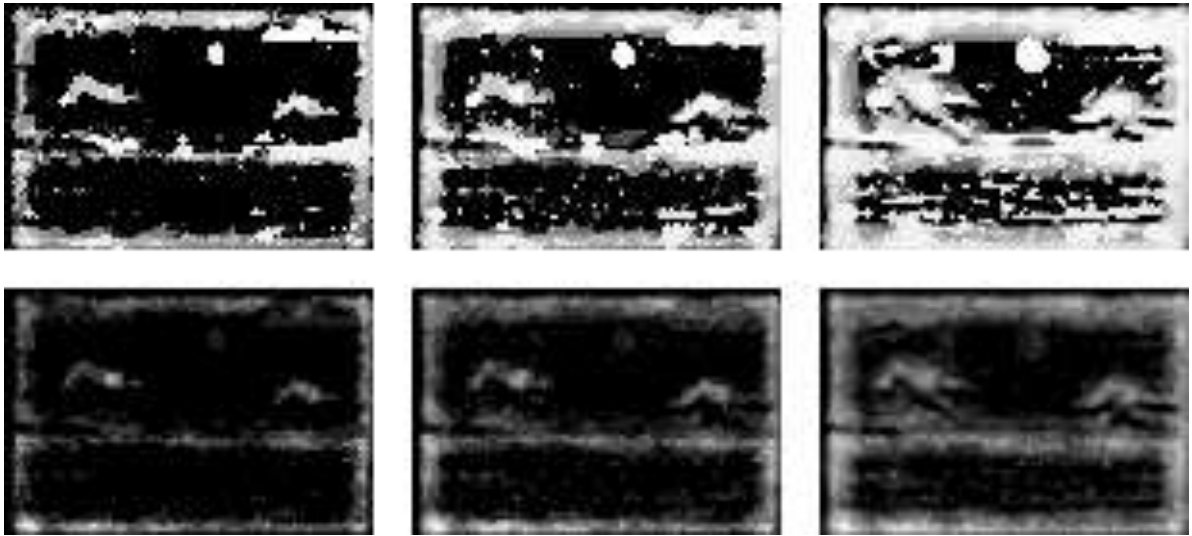
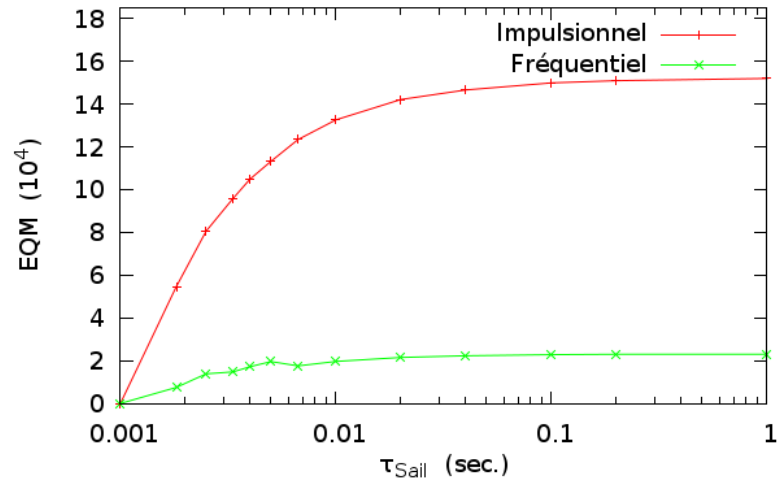


FIGURE 5.10 Influence du paramètre τ des neurones de la carte \mathcal{S} sur l'extraction de saillance. En haut, l'EQM mesurée entre l'image des saillances obtenues avec la valeur de τ la plus basse (les neurones se comportent alors comme des détecteurs de coïncidence) et le résultat obtenu avec une autre valeur de τ . En bas, les images résultantes obtenues avec $\tau = 0.001$ s (à gauche), $\tau = 0.0018$ s (au centre) et $\tau = 1$ s (à droite).

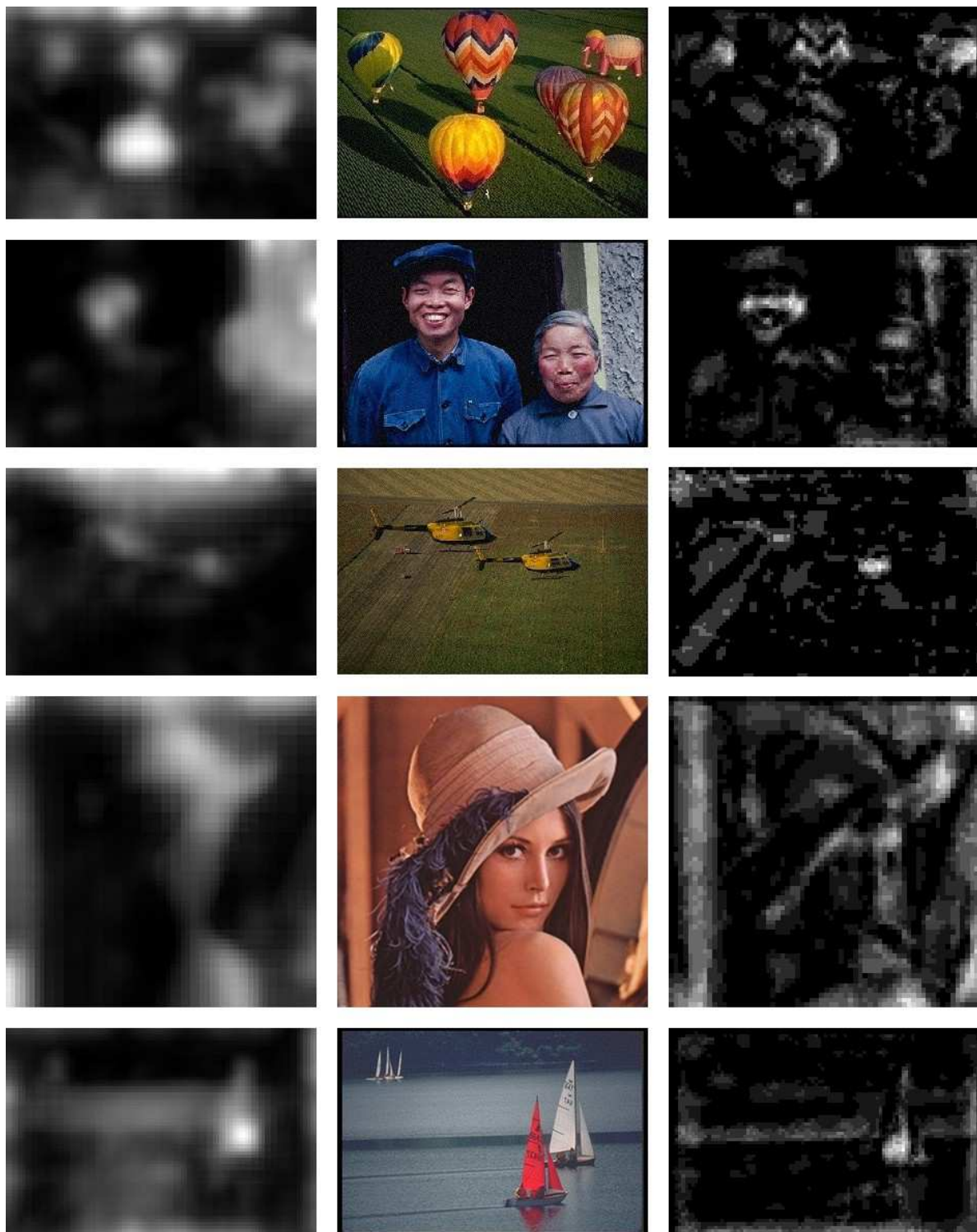


FIGURE 5.11 Comparaison des saillances extraites pour différentes images selon les informations de luminance. Les images d'entrée sont présentées au centre, les résultats obtenus avec l'algorithme d'Itti sont à gauche et ceux obtenus avec le codage fréquentiel sont à droite.

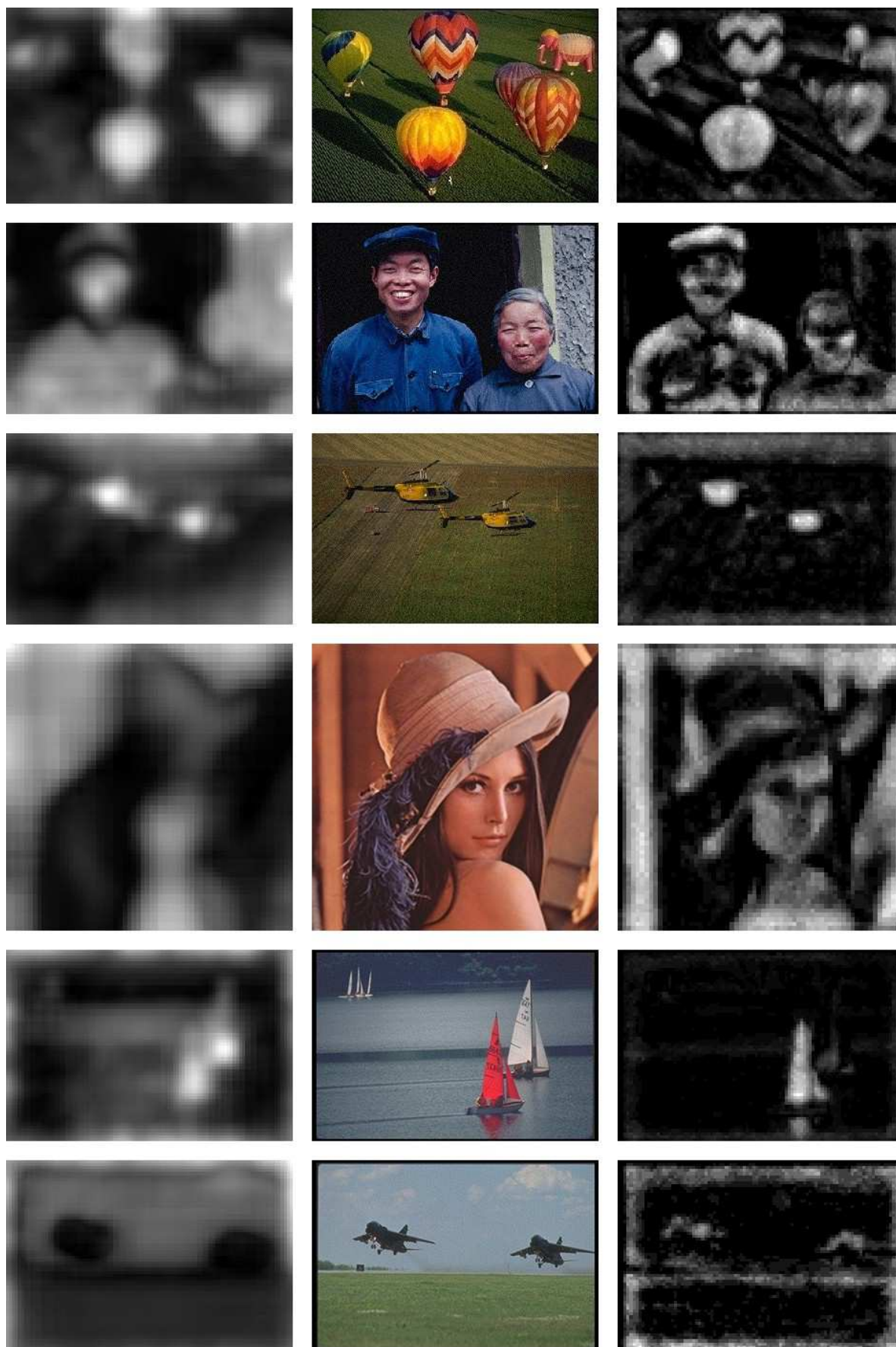


FIGURE 5.12 Comparaison des saillances extraites pour différentes images avec les couleurs. La disposition est identique à celle de la figure 5.11.

figures [A-1](#) et [A-2](#), dans l'annexe [D](#).

L'architecture proposée permet d'extraire les saillances d'une image et détecte des saillances comparables à celles obtenus avec la méthode de [\[Itti, 2007\]](#). Elle tire partie du caractère *anytime* du filtrage neuronal : les premières régions saillantes sont détectées avant que l'image ne soit intégralement filtrée. De plus, les saillances apparaissent dans l'ordre de leur importance : les régions les plus importantes sont détectées très vite et ensuite, plus le temps simulé s'écoule plus les résultats sont précis, au sens où chaque région converge vers la valeur de saillance qui la caractérise, et complet, puisque toutes les saillances sont détectées. La partie suivante propose un mécanisme neuronal permettant de porter le focus d'attention sur la région la plus saillante.

5.3 Focalisation préattentionnelle

Nous avons vu que les potentiels d'action émis par les neurones de \mathcal{S} codent temporellement l'intensité des saillances. Nous proposons d'implémenter un processus qui se focalise sur la région la plus saillante et qui conserve ce focus de façon robuste, même lorsque la cible bouge ou qu'elle se situe dans un environnement bruité. L'implémentation que nous avons choisie s'inspire des compétitions locales dues aux interactions latérales qu'il est possible d'observer dans les aires corticales. Ces connexions latérales peuvent se modéliser sous la forme de différences de gaussiennes.

L'émergence du focus d'attention prend place sur \mathcal{F} , une carte neuronale dotée d'une connexion récurrente en différence de gaussiennes (voir la figure 5.13). L'activité de \mathcal{F} est ainsi auto-régulée et reste concentrée sur une région particulière, le focus, pendant que le reste de la carte est inhibé.

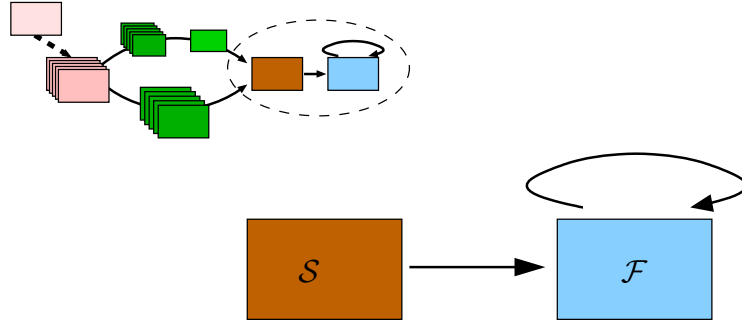


FIGURE 5.13 Les saillances extraites sur la carte \mathcal{S} sont envoyées à la carte de focus (\mathcal{F}), qui possède une connexion récurrente. L'activité de \mathcal{F} est auto-régulée par cette connexion qui implémente un algorithme de type *winner-takes-all*.

Cette approche, qui définit le focus d'attention comme une activité persistante sur une carte de neurones dotée d'une connexion récurrente, est inspirée de [Rougier et Vitay, 2006]. Les auteurs utilisent un type de modèle neuronal différent : les champs neuronaux dynamiques. Ce type de réseau de neurones est présenté dans la partie 2.1.4. Il modélise les variations d'activité d'une population de neurones, considérées comme des champs continus de potentiel membranaire. Nous proposons dans cette partie d'étudier le comportement d'une carte de neurones impulsionnels dotée du même type de connectivité. La stabilité du focus attentionnel est évalué pour deux types de conditions expérimentales : en présence de bruit et en présence de distracteurs. Les mesures choisies nous permettent de proposer une comparaison avec les résultats obtenus sur des champs neuronaux dynamiques dans [Rougier et Vitay, 2006]. Nous terminons cette partie en utilisant les saillances comme entrées de la carte de focus. La stabilité du focus d'attention est ainsi évaluée sur une séquence d'images naturelles.

5.3.1 Sélection et suivi d'une région saillante

Nous utilisons la dynamique des interactions latérales pour conserver le focus sur la région la plus saillante, même lorsque celle-ci bouge, lorsqu'elle est très fortement bruitée ou lorsqu'elle est entourée de distracteurs. En effet, seule la cible conserve une cohérence temporelle qui lui permet de conserver le focus d'attention, puisqu'elle se déplace doucement mais reste toujours active. Toutes les autres positions spatiales ont une activité qui n'est pas suffisamment constante dans le temps pour faire changer le focus d'attention.

Les interactions locales permettent au focus d'attention de suivre un stimulus en déplacement. Cependant, ce suivi est possible seulement si le déplacement du stimulus entre deux images n'est pas trop important. Il est nécessaire que d'une image à l'autre le stimulus reste dans la partie positive de la différence de gaussiennes pour que le focus puisse suivre le stimulus.

Pour évaluer la stabilité du focus en présence de bruit ou de distracteurs, nous avons utilisé un réseau composé de deux cartes neuronales (voir figure 5.14) : une carte d'entrée et une carte de focus. La carte d'entrée réalise une étape de transduction en transformant les valeurs de luminance d'une image d'entrée en potentiels d'action. Elle est connectée à la carte de focus avec un masque de connexion gaussien. Les interactions latérales des neurones de la carte de focus sont implémentées à l'aide d'un masque de connexion de type différence de gaussiennes, qui connecte la carte de focus sur elle-même.

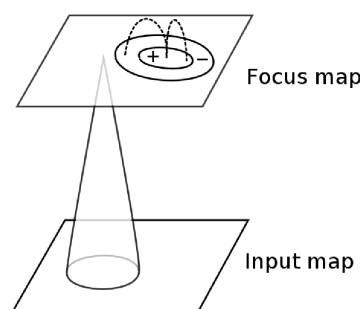


FIGURE 5.14 Schéma de l'architecture utilisée pour évaluer la stabilité du focus en présence de bruit ou de distracteurs. L'activation de la carte de Focus dépend des connexions latérales et de l'activité de la carte d'entrée (Input map).

La robustesse de la focalisation a été mesurée pour deux conditions expérimentales, illustrées sur la figure 5.15) :

- lorsque la cible se déplace et que l’image est bruitée,
- lorsque la cible se déplace et que l’image contient des distracteurs.

Le stimulus $s_{r,\theta,W,I}$ est un patch gaussien, centré sur $(r\sin\theta, r\cos\theta)$, de largeur W et d’intensité I :

$$s_{r,\theta,W,I}(x, y) = I e^{\frac{(x-x_C)^2}{W^2}} e^{\frac{(y-y_C)^2}{W^2}} \quad (5.10)$$

$$(x_C, y_C) = (r\sin\theta, r\cos\theta) \quad (5.11)$$

Une fois le stimulus placé sur l’image, on ajoute soit du bruit, soit des distracteurs. Les valeurs du bruit sont tirées d’une distribution gaussienne de moyenne nulle et de variance σ . Les neurones de la carte d’entrée reçoivent comme entrée I :

$$I = s_{r,\theta,W,I} + \eta \quad (5.12)$$

$$\eta \sim N(0, \sigma) \quad (5.13)$$

Nous avons utilisé différents niveaux de bruit, en prenant comme variance σ parmi l’ensemble des valeurs $\{0, 1; 0, 25; 0, 5; 0, 75; 1\}$. L’ajout de distracteurs dans l’image consiste à ajouter un certain nombre D de patches gaussiens, chacun d’entre eux étant exactement semblable au stimulus, avec $D \in \{1; 2; 3; 5; 10; 25\}$.

Les simulations réalisées suivent toutes le protocole suivant :

1. on présente tout d’abord, en entrée du réseau, l’image initiale qui ne contient que le stimulus (sans bruit ni distracteurs)
2. le système calcule 20 pas de temps (soit 2ms de temps simulé) pour amorcer l’activité dans le réseau, ce qui permet aux neurones de la carte de focus de commencer à décharger
3. on construit l’image suivante en ajoutant tout d’abord le stimulus à la position spatiale souhaitée, puis on bruite l’image ou bien on ajoute les distracteurs,
4. l’image courante est présentée au réseau
5. le simulateur calcule 10 pas de temps (1ms de temps simulé)
6. on recommence une nouvelle itération à l’étape 3

La figure 5.15 illustre la construction d’une image et l’ajout de bruit. Lorsque l’image bruitée est présentée au réseau, on peut observer que les neurones de la carte d’entrée sont très sensibles au bruit alors que l’activité de la carte de focus reste focalisée sur le stimulus.

La robustesse du réseau est évaluée en mesurant la distance qui sépare le centre du stimulus du centre du focus d’attention. On note i l’intervalle de temps simulé pendant

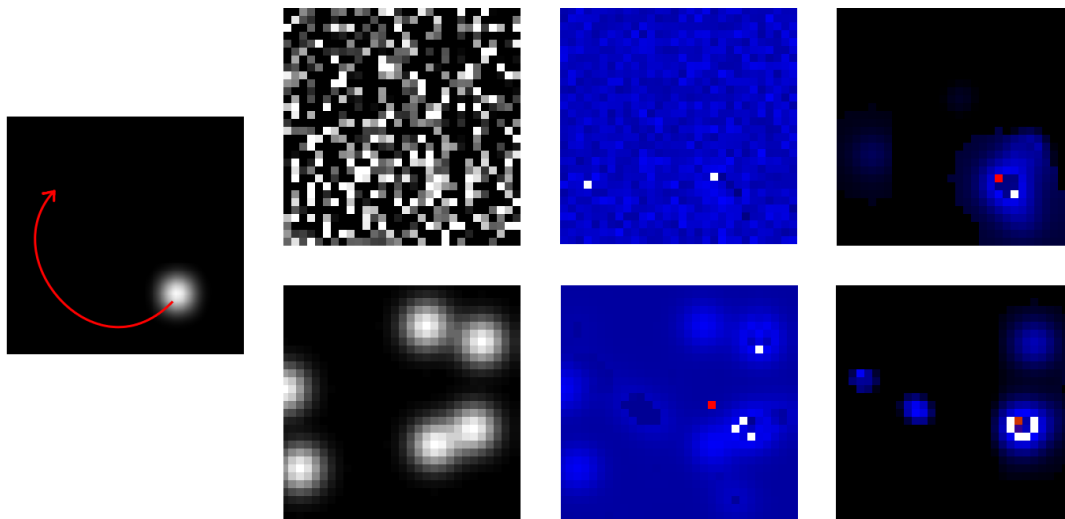


FIGURE 5.15 Construction de l'image d'entrée. À gauche, on ajoute tout d'abord le stimulus, qui est un patch gaussien et qui suit une trajectoire circulaire. On ajoute ensuite du bruit ou des distracteurs. En haut, il s'agit d'un bruit gaussien de variance $\sigma = 0,5$. En bas, on ajoute 5 distracteurs, qui sont des patches gaussiens similaires à la cible. À droite, les potentiels de membrane des neurones des cartes d'entrée et de focus pendant la simulation. Les potentiels sont représentés par des pixels colorés de la façon suivante : plus la valeur est proche du seuil, plus le pixel est bleu. Les neurones fortement inhibés, dont le potentiel est très bas, apparaissent en noir. Un pixel blanc indique que le neurone émet un potentiel d'action. Alors que les neurones des cartes d'entrée sont sensibles aux bruits et aux distracteurs, les activités des neurones des cartes de focus sont focalisées sur le stimulus et implémentent un *winner-takes-all*. Le pixel rouge indique la position de $\bar{\mathbf{X}}_i$ (voir le texte).

lequel le centre du stimulus est à la position \mathbf{S}_i , soit 10 pas de temps. Le centre du focus d'attention est déterminé comme le barycentre des potentiels d'action émis durant i :

$$\bar{\mathbf{F}}_i = \frac{1}{K} \sum_{j=1}^K \mathbf{f}_{i,j} \quad (5.14)$$

avec $\mathbf{f}_{i,j}$ la position spatiale du neurone de la carte de focus qui a émis le j^e potentiel d'action dans l'intervalle i et K le nombre total de potentiels d'action émis pendant i . Nous avons procédé de la même façon pour la carte d'entrée :

$$\bar{\mathbf{E}}_i = \frac{1}{L} \sum_{j=1}^L \mathbf{e}_{i,j} \quad (5.15)$$

avec $\mathbf{e}_{i,j}$ la position spatiale du neurone de la carte d'entrée qui a émis le j^e potentiel d'action dans l'intervalle i et L le nombre total de potentiels d'action émis pendant i .

Nous avons utilisé la moyenne des distances entre le centre du stimulus et le barycentre des potentiels d'action émis comme mesure d'erreur. L'erreur de la carte de focus est exprimée comme :

$$E_{\text{focus}} = \frac{1}{N} \sum_{i=1}^N d(\mathbf{S}_i, \bar{\mathbf{F}}_i) \quad (5.16)$$

avec N le nombre total d'intervalles i écoulés pendant la simulation. La carte de focus est toroïdale c'est pourquoi nous avons utilisé la distance suivante :

$$\begin{aligned} d(\mathbf{r}, \mathbf{r}') &= \min \left(\left(\frac{x_{\mathbf{r}} - x_{\mathbf{r}'}}{n} \right)^2, \left(1 - \frac{x_{\mathbf{r}} - x_{\mathbf{r}'}}{n} \right)^2 \right) \\ &+ \min \left(\left(\frac{y_{\mathbf{r}} - y_{\mathbf{r}'}}{n} \right)^2, \left(1 - \frac{y_{\mathbf{r}} - y_{\mathbf{r}'}}{n} \right)^2 \right) \end{aligned} \quad (5.17)$$

pour une carte de taille $n \times n$ pixels.

L'erreur $E_{\text{entrée}}$ est définie de la même façon que E_{focus} , comme la moyenne des distances entre \mathbf{S}_i et le barycentre $\bar{\mathbf{E}}_i$ des potentiels d'action émis sur la carte d'entrée :

$$E_{\text{entrée}} = \frac{1}{N} \sum_{i=1}^N d(\mathbf{S}_i, \bar{\mathbf{E}}_i) \quad (5.18)$$

5.3.2 Évaluation de la robustesse de la focalisation

La figure 5.16 montre l'évolution des barycentres de l'activité des cartes d'entrée et de focus, respectivement $\bar{\mathbf{E}}_i$ et $\bar{\mathbf{F}}_i$, pour des images bruitées avec $\sigma = 0.5$. Il apparaît

que $\bar{\mathbf{E}}_i$ est très fortement influencé par le bruit et ne suit pas du tout le déplacement du stimulus, au contraire de $\bar{\mathbf{F}}_i$ qui montre que l'activité reste focalisée sur le stimulus.

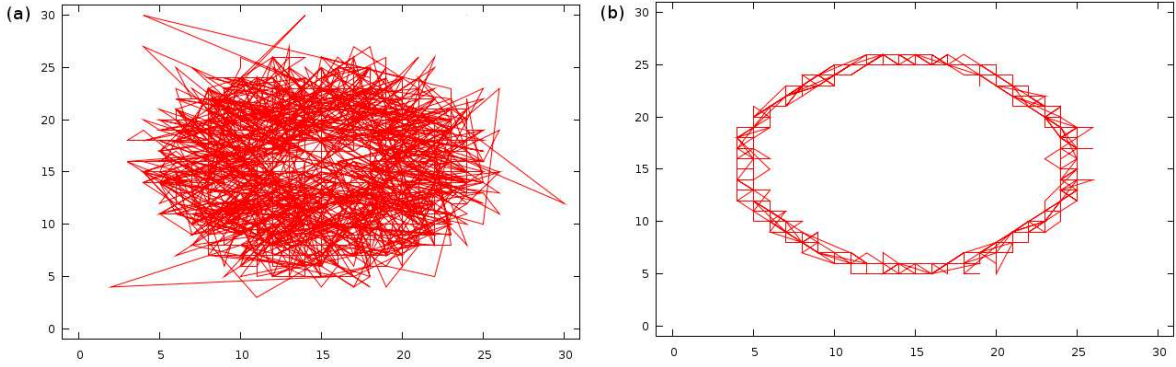


FIGURE 5.16 Représentation des barycentres de l'activité de la carte d'entrée ($\bar{\mathbf{E}}_i$, à gauche) et de la carte de focus ($\bar{\mathbf{F}}_i$, à droite). Un stimulus se déplace en suivant une trajectoire circulaire sur une image bruitée ($\sigma = 0,5$) qui est mise en entrée de la carte d'entrée. La position du stimulus est décodée sur la carte de focus. Même en présence d'un niveau de bruit important, le focus d'attention reste stable et suit le stimulus.

Les résultats obtenus sont comparables à ceux de [Rougier et Vitay, 2006], comme l'illustre la figure 5.17. Le haut de la figure reproduit les résultats obtenus avec un champ neuronal et le bas de la figure récapitule les résultats obtenus avec le réseau de neurones impulsionnels. Les histogrammes représentent les valeurs calculées de E_{focus} et $E_{\text{entrée}}$ pour 100 simulations. Les deux conditions expérimentales testées sont le suivi de la cible lorsqu'elle se déplace dans une image bruitée, avec comme valeur $\sigma = \{0, 1; 0, 25; 0, 5; 0, 75; 1\}$ (voir l'équation (5.13)) et le suivi de la cible lorsqu'elle se déplace dans une image contenant D distracteurs, avec $D = \{1; 2; 3; 5; 10; 25\}$. Pour les deux conditions expérimentales, les erreurs $E_{\text{entrée}}$ obtenues avec les neurones impulsionnels sont quasiment identiques à celles obtenues avec les champs neuronaux. Les erreurs E_{focus} pour un grand nombre de distracteurs sont très légèrement supérieures pour les neurones impulsionnels.

Il est intéressant de noter que, pour le réseau de neurones impulsionnels, les erreurs mesurées sur la carte d'entrée $E_{\text{entrée}}$ sont bien plus faibles que celles mesurées avec le champ neuronal. Ceci est imputable au fait que les neurones de la carte d'entrée réalisent un filtrage spatio-temporel de l'image : pour $\sigma = \{0; 0, 1; 0, 25\}$, $s_{r,\theta,W,I} \ll \eta$ (équation (5.12)), c'est-à-dire que l'intensité du bruit n'est pas assez importante pour amener les neurones à décharger. Il en est de même lorsque le nombre de distracteurs est faible : les distracteurs n'occupent pas suffisamment longtemps la même localisation spatiale pour amener les neurones de la carte d'entrée à décharger.

Les simulations du réseau de neurones sont rapides : les calculs nécessaires à la mise

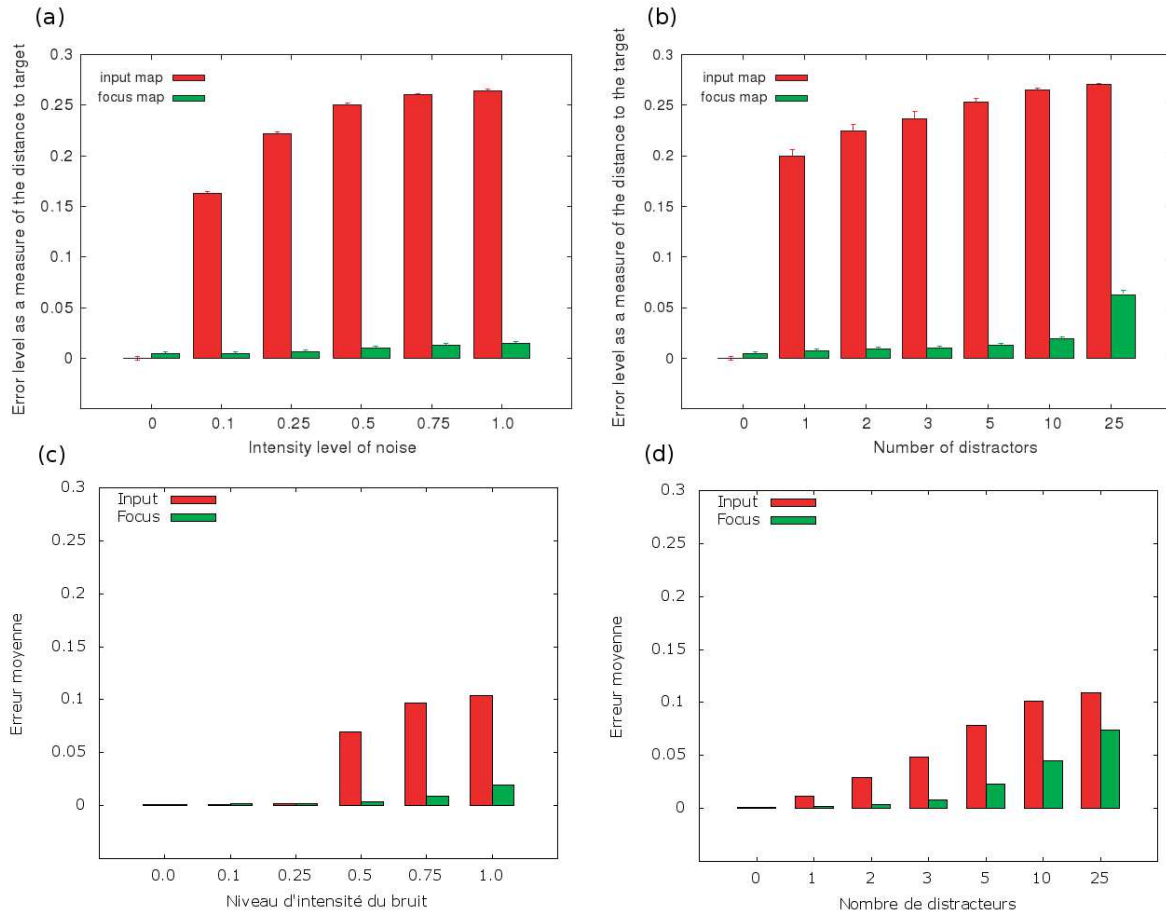


FIGURE 5.17 Comparaison de la robustesse de la focalisation attentionnelle entre les résultats obtenus par [Rougier et Vitay, 2006] et ceux obtenus avec un réseau de neurones impulsionnels. Les histogrammes (a) et (b) présentent les erreurs $E_{\text{entrée}}$ et E_{focus} mesurées dans [Rougier et Vitay, 2006] lorsque les images sont bruitées ((a), pour $\sigma = \{0, 1; 0, 25; 0, 5; 0, 75; 1\}$) ou contiennent des distracteurs ((b), pour 1, 2, 3, 4, 5, 10 ou 25 distracteurs). (c) Mêmes conditions expérimentales que (a) avec un réseau de neurones impulsionnels. (d) Mêmes conditions expérimentales que (b) avec un réseau de neurones impulsionnels. Les histogrammes de (a) et de (b) sont tirés de [Rougier et Vitay, 2006].

à jour (un pas de temps Δt) prennent entre 1 et 9 millisecondes en fonction de l'activité du réseau sur une machine équipée d'un processeur Intel Core2Duo (1,86GHz).

Nous avons donc vu dans cette partie qu'il est possible de reproduire avec un réseau de neurones impulsionnels le mécanisme de focalisation attentionnelle proposé par [Rougier et Vitay, 2006]. La partie suivante présente les résultats obtenus en utilisant l'architecture complète, c'est-à-dire en utilisant le mécanisme de focalisation attentionnelle sur la carte de saillances (voir figure 5.1).

5.3.3 Focalisation sur une séquence d'images naturelles

Nous proposons dans cette partie d'alimenter la carte de focus avec les saillances détectées par l'architecture de vision préattentionnelle. Comme cette architecture est conçue pour être intégrée à terme dans un système de perception artificiel, nous avons évalué la robustesse de la focalisation attentionnelle sur une séquence d'images naturelles.

La séquence d'images utilisée montre un robot Khepera (le stimulus) qui se déplace en partant de la partie gauche de l'image, passe derrière une pile de livres et réapparaît dans la partie droite (figure 5.18). Les 30 images de la séquence ont été acquises avec une caméra Sony EVID31 et réduites à une taille de 76x56 pixels. Le réseau utilisé ne traite que les informations de luminance. Comme le but de cette expérience est d'étudier la stabilité du focus attentionnel, l'image d'entrée est passée en négatif ce qui permet que le robot soit l'élément le plus saillant de la scène.

La mise à jour des images est similaire à celle décrite dans la partie 5.3.1 :

1. la première image est mise en entrée du réseau
2. le simulateur calcule 20 pas de temps Δt pour amorcer l'activité dans le réseau, ce qui permet aux neurones de la carte de focus de commencer à décharger
3. l'image suivante est mise en entrée du réseau
4. le simulateur calcule N pas de temps Δt
5. on recommence une nouvelle itération à partir de l'étape 3

Dans les résultats présentés sur la figure 5.19, nous avons fait varier N , le nombre de pas de temps Δt simulés à chaque itération.

Au début de la séquence d'images, le focus d'attention se stabilise rapidement sur le robot et lorsque le robot se déplace, le focus d'attention le suit. Quand le robot passe derrière la pile de livres, le focus d'attention se déplace vers un autre endroit saillant

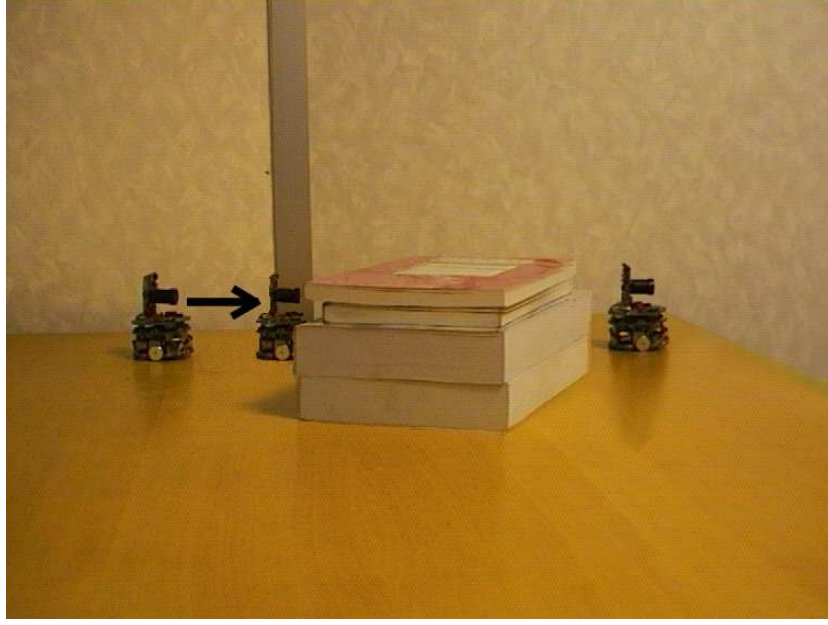


FIGURE 5.18 Cette figure récapitule la séquence d’images utilisée pour cette évaluation. Un robot Khepera se déplace dans un contexte fixe. Trois positions différentes du Khepera sont représentées sur cette image, le sens de déplacement du robot est indiqué par une flèche.

de la scène. Dès lors que le robot réapparaît de l’autre côté de la pile de livres, il est suffisamment saillant pour recapturer le focus d’attention.

Nous avons utilisé la moyenne des distances entre \mathbf{S}_i , le centre du stimulus, et $\bar{\mathbf{F}}_i$, le centre du focus d’attention, comme mesure d’erreur. Le centre du stimulus est déterminé comme le barycentre des pixels appartenant au robot et le centre du focus d’attention est calculé de la même façon que dans la partie précédente pour chaque intervalle i (voir équation (5.14)). L’erreur mesurée est :

$$E_{\text{focus}} = \frac{1}{K} \sum_{i=1}^K d(\mathbf{S}_i, \bar{\mathbf{F}}_i) \quad (5.19)$$

avec K le nombre total d’intervalles i écoulés pendant la simulation.

Les histogrammes de la figure 5.19 montrent l’erreur moyenne E_{focus} calculée pour la séquence d’images naturelles. Nous avons fait varier le nombre de pas de temps Δt simulé lors de la présentation de chaque image. On pourra constater que lorsque le réseau dispose d’un plus grand nombre de pas de temps Δt pour traiter chaque image, la réponse devient d’autant plus précise.

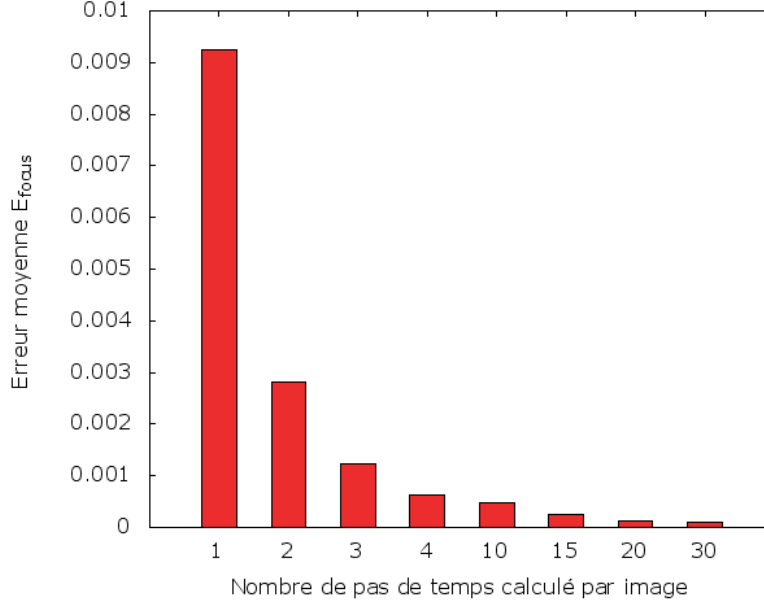


FIGURE 5.19 Les histogrammes indiquent l'erreur moyenne E_{focus} calculée pour la séquence d'images naturelles pour différentes valeurs de N , le nombre de pas de temps Δt simulés à chaque itération.

5.4 Conclusion

Nous avons détaillé dans ce chapitre une architecture neuronale qui permet d'extraire les saillances d'une image ou d'une séquence d'images. Les régions saillantes sont déterminées comme étant des contrastes locaux, globalement rares, qui combinent la présence de traits caractéristiques sur plusieurs échelles spatiales. La séparation des informations dans les hautes et basses fréquences spatiales s'inspire de la ségrégation des informations dans les voies magno- et parvocellulaire. Les informations issues des basses fréquences spatiales sont utilisées pour préactiver la carte de saillances.

L'implémentation avec un réseau de neurones impulsionnels tire parti du caractère *anytime* du filtrage neuronal présenté dans le chapitre précédent. Les saillances apparaissent dans l'ordre de leur importance et les premières régions saillantes sont déterminées avant que l'ensemble des traitements soit fini. D'autre part, la détection de saillance est réalisée par des neurones se comportant comme des détecteurs de coïncidences, ce qui permet de moduler le nombre de régions considérées comme saillantes. Cette approche donne des résultats comparables à ceux obtenus avec la méthode proposée par [Itti *et al.*, 1998] et les saillances extraites sont plus précises spatialement.

Dans la dernière partie du chapitre, nous avons proposé d'implémenter un mécanisme de *winner-takes-all* avec une carte neuronale dotée d'une connexion récurrente. L'activité qui émerge de cette carte permet de déterminer la position du focus attentionnel. Nous avons évalué la stabilité de ce focus attentionnel et les résultats expérimentaux

montrent qu'il est robuste aux bruits et aux distracteurs. Ces résultats sont globalement meilleurs que ceux obtenus avec des champs de neurones dynamiques lorsque l'on évalue la robustesse par rapport au bruit. Cette focalisation attentionnelle peut être couplée avec les sorties de la carte de saillances. Ceci permet de focaliser l'attention sur la région la plus saillante d'une scène visuelle et de suivre cette région si elle se déplace.

Chapitre 6

Conclusions et perspectives

6.1 Conclusions

Les mécanismes attentionnels ascendants, et plus particulièrement les calculs de saillances, permettent de réduire la complexité des traitements en vision artificielle. Ils offrent une heuristique pour choisir des régions de l'espace perceptif à traiter prioritairement. Ce type de mécanismes est donc particulièrement utile, par exemple, pour orienter des processus de reconnaissance d'objets ou d'amers vers les régions les plus susceptibles de contenir des informations intéressantes.

Nous avons proposé une architecture pour la vision préattentionnelle qui s'inspire du fonctionnement de la perception visuelle des systèmes biologiques, tant au niveau de la structure que des traitements réalisés. Les propriétés des neurones biologiques et les principales structures qu'ils forment dans le système visuel ainsi que les principales théories de l'attention visuelle en psychologie cognitive sont détaillées dans le chapitre 1. Il apparaît que même si les propriétés et les mécanismes mis en jeu dans certaines structures sont relativement bien cernés, nous sommes encore loin de pouvoir caractériser le fonctionnement de l'ensemble du système visuel. Ceci peut s'expliquer tout d'abord par la grande diversité des types cellulaires et des structures entrant en jeu dans la vision, ce qui rend difficile leur étude systématique. Une seconde cause concerne l'évolution des techniques expérimentales récentes, qui permettent des observations beaucoup plus fines et dans des conditions plus proches de situations réelles, dont les résultats sont très différents de ceux précédemment obtenus. Nous avons choisi de développer une approche bio-inspirée, qui tire parti des connaissances actuelles sur les mécanismes mis en jeu dans la vision sans chercher à répliquer exactement les observations expérimentales, pour proposer une implémentation informatique efficace.

Nous avons détaillé une implémentation des mécanismes préattentionnels qui repose

sur l'utilisation de neurones impulsionnels. Ce type de modèle neuronal, détaillé dans le chapitre 2, permet d'expliciter le temps d'émission précis des potentiels d'action en fonction des entrées. Les modèles impulsionnels permettent de recoder les informations sous une forme différente : temporelle et discrète. Ils proposent un compromis intéressant, permettant de capturer des dynamiques temporelles fines, sans requérir un coût computationnel aussi important que des modèles explicatifs, qui reproduisent très fidèlement les observations neurobiologiques au prix d'une complexité bien supérieure.

Les mécanismes préattentionnels ont déjà fait l'objet de plusieurs implémentations, récapitulées dans le chapitre 3, et qu'il est possible de répartir en trois catégories : les modèles neuropsychologiques, les systèmes de traitement d'images et les approches bio-inspirées. Les implémentations de modèles neuropsychologiques sont conçues pour valider les théories attentionnelles et proposer des pistes de recherches expérimentales. Elles permettent de traiter des stimulus expérimentaux mais ne sont pas particulièrement adaptées pour traiter des images naturelles. Les systèmes de traitement d'images regroupent les approches qui définissent différentes sortes de descripteurs d'images. Utilisés pour des tâches d'indexation ou de reconnaissance d'objets, ils permettent de réduire l'espace de recherche en sélectionnant une collection de points sur chaque image, ce qui réduit la complexité des traitements à effectuer pour réaliser la tâche choisie. Les approches bio-inspirées adaptent des mécanismes observés chez les organismes biologiques pour proposer des implémentations efficaces. Comparativement aux systèmes de traitement d'images qui reposent le plus souvent sur un seul trait caractéristique, elles proposent d'utiliser plusieurs traits caractéristiques et de les combiner ensuite pour déterminer les zones saillantes d'une image. Les approches bio-inspirées ont donné lieu à plusieurs implémentations, qui utilisent des techniques classiques de traitement d'images ou des champs neuronaux.

Les neurones impulsionnels permettent de traiter les informations visuelles d'une façon différente des traitements d'image classiques et offrent la possibilité d'implémenter d'une manière originale une architecture préattentionnelle et bio-inspirée. Dans le chapitre 4, nous avons présenté une méthode neuronale de filtrage d'images qui permet d'obtenir le résultat d'un filtrage par convolution par approximations successives. Les valeurs (de luminance par exemple) des pixels sont traduites en trains de potentiels d'action par les neurones impulsionnels de la carte d'entrée, ce qui nous permet d'encoder la luminance sous une forme temporelle. La luminance est utilisée pour déterminer les temps d'émission des potentiels d'action, ce qui définit implicitement l'ordre utilisé pour traiter les pixels : les valeurs de luminance les plus élevées sont traitées les premières. Les potentiels d'action d'entrée sont envoyés à une carte de filtrage au travers d'un masque de connexion qui fixe les poids synaptiques à partir de la réponse impul-

sionnelle du filtre utilisé. Les potentiels d'action émis par les neurones de la carte de filtrage sont interprétés pour obtenir le résultat du filtrage. Pour obtenir rapidement une première approximation nous avons proposé d'interpréter ces potentiels d'actions avec un codage impulsionnel. Ce codage ne tient compte que du premier potentiel d'action émis pour chaque neurone et consiste à ordonner ces potentiels d'action en fonction de leur instant d'émission. Pour obtenir une approximation plus précise du filtrage nous avons utilisé le codage fréquentiel, qui requiert un temps de traitement plus important car les neurones doivent avoir émis plusieurs potentiels d'actions. Pour interpréter les potentiels d'action avec un codage fréquentiel, nous avons mesuré l'intervalle inter-spike moyen de chaque neurone. L'approximation du filtrage est d'autant plus précise que le nombre de potentiels d'action considérés est important.

Le premier intérêt de cette approche neuronale est de permettre, avec une implémentation adéquate, de traiter uniquement les pixels suffisamment lumineux (ou colorés). Le nombre de calculs requis dans notre implémentation est fonction du nombre de potentiels d'action émis. Les valeurs de luminance (ou de couleur) inférieures à un seuil paramétrable ne suscitent pas de potentiels d'action, ce qui réduit donc le coût computationnel. L'apport principal du filtrage neuronal est d'implémenter un processus *anytime*. Le filtrage neuronal produit une approximation du filtrage par convolution qui est d'autant plus précise que l'algorithme dispose de temps processeur. Plus le système peut calculer de temps simulé, plus les neurones émettent de potentiels d'action, ce qui permet d'obtenir une approximation plus précise du filtrage. Le codage impulsionnel permet d'obtenir rapidement un premier résultat et le codage fréquentiel fournit une approximation de meilleure qualité. Il est ainsi possible d'obtenir un résultat à tout instant. Ceci est particulièrement intéressant lorsque l'on combine les résultats de plusieurs opérations de filtrage, que ce soit avec des filtres différents ou à des échelles spatiales différentes, car il est possible d'obtenir une approximation du résultat final de façon précoce. Cette approche permet d'échanger de la précision de résultat contre du temps processeur.

Nous avons proposé de tirer parti de ces propriétés du filtrage neuronal dans l'architecture préattentionnelle (chapitre 5), qui extrait les saillances d'une image. Dans cette architecture, la luminance et les couleurs d'une image d'entrée sont utilisées pour réaliser plusieurs filtres qui permettent de décomposer cette image selon différentes modalités visuelles (contrastes de luminance, de couleurs et d'orientations) pour deux échelles spatiales, en s'inspirant des traitements réalisés par les voies parvo- et magno-cellulaire. Les contrastes locaux obtenus dans chaque modalité sont combinés sur une carte de saillances. La détermination des régions saillantes repose sur un mécanisme qui exploite le comportement de détecteur de coïncidences des neurones impulsionnels et

qui ne pourrait pas actuellement être implémenté autrement.

Le mécanisme d'extraction de saillance présente plusieurs propriétés intéressantes. Tout d'abord, les contrastes localisés dans les basses fréquences spatiales préactivent la carte de saillances, ce qui sélectionne un nombre restreint de régions sur lesquelles la recherche de saillances est effectuée. La recherche de saillance est concentrée sur les régions qui portent des informations dans les deux fréquences spatiales. Ceci permet de faire ressortir des différences fines dans chaque région sélectionnée : les saillances obtenues sont spatialement plus précises que celles obtenues avec l'algorithme d'Itti. Ensuite, les neurones de la carte de saillances sont paramétrés pour détecter les coïncidences dans les trains de potentiels d'action ; il est ainsi possible de choisir la sélectivité de ces neurones, celle-ci agit comme une sorte de gain permettant de réduire ou d'étendre les régions considérées comme saillantes. Enfin, cette architecture implémente un processus *anytime* qui permet rapidement d'obtenir des saillances sommaires puis de raffiner le résultat pour obtenir des valeurs de plus en plus précises pour chaque pixel.

L'architecture que nous avons décrite réalise une focalisation attentionnelle : les saillances extraites alimentent un réseau récurrent qui implémente un mécanisme de *winner-takes-all*. Ce mécanisme est utilisé pour fixer le focus d'attention sur la région la plus saillante et pour maintenir cette fixation lorsque la région saillante se déplace. Le principe de fonctionnement est semblable à la mise en compétition locale réalisée par un champ neuronal. Les résultats expérimentaux montrent que la robustesse aux distracteurs est aussi bonne que celle obtenue avec un champ neuronal et que la robustesse aux bruits est globalement meilleure que celle des champs neuronaux.

L'architecture préattentionnelle a été conçue dans le cadre d'une réflexion sur la perception située, où la vision est définie comme le résultat d'une constante interaction entre la perception et l'action. Les traitements réalisés sur les entrées perceptives doivent donc être relativement souples pour s'adapter à la dynamique des interactions, par exemple pour donner rapidement des indications sommaires lorsqu'il faut déterminer un contexte ou au contraire affiner les résultats des traitements lorsqu'il faut caractériser une région perceptive pour l'apprendre ou la reconnaître. L'implémentation avec des neurones impulsionnels permet d'extraire très rapidement de la scène visuelle une première information, éventuellement assez grossière mais significative, puis de raffiner de plus en plus la précision de la prise d'information au cours des instants qui suivent. Elle offre beaucoup de souplesse au processus de prise de décision, qui peut ainsi moduler sa réponse en fonction de critères temporels ou qualitatifs.

6.2 Perspectives

Les travaux expérimentaux présentés dans ce manuscrit mettent en évidence les propriétés temporelles du filtrage neuronal et montrent qu'il est tout à fait possible d'obtenir des résultats équivalents à ceux utilisant des méthodes de convolution. Nous avons caractérisé mathématiquement la partie utilisant le codage impulsionnel et nous travaillons à étendre ce résultat au codage fréquentiel dans l'optique de pouvoir proposer une comparaison plus approfondie avec le filtrage par convolution. Ces résultats théoriques permettent d'orienter les choix algorithmiques pour améliorer l'efficacité de l'implémentation utilisée. D'autre part, nous avons commencé une étude expérimentale utilisant une base d'images beaucoup plus large sur laquelle nous pourrions utiliser des mesures statistiques de distance pour produire une évaluation des différences entre filtrage neuronal et filtrage par convolution qualitativement différente de celle proposée dans ce manuscrit.

L'approfondissement des apports des cartes récurrentes impulsionnelles comparativement aux champs neuronaux constitue le second axe de poursuite à court terme de ce travail. Si l'on considère que les champs neuronaux décrivent une activité qui est liée à la fréquence de décharge, il est intéressant d'explorer les apports du codage impulsionnel pour des tâches sur lesquelles sont utilisés les champs neuronaux. Il s'agit tout d'abord d'évaluer les différences qualitatives entre neurones impulsionnels et champs neuronaux sur la dynamique de l'activité du réseau. Dans un second temps, il sera intéressant d'examiner les apports possibles du codage impulsionnel pour désambiguïser des situations de compétitions locales entre plusieurs régions actives.

Cette architecture neuronale pourra être intégrée à terme sur une plateforme robotique dotée d'un système de perception visuelle. L'architecture sera une étape de filtrage des entrées visuelles qui permettra de concentrer les traitements de plus haut niveau sur les régions saillantes. Pour les traitements de plus haut niveau, il pourra être intéressant d'utiliser une approche du type *reservoir computing*, qui peut être implémentée avec des neurones impulsionnels : ceci permettrait de conserver une représentation de l'information sous une forme unique (les potentiels d'action) dans le système. L'intégration d'influence descendante sur l'architecture de vision préattentionnelle serait ainsi facilitée. Par exemple, il est envisageable d'utiliser un retour proprioceptif pour biaiser la détection de saillance, de proposer un modèle d'inhibition de retour prenant en compte des modulations descendantes ou bien de modifier directement l'activité des cartes d'entrée en fonction d'informations contextuelles.

Publications

Conférences internationales avec comité de relecture

- S. Chevallier and P. Tarroux. Covert attention with a spiking neural network. In A. Gasteratos, M. Vincze, and J.K. Tsotsos, editors, *Int. Conf. on Computer Vision Systems (ICVS)*, volume 5008 of *Lecture Notes in Computer Science*, pages 56–65. Springer, 2008.
- S. Chevallier and P. Tarroux. Visual focus with spiking neurons. In M. Verleysen, editor, *European Symposium on Artificial Neural Networks (ESANN)*, pages 385–390, 2008.
- S. Chevallier, P. Tarroux, and H. Paugam-Moisy. Saliency extraction with a distributed spiking neural network. In M. Verleysen, editor, *European Symposium on Artificial Neural Networks (ESANN)*, pages 209–214, Bruges, Belgium, 2006.
- S. Chevallier, H. Paugam-Moisy, and F. Lemaître. Distributed processing for modelling real-time multimodal perception in a virtual robot. In *Int. Conf. on Parallel and Distributed Computing and Networks (PDCN)*, pages 393–398, Innsbruck, Austria, 2005. ACTA Press.

Conférence nationale avec comité de relecture

- S. Chevallier and P. Tarroux. Implémentation d’un mécanisme de “covert attention” avec un réseau de neurones impulsionnels. In L. Perrinet and E. Dauce, editors, *NeuroComp’08 – conférence française de Neurosciences Computationnelles*, pages 78–81, Marseille, France, Octobre 2008.
- S. Chevallier and P. Tarroux. Extraction de saillances par un réseau de neurones impulsionnels. In *NeuroComp’06 – conférence française de Neurosciences Computationnelles*, pages 154–157, Nancy, France, 2006.

Conférence nationale de jeunes chercheurs

- S. Chevallier and H. Paugam-Moisy. Modélisation de processus attentionnels dans la perception multimodale d'un robot virtuel. In *Actes du VIème Colloque Jeunes Chercheurs en Sciences Cognitives*, Bordeaux, France, 2005.

Bibliographie

- [Abbott et Nelson, 2000] L.F. Abbott et S.B. Nelson. Synaptic plasticity : taming the beast. *Nature Neuroscience*, 3 Suppl :1178–1183, Novembre 2000.
- [Abbott, 1999] L.F. Abbott. Lapique’s introduction of the integrate-and-fire model neuron (1907). *Brain Research Bulletin*, 50(5-6) :303–304, 1999.
- [Abu-Mostafa et Schweizer, 1990] Y. Abu-Mostafa et D. Schweizer. Neural networks. In *VLSI and Parallel Computation*, éditeurs R. Suaya et G. Birtwistle, pages 390–415. Morgan Kaufmann, 1990.
- [Abu-Mostafa, 1989] Y. Abu-Mostafa. Information theory, complexity and neural networks. *IEEE Communications Magazine*, 27(11) :312–317, 1989.
- [Ahrns et Neumann, 1999] I. Ahrns et H. Neumann. Space-variant dynamic neural fields for visual attention. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 2, page 318. IEEE Computer Society, 1999.
- [Alexandre, 1997] F. Alexandre. *Intelligence Neuromimétique*. Habilitation à diriger des recherches, Université Henri Poincaré, Nancy, France, 1997.
- [Alleysson *et al.*, 2005] D. Alleysson, S. Susstrunk, et J. Hérault. Linear demosaicing inspired by the human visual system. *IEEE Transactions on Image Processing*, 14(4) :439–449, 2005.
- [Amari, 1977] S-I. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2) :77–87, Juin 1977.
- [Amit et Tsodyks, 1991] D.J. Amit et M. Tsodyks. Quantitative study of attractor neural networks retrieving at low spike rates : li. low-rate retrieval in symmetric networks. *Network : Computation in Neural Systems*, pages 275–294, Aout 1991.
- [Arbib, 1998] éditeur M.A. Arbib. *The Handbook of Brain Theory and Neural Networks*. MIT Press, Juillet 1998.
- [Baluja et Pomerleau, 1997] S. Baluja et D.A. Pomerleau. Expectation-based selective attention for visual monitoring and control of a robot vehicle. *Robotics and Autonomous Systems*, pages 329–344, Décembre 1997.
- [Barlow et Levick, 1965] H.B. Barlow et W.R. Levick. The mechanism of directionally selective units in rabbit’s retina. *The Journal of physiology*, 178(3) :477–504, Juin 1965.
- [Beaudot *et al.*, 1993] W. Beaudot, P. Palagi, et J. Hérault. Realistic simulation tool for early visual processing including space, time and colour data. In *Proc. of the Int. Workshop on Artificial Neural Networks (IWANN)*, pages 370–375, London, UK, 1993. Springer-Verlag.

- [Beaudot, 1994] W. Beaudot. *Le traitement neuronal de l'information dans la rétine des vertébrés : un creuset d'idées pour la vision artificielle*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, Grenoble, France, Décembre 1994.
- [Benoit, 2007] A. Benoit. *Le système visuel humain au secours de la vision par ordinateur*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, Grenoble, France, Février 2007.
- [Bi et Poo, 1998] G.Q. Bi et M.M. Poo. Synaptic modifications in cultured hippocampal neurons : dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24) :10464–10472, Décembre 1998.
- [Bi et Poo, 2001] G.Q. Bi et M.M. Poo. Synaptic modification by correlated activity : Hebb's postulate revisited. *Annual Review of Neuroscience*, 24 :139–166, 2001.
- [Bi, 2002] G.Q. Bi. Spatiotemporal specificity of synaptic plasticity : cellular rules and mechanisms. *Biological Cybernetics*, 87(5-6) :319–332, Décembre 2002.
- [Bishop, 1995] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [Boddy et Dean, 1989] M. Boddy et T.L. Dean. Solving time-dependent planning problems. In *Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 1989.
- [Boddy et Dean, 1994] M. Boddy et T.L. Dean. Deliberation scheduling for problem solving in time-constrained environments. *Artificial Intelligence*, 67(2) :245–285, 1994.
- [Bohte, 2004] S. Bohte. The evidence for neural information processing with precise spike-times : A survey. *Natural Computing*, 3 :195–206, 2004.
- [Bower et Beeman, 1998] J.M. Bower et D. Beeman. *The book of GENESIS : exploring realistic neural models with the GEneral NEural SIMulation System*. Springer, New York, NY, USA, 2e édition, 1998.
- [Brette et al., 2007] R. Brette, M. Rudolph, N.T. Carnevale, M.L. Hines, D. Beeman, J. Bower, M. Diesmann, A. Morrison, P. Goodman, F. Harris Jr, M. Zirpe, T. Natschläger, D. Pecevski, G.B. Ermentrout, M. Djurfeldt, A. Lansner, O. Rochel, T. Viéville, E. Muller, A.P. Davison, S. El Boustani, et A. Destexhe. Simulation of networks of spiking neurons : A review of tools and strategies. *Journal of Computational Neuroscience*, 23(3) :349–98, 2007.
- [Brette, 2006] R. Brette. Exact simulation of integrate-and-fire models with synaptic conductances. *Neural Computation*, 18(8) :2004–2027, 2006.
- [Brette, 2007] R. Brette. Exact simulation of integrate-and-fire models with exponential currents. *Neural Computation*, 19(10) :2604–2609, Octobre 2007.
- [Burr et al., 1994] D.C. Burr, M.C. Morrone, et J. Ross. Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, 371(6497) :511–513, Octobre 1994.
- [Burt et Adelson, 1983] P. Burt et E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4) :532–540, 1983.
- [Carmi et Itti, 2006] R. Carmi et L. Itti. The role of memory in guiding attention during natural vision. *Journal of vision*, 6(9) :898–914, 2006.
- [Catterall, 1988] W.A. Catterall. Structure and function of voltage-sensitive ion channels. *Science*, 242 :50–61, 1988.

- [Cave, 1999] K.R. Cave. The featuregate model of visual selection. *Psychological Research*, 62(2-3) :182–194, 1999.
- [Cessac et Samuelides, 2006] B. Cessac et M. Samuelides. From neuron to neural networks dynamics. *European Physical Journal Special Topics*, 142 :7–88, Septembre 2006.
- [Chalupa et Werner, 2003] éditeurs L.M. Chalupa et J.S. Werner. *The Visual Neurosciences*. MIT Press, 1ère édition, Novembre 2003.
- [Chariot *et al.*, 2006] A. Chariot, R. Keriven, et R. Brette. Simulation rapide de modèles de neurones impulsionnels sur carte graphique. In *NeuroComp 1ère conférence française de Neurosciences Computationnelles*, pages 150–153, Nancy, France, 2006.
- [Churchland et Sejnowski, 1994] P.S. Churchland et T.J. Sejnowski. *The computational brain*. MIT Press, Cambridge, USA, 1994.
- [Churchland, 1986] P.S. Churchland. *Neurophilosophy : Toward a Unified Science of the Mind-brain*. MIT Press, Cambridge, USA, 1986.
- [Crick, 1984] F. Crick. Function of the thalamic reticular complex : the searchlight hypothesis. *Proceedings of the National Academy of Sciences*, 81(14) :4586–4590, Juillet 1984.
- [Cuperlier *et al.*, 2007] N. Cuperlier, M. Quoy, et P. Gaussier. Neurobiologically inspired mobile robot navigation and planning. *Frontiers in Neurobotics*, 1(3), Novembre 2007.
- [Dan et Poo, 2004] Y. Dan et M.M. Poo. Spike timing-dependent plasticity of neural circuits. *Neuron*, 44(1) :23–30, Septembre 2004.
- [Daugman et Downing, 1998] J. Daugman et C. Downing. Gabor wavelets for statistical pattern recognition. In Arbib [1998], pages 414–420.
- [de Brecht et Saiki, 2006] M. de Brecht et J. Saiki. A neural network implementation of a saliency map model. *Neural Networks*, 19(10) :1467–1474, Décembre 2006.
- [Dean et Boddy, 1988] T.L. Dean et M. Boddy. An analysis of time-dependent planning. In *Proc. of the AAAI*, AAAI, pages 49–54, 1988.
- [Deco *et al.*, 2002] G. Deco, O. Pollatos, et J. Zihl. The time course of selective visual attention : theory and experiments. *Vision Research*, 42(27) :2925–2945, Décembre 2002.
- [Deco et Zihl, 2001] G. Deco et J. Zihl. Top-down selective visual attention : A neurodynamical approach. *Visual Cognition*, pages 118–139, Février 2001.
- [Delorme *et al.*, 1999] A. Delorme, J. Gautrais, R. VanRullen, et S.J. Thorpe. Spikenet : a simulator for modeling large networks of integrate-and-fire neurons. *Neurocomputing*, 26-27 :989–996, 1999.
- [Desimone et Duncan, 1995] R. Desimone et J. Duncan. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18 :193–222, 1995.
- [Destexhe *et al.*, 1994] A. Destexhe, Z.F. Mainen, et T.J. Sejnowski. An efficient method for computing synaptic conductances based on a kinetic model of receptor binding. *Neural Computation*, 6(1) :14–18, Janvier 1994.
- [Diesmann et Gewaltig, 2003] M. Diesmann et M-O. Gewaltig. NEST : An environment for neural systems simulations. In *Forschung und wissenschaftliches Rechnen, Beiträge zum Heinz-Billing-Preis*, éditeur T. Plesser, numéro 58 dans GWDG Bericht. Ges. für Wiss. Datenverarbeitung, Göttingen, Germany, 2003.

- [Douglas et Martin, 2004] R.J. Douglas et K.A. Martin. Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, 27 :419–451, 2004.
- [Dreyfus *et al.*, 2002] G. Dreyfus, M. Samuelides, J.-M. Martinez, M.B. Gordon, F. Bar-dran, S. Thiria, et L. Hérault. *Réseaux de neurones : Méthodologies et applications*. Eyrolles, 2002.
- [Duncan et Humphreys, 1989] J. Duncan et G.W. Humphreys. Visual search and stimulus similarity. *Psychological Review*, 96(3) :433–458, Juillet 1989.
- [Duncan, 1989] J. Duncan. Boundary conditions on parallel processing in human vision. *Perception*, 18(4) :457–469, 1989.
- [Edeline, 1996] J.-M. Edeline. Does Hebbian synaptic plasticity explain learning-induced sensory plasticity in adult mammals ? *Journal of Physiology (Paris)*, 90(3-4) :271–276, 1996.
- [Enroth-Cugell et Robson, 1966] C. Enroth-Cugell et J.G. Robson. The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, 187(3) :517–552, Décembre 1966.
- [Ermentrout et Kopell, 1986] G.B. Ermentrout et N. Kopell. Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM Journal on Applied Mathematics*, 46(2) :233–253, 1986.
- [Ermentrout, 1996] G.B. Ermentrout. Type i membranes, phase resetting curves, and synchrony. *Neural Computation*, 8(5) :979–1001, Juillet 1996.
- [Fecteau et Munoz, 2006] J.H. Fecteau et D.P. Munoz. Saliency, relevance, and firing : a priority map for target selection. *Trends in Cognitive Sciences*, 10(8) :382–390, 2006.
- [Feldman et Cowan, 1975] J. Feldman et J.D. Cowan. Large-scale activity in neural nets i : Theory with application to motoneuron pool responses. *Biological Cybernetics*, 17(1) :29–38, Janvier 1975.
- [Felleman et Van Essen, 1991] D.J. Felleman et D.C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1) :1–47, 1991.
- [Feng, 2001] J. Feng. Is the integrate-and-fire model good enough ?—a review. *Neural Networks*, 14(6-7) :955–975, Juillet 2001.
- [Feng, 2004] éditeur J. Feng. *Computational Neuroscience : A Comprehensive Approach*. Chapman & Hall/CRC, 2004.
- [Fink *et al.*, 1997] G. Fink, R. Dolan, P. Halligan, J. Marshall, et C. Frith. Space-based and object-based visual attention : shared and specific neural domains. *Brain*, 120(11) :2013–2028, Novembre 1997.
- [Fix *et al.*, 2007] J. Fix, J. Vitay, et N.P. Rougier. A distributed computational model of spatial memory anticipation during a visual search task. In *Anticipatory Behavior in Adaptive Learning Systems*, Lecture Notes in Artificial Intelligence, pages 170–188, 2007.
- [Fix, 2007] J. Fix. Saliency. 2007. Disponible en ligne : <http://jeremy.fix.free.fr/software.php?soft=Saliency>.
- [Fix, 2008] J. Fix. *Mécanismes numériques et distribués de l'anticipation motrice*. Thèse de Doctorat, Université Henri Poincaré, Nancy, France, Octobre 2008.

- [Fouquier *et al.*, 2008] G. Fouquier, J. Atif, et I. Bloch. Incorporating a pre-attention mechanism in fuzzy attribute graphs for sequential image segmentation. In *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 840–847, 2008.
- [Fourcaud-Trocmé *et al.*, 2003] N. Fourcaud-Trocmé, D. Hansel, C. van Vreeswijk, et N. Brunel. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *Journal of Neuroscience*, 23(37) :11628–11640, Décembre 2003.
- [Fourcaud-Trocmé, 2003] N. Fourcaud-Trocmé. *Fonction de transfert neuronale en présence de bruit*. Thèse de Doctorat, Université Paris VI, Paris, France, 2003.
- [Freeman et Skapura, 1991] J.A. Freeman et D.M. Skapura. *Neural networks : algorithms, applications, and programming techniques*. Addison Wesley, 1991.
- [Frégnac *et al.*, 2005] Y. Frégnac, P. Baudot, M. Levy, et O. Marre. An intracellular view of time coding and sparseness of cortical representation in V1 neurons during virtual oculomotor exploration of natural scenes. In *Computational and Systems Neuroscience (CoSyNe)*, page 17, 2005.
- [Frintrop *et al.*, 2006] S. Frintrop, P. Jensfelt, et H.I. Christensen. Attentional landmark selection for visual slam. In *Int. Conf. on Intelligent Robots and Systems*, pages 2582–2587. IEEE Computer Society, 2006.
- [Frintrop, 2005] S. Frintrop. *VOCUS : A Visual Attention System for Object Detection and Goal-directed Search*. Thèse de Doctorat, University of Bonn, Bonn, Germany, Juillet 2005.
- [Fukushima, 1980] K. Fukushima. Neocognitron : a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4) :193–202, 1980.
- [Gaussier *et al.*, 2003] P. Gaussier, P. Andry, J-P. Banquet, M. Quoy, J. Nadel, et A. Revel. Robots as models of the brain : What can we learn from modelling rat navigation and infant imitation games ? In *Conference on Artificial Intelligence in Medicine in Europe (AIME)*, Lecture Notes in Artificial Intelligence, pages 377–385, 2003.
- [Gerstner et Kistler, 2002] W. Gerstner et W. Kistler. *Spiking Neuron Models : Single Neurons, Population, Plasticity*. Cambridge University Press, New York, NY, USA, 2002.
- [Gerstner et van Hemmen, 1992] W. Gerstner et J. van Hemmen. Associative memory in a network of ‘spiking’ neurons. *Network : Computation in Neural Systems*, 3(2) :139–164, 1992.
- [Gerstner, 2000] W. Gerstner. Population dynamics of spiking neurons : fast transients, asynchronous states, and locking. *Neural Computation*, 12(1) :43–89, Janvier 2000.
- [Gilbert et Wiesel, 1989] C.D. Gilbert et T.N. Wiesel. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, 9(7) :2432–2442, Juillet 1989.
- [Goldman, 1953] D.E. Goldman. Potential impedance, and rectification in membranes. *Journal of General Physiology*, 27 :37–60, 1953.
- [Gollisch et Meister, 2008a] T. Gollisch et M. Meister. Modeling convergent on and off pathways in the early visual system. *Biological Cybernetics*, 99(4) :263–278, November 2008.

- [Gollisch et Meister, 2008b] T. Gollisch et M. Meister. Rapid neural coding in the retina with relative spike latencies. *Science*, 319(5866) :1108–1111, February 2008.
- [Goodman et Brette, 2008] D. Goodman et R. Brette. Brian : a simulator for spiking neural networks in python. *Frontiers in Neuroinformatics*, 2, 2008.
- [Gottlieb, 2007] J. Gottlieb. From thought to action : The parietal cortex as a bridge between perception, action, and cognition. *Neuron*, 53(1) :9–16, Janvier 2007.
- [Grassmann et Anlauf, 1998] C. Grassmann et J. Anlauf. Distributed, event driven simulation of spiking neural networks. In *NC’98 - Int. ICSC/IFAC Symp. on Neural Computation*, pages 100–105, 1998.
- [Gray et Singer, 1989] C.M. Gray et W. Singer. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*, 86(5) :1698–1702, Mars 1989.
- [Guillaume, 2009] H. Guillaume. *Explorer pour reconnaître : approche probabiliste pour la reconnaissance visuelle de lieux par un robot mobile*. Thèse de Doctorat, Université Paris-Sud - LIMSIS, 2009. À paraître.
- [Guillery et Sherman, 2002] R.W. Guillery et S.M. Sherman. Thalamic relay functions and their role in corticocortical communication : generalizations from the visual system. *Neuron*, 33(2) :163–175, Janvier 2002.
- [Hammarlund et Ekeberg, 1998] P. Hammarlund et O. Ekeberg. Large neural network simulations on multiple hardware platforms. *Journal of Computational Neuroscience*, 5(4) :443–459, Décembre 1998.
- [Harris et Stephens, 1988] C. Harris et M. Stephens. A combined corner and edge detector. In *Fourth Alvey Vision Conference*, pages 147–151, 1988.
- [Harris, 1993] C. Harris. Geometry from visual motion. In *Active Vision*, pages 263–284. MIT Press, Cambridge, MA, USA, 1993.
- [Haykin, 1994] S. Haykin. *Neural Networks : A Comprehensive Foundation*. Prentice Hall, 1994.
- [Healey, 2007] C.G. Healey. Perception in visualization. 2007. Disponible en ligne : <http://www.csc.ncsu.edu/faculty/healey/PP/index.html>.
- [Hebb, 1949] D.O. Hebb. *The organization of behavior*. John Wiley & Sons Inc, New York, NYC, 1949.
- [Heinke et Humphreys, 2003] D. Heinke et G.W. Humphreys. Attention, spatial representation, and visual neglect : simulating emergent attention and spatial memory in the selective attention for identification model (saim). *Psychological Review*, 110(1) :29–87, Janvier 2003.
- [Heinke et Humphreys, 2005] D. Heinke et G.W. Humphreys. Computational models of visual selective attention : A review. In *Connectionist Models in Cognitive Psychology*, pages 273–312. Routledge, 2005.
- [Hendry et Reid, 2000] S.H. Hendry et R.C. Reid. The koniocellular pathway in primate vision. *Annual Review of Neuroscience*, 23 :127–153, 2000.
- [Hérault, 1999] J. Hérault. *Rétine et cortex visuel : formalisation et application au traitement des images*, chapitre 17. Hermès, 1999.
- [Hines et Carnevale, 1997] M.L. Hines et N.T. Carnevale. The neuron simulation environment. *Neural Computation*, 9(6) :1179–1209, 1997.

- [Hines et Carnevale, 2004] M.L. Hines et N.T. Carnevale. Discrete event simulation in the neuron environment. *Neurocomputing*, 58-60 :1117–1122, Juin 2004.
- [Hirsch et Martinez, 2006] J.A. Hirsch et L.M. Martinez. Circuits that build visual cortical receptive fields. *Trends in Neurosciences*, 29(1) :30–39, Janvier 2006.
- [Hodgkin *et al.*, 1952] A.L. Hodgkin, A.F. Huxley, et B. Katz. Measurements of current-voltage relations in the membrane of the giant axon of loligo. *Journal of Physiology (Londres)*, 116 :424–448, 1952.
- [Hodgkin et Huxley, 1952a] A.L. Hodgkin et A.F. Huxley. The components of membrane conductance in the giant axon of loligo. *Journal of Physiology (Londres)*, 116 :473–496, 1952.
- [Hodgkin et Huxley, 1952b] A.L. Hodgkin et A.F. Huxley. Currents carried by sodium and potassium ions through the membrane of the giant axon of loligo. *Journal of Physiology (Londres)*, 116 :449–472, 1952.
- [Hodgkin et Huxley, 1952c] A.L. Hodgkin et A.F. Huxley. The dual effect of membrane potential on sodium conductance in the giant axon of loligo. *Journal of Physiology (Londres)*, 116 :497–506, 1952.
- [Hodgkin et Huxley, 1952d] A.L. Hodgkin et A.F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology (Londres)*, 116 :507–544, 1952.
- [Hong et Brady, 2003] B.-W. Hong et M. Brady. A topographic representation for mammogram segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, pages 730–737. Springer, 2003.
- [Hopfield, 1984] J.J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10) :3088–3092, Mai 1984.
- [Horvitz, 1987] E.J. Horvitz. Reasoning about beliefs and actions under computational resource constraints. In *Proc. of the Workshop on Uncertainty in Artificial Intelligence*, pages 429–444, 1987.
- [Hubel et Wiesel, 1960] D.H. Hubel et T.N. Wiesel. Receptive fields of optic nerve fibres in the spider monkey. *The Journal of physiology*, 154 :572–580, Décembre 1960.
- [Hubel et Wiesel, 1962] D.H. Hubel et T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *Journal of Physiology*, 160 :106–154, Janvier 1962.
- [Hubel et Wiesel, 1968] D.H. Hubel et T.N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195(1) :215–243, Mars 1968.
- [Hubel, 1988] D.H. Hubel. *Eye, Brain, and Vision*. W.H. Freeman and Compagny, 1988.
- [Humphreys et Müller, 1993] G.W. Humphreys et H.J. Müller. Search via recursive rejection (serr) : A connectionist model of visual search. *Cognitive Psychology*, 25(1) :43–110, Janvier 1993.
- [Humphreys, 1998] G.W. Humphreys. Neural representation of objects in space : a dual coding account. *Philosophical Transactions of the Royal Society London : B Biological sciences*, 353(1373) :1341–1351, Aout 1998.

- [Itti *et al.*, 1998] L. Itti, C. Koch, et E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 20(11) :1254–1259, Novembre 1998.
- [Itti *et al.*, 2005] éditeurs L. Itti, G. Rees, et J.K. Tsotsos. *Neurobiology of Attention*. Elsevier, San Diego, USA, 2005.
- [Itti et Koch, 2000] L. Itti et C. Koch. A saliency based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12) :1489–1506, 2000.
- [Itti et Koch, 2001] L. Itti et C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3) :194–203, Mars 2001.
- [Itti, 2000] L. Itti. *Models of Bottom-Up and Top-Down Visual Attention*. Thèse de Doctorat, California Institute of Technology, Pasadena, USA, Janvier 2000.
- [Itti, 2002] L. Itti. iLab Neuromorphic Vision C++ Toolkit (iNVT). 2002. Disponible en ligne : <http://ilab.usc.edu/toolkit/>.
- [Itti, 2004] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, 13(10) :1304–1318, 2004.
- [Itti, 2007] L. Itti. Visual salience. *Scholarpedia*, 2(9) :3327, 2007.
- [Izhikevich, 2003] E.M. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6) :1569–1572, 2003.
- [Izhikevich, 2004] E.M. Izhikevich. Which model to use for cortical spiking neurons ? *IEEE Transactions on Neural Networks*, 15(5) :1063–1070, 2004.
- [James, 1890] W. James. *The Principles of Psychology, Vol. 1*. Dover Publications, 1890.
- [Jones *et al.*, 1987] J.P. Jones, A. Stepnoski, et L.A. Palmer. The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6) :1212–1232, Décembre 1987.
- [Jones et Palmer, 1987] J.P. Jones et L.A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6) :1233–1258, 1987.
- [Julész, 1981] B. Julész. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics*, 41(2) :131–138, 1981.
- [Julész, 1986] B. Julész. Texton gradients : The texton theory revisited. *Biological Cybernetics*, 54(4) :245–251, 1986.
- [Kandel *et al.*, 2000] E.R. Kandel, J.H. Schwartz, et T.M. Jessell. *Principles of Neural Science*. McGraw-Hill Medical, Janvier 2000.
- [Kepler *et al.*, 1992] T. Kepler, L.F. Abbott, et E. Marder. Reduction of conductance-based neuron models. *Biological Cybernetics*, 66(5) :381–387, Mars 1992.
- [Koch et Ullman, 1985] C. Koch et S. Ullman. Shifts in selective visual attention : towards the underlying neural circuitry. *Human Neurobiology*, 4(4) :219–227, 1985.
- [Kolb *et al.*, 2008] H. Kolb, E. Fernandez, et R. Nelson. Webvision : The organisation of the retina and visual system. Février 2008. Disponible en ligne : <http://webvision.med.utah.edu/>.

- [König *et al.*, 1996] P. König, A.K. Engel, et W. Singer. Integrator or coincidence detector ? the role of the cortical neuron revisited. *Trends in Neurosciences*, 19(4) :130–137, 1996.
- [Krüger *et al.*, 2000] V. Krüger, S. Bruns, et G. Sommer. Efficient head pose estimation with gabor wavelet networks. In *Proc. of the British Machine Vision Conference (BMVC)*. British Library, 2000.
- [Krüger et Sommer, 2002a] V. Krüger et G. Sommer. Gabor wavelet networks for efficient head pose estimation. *Image and Vision Computing*, 20(9-10) :665–672, 2002.
- [Krüger et Sommer, 2002b] V. Krüger et G. Sommer. Wavelet networks for face processing. *Journal of the Optical Society of America. A : Optics, image science, and vision*, 19(6) :1112–1119, Juin 2002.
- [Kuffler, 1953] S.W. Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1) :37–68, Janvier 1953.
- [Kustov et Robinson, 1996] A.A. Kustov et D.L. Robinson. Shared neural control of attentional shifts and eye movements. *Nature*, 384(6604) :74–77, Novembre 1996.
- [Last *et al.*, 2001] M. Last, A. Kandel, O. Maimon, et E. Eberbach. Anytime algorithm for feature selection. In *Rough Sets and Current Trends in Computing*, Lecture Notes in Computer Science, pages 532–539. Springer, 2001.
- [Latham *et al.*, 2000] P.E. Latham, B.J. Richmond, P.G. Nelson, et S. Nirenberg. Intrinsic dynamics in neuronal networks. i. theory. *Journal of Neurophysiology*, 83(2) :808–827, Février 2000.
- [Le Meur *et al.*, 2006] O. Le Meur, P. Le Callet, D. Barba, et D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 28(5) :802–817, Mai 2006.
- [LeCun et Bengio, 1998] Y. LeCun et Y. Bengio. Convolutional networks for images, speech, and time series. In [Arbib \[1998\]](#).
- [Levick, 1967] W.R. Levick. Receptive fields and trigger features of ganglion cells in the visual streak of the rabbits retina. *The Journal of physiology*, 188(3) :285–307, Février 1967.
- [Li, 2002] Z. Li. A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1) :9–16, Janvier 2002.
- [Loudni et Boizumault, 2003] S. Loudni et P. Boizumault. Solving constraint optimization problems in anytime contexts. In *Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 251–257, Aout 2003.
- [Lowe, 1999] D.G. Lowe. Object recognition from local scale-invariant features. In *Int. Conf. on Computer Vision (ICCV)*, Washington, DC, USA, 1999. IEEE Computer Society.
- [Lowe, 2004] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :91–110, Novembre 2004.
- [Ma et Zhang, 2003] Y.-F. Ma et H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *Proc. of the eleventh ACM international conference on Multimedia (MULTIMEDIA)*, pages 374–381, New York, USA, 2003. ACM.
- [Maass, 1997] W. Maass. Networks of spiking neurons : the third generation of neural network models. *Neural Networks*, 10 :1659–1671, 1997.

- [Maeder *et al.*, 1996] A.J. Maeder, J. Diederich, et E. Niebur. Limiting human perception for image sequences. In *Proc. of the SPIE, Human Vision and Electronic Imaging*, éditeurs B.E. Rogowitz et J.P. Allebach, volume 2657, pages 330–337. SPIE, 1996.
- [Maillard *et al.*, 2005] M. Maillard, O. Gapenne, P. Gaussier, et L. Hafemeister. Perception as a dynamical sensori-motor attraction basin. In *Advances in Artificial Life*, volume 3630 de *Lecture Notes in Computer Science*, pages 37–46. Springer, 2005.
- [Maillard, 2007] M. Maillard. *Formalisation de la perception comme dynamique sensori-motrice : application dans un cadre de reconnaissance d’objets par un robot autonome*. Thèse de Doctorat, Université de Cergy-Pontoise, Cergy, France, Décembre 2007.
- [Markram *et al.*, 1995] H. Markram, P.J. Helm, et B. Sakmann. Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons. *The Journal of physiology*, 485 (Pt 1) :1–20, Mai 1995.
- [Markram *et al.*, 1997] H. Markram, J. Lübke, M. Frotscher, et B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297) :213–215, Janvier 1997.
- [Markram *et al.*, 1998] H. Markram, Y. Wang, et M. Tsodyks. Differential signaling via the same axon of neocortical pyramidal neurons. *Proceedings of the National Academy of Sciences*, 95(9) :5323–5328, Avril 1998.
- [Markram et Tsodyks, 1996] H. Markram et M. Tsodyks. Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382(6594) :807–810, Aout 1996.
- [Marr, 1982] D. Marr. *Vision : a computational investigation into the human representation and processing of visual information*. W.H. Freeman and Compagny, New-York, USA, 1982.
- [Marre *et al.*, 2005] O. Marre, P. Baudot, M. Levy, et Y. Frégnac. High timing precision and reliability, low redundancy, and low entropy code in V1 neurons during visual processing of natural scenes. In *Annual Meeting of the Society For Neuroscience*, 2005.
- [Masland et Martin, 2007] R.H. Masland et P.R. Martin. The unsolved mystery of vision. *Current Biology*, 17(15) :R577–R582, Aout 2007.
- [Masland, 2001a] R.H. Masland. The fundamental plan of the retina. *Nature Neuroscience*, 4(9) :877–886, 2001.
- [Masland, 2001b] R.H. Masland. Neuronal diversity in the retina. *Current Opinion in Neurobiology*, 11(4) :431–436, 2001.
- [Masland, 2005] R.H. Masland. Sensory systems : fine-tuning the visual scene. *Current Biology*, 15(19), Octobre 2005.
- [Mazer et Gallant, 2003] J.A. Mazer et J.L. Gallant. Goal-related activity in v4 during free viewing visual search. evidence for a ventral stream visual salience map. *Neuron*, 40(6) :1241–1250, Décembre 2003.
- [McCulloch et Pitts, 1943] W.S. McCulloch et W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5 :115–133, 1943.
- [Meister et Berry, 1999] M. Meister et M.J. Berry. The neural code of the retina. *Neuron*, 22(3) :435–450, Mars 1999.

- [Meister, 1996] M. Meister. Multineuronal codes in retinal signaling. *Proceedings of the National Academy of Sciences*, 93(2) :609–614, Janvier 1996.
- [Meunier et Paugam-Moisy, 2005] D. Meunier et H. Paugam-Moisy. Evolutionary supervision of a dynamical neural network allows learning with on-going weights. In *Int. Joint Conf. on Neural Networks (IJCNN)*, Montréal, Canada, 2005.
- [Meunier et Paugam-Moisy, 2008] D. Meunier et H. Paugam-Moisy. Neural networks for computational neuroscience. In *European Symposium on Artificial Neural Networks (ESANN)*, éditeur M. Verleysen, pages 367–378, 2008.
- [Michalke *et al.*, 2007] T. Michalke, A. Gepperth, M. Schneider, J. Fritsch, et C. Goerick. Towards a human-like vision system for resource-constrained intelligent cars. In *Int. Conf. on Computer Vision Systems (ICVS)*, 2007.
- [Michalke *et al.*, 2008] T. Michalke, J. Fritsch, et C. Goerick. Enhancing robustness of a saliency-based attention system for driver assistance. In *Int. Conf. on Computer Vision Systems (ICVS)*, volume 5008 de *Lecture Notes in Computer Science*, pages 43–55. Springer, 2008.
- [Mikolajczyk et Schmid, 2005] K. Mikolajczyk et C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(10) :1615–1630, Octobre 2005.
- [Morrison *et al.*, 2005] A. Morrison, C. Mehring, T. Geisel, A. Aertsen, et M. Diesmann. Advancing the boundaries of high-connectivity network simulation with distributed computing. *Neural Computation*, 17(8) :1776–1801, August 2005.
- [Mouraud *et al.*, 2006] A. Mouraud, H. Paugam-Moisy, et D. Puzenat. A distributed and multithreaded neural event driven simulation framework. In *Int. Conf. on Parallel and Distributed Computing and Networks (PDCN)*, pages 212–217. ACTA Press, 2006.
- [Mozer et Sitton, 1998] M.C. Mozer et M. Sitton. Computational modeling of spatial attention. In Pashler [1998], pages 341–394.
- [Mozer, 1991] M.C. Mozer. *The Perception of Multiple Objects : A Connectionist Approach (Neural Network Modelling and Connectionism)*. MIT Press, Juin 1991.
- [Navalpakkam et Itti, 2005] V. Navalpakkam et L. Itti. Modeling the influence of task on attention. *Vision Research*, 45(2) :205–231, Janvier 2005.
- [Neisser, 1967] U. Neisser. *Cognitive Psychology*. Prentice Hall, 1ère édition, 1967.
- [Niebur, 2007] E. Niebur. Saliency map. *Scholarpedia*, 2(8) :2675, 2007.
- [Noë, 2002] A. Noë. Is the visual world a grand illusion? *Journal of Consciousness Studies*, pages 1–12, 2002.
- [Oliva et Torralba, 2001] A. Oliva et A. Torralba. Modeling the shape of the scene : A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3) :145–175, Mai 2001.
- [Olshausen *et al.*, 1993] B.A. Olshausen, C.H. Anderson, et D.C. Van Essen. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11) :4700–4719, Novembre 1993.
- [Olshausen et Field, 2004] B.A. Olshausen et D.J. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4) :481–487, 2004.

- [Olshausen et Field, 2005] B.A. Olshausen et D.J. Field. How close are we to understanding V1 ? *Neural Computation*, 17(8) :1665–1699, Aout 2005.
- [O’Regan *et al.*, 2000] K.J. O’Regan, H. Deubel, J.J. Clark, et R.A. Rensink. Picture changes during blinks : Looking without seeing and seeing without looking. *Visual Cognition*, pages 191–211, Janvier 2000.
- [O’Regan et Noë, 2000] K.J. O’Regan et A. Noë. Experience is not something we feel but something we do : a principled way of explaining sensory phenomenology, with change blindness and other empirical consequences. 2000. Disponible en ligne : <http://nivea.psych.univ-paris5.fr/ASSChtml/ASSC.html>.
- [O’Regan et Noë, 2001] K.J. O’Regan et A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24 :939–1031, 2001.
- [O’Regan, 1992] K.J. O’Regan. Solving the “real” mysteries of visual perception : the world as an outside memory. *Canadian Journal of Psychology*, 46(3) :461–488, Septembre 1992.
- [Osterberg, 1935] G. Osterberg. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica*, suppl. 6 :1–103, 1935.
- [Pashler, 1998] éditeur H. Pashler. *Attention*. Psychology Press, 1998.
- [Paugam-Moisy, 2006] H. Paugam-Moisy. Spiking neuron networks : a survey. Rapport Technique RR-11, IDIAP, Martigny, Switzerland, 2006.
- [Perrinet *et al.*, 2001] L. Perrinet, A. Delorme, M. Samuelides, et S.J. Thorpe. Networks of integrate-and-fire neuron using rank order coding a : How to implement spike time dependent hebbian plasticity. *Neurocomputing*, 38-40 :817–822, Juin 2001.
- [Perrinet, 2003] L. Perrinet. *Comment déchiffrer le code impulsionnel de la Vision ? Étude du flux parallèle, asynchrone et épars dans le traitement visuel ultra-rapide*. Thèse de Doctorat, Université Paul Sabatier, Toulouse, France, 2003.
- [Perrinet, 2004] L. Perrinet. Finding independent components using spikes : a natural result of hebbian learning in a sparse spike coding scheme. *Natural Computing*, 3(2) :159–175, 2004.
- [Pfrieger, 2002] F.W. Pfrieger. Role of glia un synapse development. *Current Opinion in Neurobiology*, 12 :486–490, 2002.
- [Phaf *et al.*, 1990] R.H. Phaf, A.H. Van der Heijden, et P.T. Hudson. Slam : a connectionist model for attention in visual selection tasks. *Cognitive Psychology*, 22(3) :273–341, Juillet 1990.
- [Pineau *et al.*, 2003] J. Pineau, G. Gordon, et S. Thrun. Point-based value iteration : An anytime algorithm for pomdps. In *Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1025–1032, Aout 2003.
- [Posner *et al.*, 1980] M.I. Posner, C.R. Snyder, et B.J. Davidson. Attention and the detection of signals. *Journal of Experimental Psychology*, 109(2) :160–174, Juin 1980.
- [Posner, 1980] M.I. Posner. Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1) :3–25, 1980.
- [Purves *et al.*, 1997] D. Purves, G.J. Augustine, D. Fitzpatrick, L.C. Katz, A.-S. LaMantia, et J.O. McNamara. *Neuroscience*. Sinauer Associates, USA, 1997.
- [Pylyshyn et Storm, 1988] Z. Pylyshyn et R.W. Storm. Tracking multiple independent targets : evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3) :179–197, 1988.

- [Pylyshyn, 1994] Z. Pylyshyn. Primitive mechanisms of spatial attention. *Cognition*, 50 :363–384, 1994.
- [Pylyshyn, 1998] Z. Pylyshyn. The role of visual indexes in spatial vision and imagery. In *Visual Attention*, éditeur R. Wright. Oxford University Press, New-York, USA, 1998.
- [Pylyshyn, 2001] Z. Pylyshyn. Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1-2) :127–158, 2001.
- [Quinlan, 2003] P.T. Quinlan. Visual feature integration theory : past, present, and future. *Psychological Bulletin*, 129(5) :643–673, Septembre 2003.
- [Rall, 1967] W. Rall. Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *Journal of Neurophysiology*, 30(5) :1138–1168, Septembre 1967.
- [Rao et al., 2002] R.P.N. Rao, G.J. Zelinsky, M.M. Hayhoe, et D.H. Ballard. Eye movements in iconic visual search. *Vision Research*, 42(11) :1447–1463, Mai 2002.
- [Rao et Ballard, 1995] R.P.N. Rao et D.H. Ballard. An active vision architecture based on iconic representations. *Artificial Intelligence*, 78(1-2) :461–505, Octobre 1995.
- [Rao et Ballard, 2005] R. Rao et D.H. Ballard. Probabilistic models of attention based on iconic representations and predictive coding. In Itti et al. [2005].
- [Rensink et al., 1997] R.A. Rensink, K.J. O’Regan, et J.J. Clark. To see or not to see : The need for attention to perceive changes in scenes. *Psychological Science*, 8(5) :368–373, 1997.
- [Rensink et al., 2000] R.A. Rensink, K.J. O’Regan, et J.J. Clark. On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, pages 127–145, Janvier 2000.
- [Rensink, 2000] R.A. Rensink. Seeing, sensing, and scrutinizing. *Vision Research*, 40(10-12) :1469–1487, 2000.
- [Revel et Gaussier, 2004] A. Revel et P. Gaussier. *Designing neural control architectures for an autonomous robot using vision to solve complex learning tasks*, chapitre 11, pages 299–350. Physica-Verlag Heidelberg, Juin 2004.
- [Riesenhuber et Poggio, 1999] M. Riesenhuber et T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11) :1019–1025, Novembre 1999.
- [Robinson et Petersen, 1992] D.L. Robinson et S.E. Petersen. The pulvinar and visual salience. *Trends in Neurosciences*, 15(4) :127–132, Avril 1992.
- [Rochel et Martinez, 2003] O. Rochel et D. Martinez. An event-driven framework for the simulation of networks of spiking neurons. In *European Symposium on Artificial Neural Networks (ESANN)*, éditeur M. Verleysen, pages 295–300, 2003.
- [Rougier et Vitay, 2006] N.P. Rougier et J. Vitay. Emergence of attention within a neural population. *Neural Networks*, 19(5) :573–581, 2006.
- [Rudolph et Destexhe, 2006] M. Rudolph et A. Destexhe. Analytical integrate-and-fire neuron models with conductance-based dynamics for event-driven simulation strategies. *Neural Computation*, 18(9) :2146–2210, Septembre 2006.
- [Rudolph et Destexhe, 2007] M. Rudolph et A. Destexhe. How much can we trust neural simulation strategies? *Neurocomputing*, 70(10-12) :1966–1969, June 2007.

- [Rumelhart et McClelland, 1986] éditeurs D.E. Rumelhart et J.L. McClelland. *Parallel distributed processing : explorations in the microstructure of cognition, vol. 1 : foundations*. MIT Press, Cambridge, USA, 1986.
- [Schaefer *et al.*, 2002] M. Schaefer, T. Schoenauer, C. Wolff, G. Hartmann, H. Klar, et U. Ruckert. Simulation of spiking neural networks - architectures and implementations. *Neurocomputing*, pages 647–679, Octobre 2002.
- [Schmid *et al.*, 2000] C. Schmid, R. Mohr, et C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2) :151–172, 2000.
- [Schmid et Mohr, 1997] C. Schmid et R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(5) :530–535, 1997.
- [Scholl, 2001] B.J. Scholl. Objects and attention : the state of the art. *Cognition*, 80(1-2) :1–46, Juin 2001.
- [Schwartz, 1990] éditeur E.L. Schwartz. *Computational Neuroscience*. MIT Press, Cambridge, USA, 1990.
- [Sejnowski *et al.*, 1988] T.J. Sejnowski, C. Koch, et P.S. Churchland. Computational neuroscience. *Science*, 241(4871) :1299–1306, 1988.
- [Sela et Levine, 1997] G. Sela et M.D. Levine. Real-time attention for robotic vision. *Real-Time Imaging*, 3 :173–194, 1997.
- [Selverston, 1988] A.I. Selverston. A consideration of invertebrate central pattern generators as computational data bases. *Neural Networks*, 1 :109–117, 1988.
- [Serre *et al.*, 2005] T. Serre, L. Wolf, et T. Poggio. Object recognition with features inspired by visual cortex. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 994–1000, Washington, DC, USA, 2005. IEEE Computer Society.
- [Shepherd, 1998] éditeur G.M. Shepherd. *The Synaptic Organization of the Brain*. Oxford University Press, 5e édition, Novembre 1998.
- [Sherman et Guillery, 2003] S.M. Sherman et R.W. Guillery. The visual relays in the thalamus. In Chalupa et Werner [2003], chapitre 35, pages 565–591.
- [Shipp, 2004] S. Shipp. The brain circuitry of attention. *Trends in Cognitive Sciences*, 8(5) :223–230, Mai 2004.
- [Siagian et Itti, 2007] C. Siagian et L. Itti. Biologically-inspired robotics vision monte-carlo localization in the outdoor environment. In *Int. Conf. on Intelligent Robots and Systems*, pages 1723–1730. IEEE Computer Society, 2007.
- [Simons, 2000] D.J. Simons. Current approaches to change blindness. *Visual Cognition*, 7(1) :1–15, 2000.
- [Sirosh *et al.*, 1996] J. Sirosh, R. Miikkulainen, et Y. Choe. Lateral interactions in the cortex : Structure and function. 1996. Disponible en ligne : <http://www.cs.utexas.edu/~nn/web-pubs/htmlbook96/>.
- [Song *et al.*, 2000] S. Song, K.D. Miller, et L.F. Abbott. Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3(9) :919–926, Septembre 2000.
- [Song et Abbott, 2001] S. Song et L.F. Abbott. Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, 32(2) :339–350, Octobre 2001.

- [Stemmler *et al.*, 1995] M. Stemmler, M. Usher, et E. Niebur. Lateral interactions in primary visual cortex : a model bridging physiology and psychophysics. *Science*, 269(5232) :1877–1880, Septembre 1995.
- [Sun et Fisher, 2003] Y. Sun et R. Fisher. Object-based visual attention for computer vision. *Artificial Intelligence*, 146(1) :77–123, Mai 2003.
- [Sun, 2003] Y. Sun. *Hierarchical Object-Based Visual Attention for Machine Vision*. Thèse de Doctorat, University of Edinburgh, 2003.
- [Taylor et Rogers, 2002] J.G. Taylor et M. Rogers. A control model of the movement of attention. *Neural Networks*, 15(3) :309–326, Avril 2002.
- [Taylor, 1999] J.G. Taylor. Neural 'bubble' dynamics in two dimensions : foundations. *Biological Cybernetics*, 80(6) :393–409, Juin 1999.
- [Theunissen, 2003] F.E. Theunissen. From synchrony to sparseness. *Trends in Neurosciences*, 26(2) :61–64, 2003.
- [Thiele *et al.*, 2002] A. Thiele, P. Henning, M. Kubischik, et K. P. Hoffmann. Neural mechanisms of saccadic suppression. *Science*, 295(5564) :2460–2462, Mars 2002.
- [Thorpe *et al.*, 1996] S.J. Thorpe, D. Fize, et C. Marlot. Speed of processing in the human visual system. *Nature*, 381 :520–522, 1996.
- [Thorpe *et al.*, 2004] S.J. Thorpe, R. Guyonneau, N. Guilbaud, J-M. Allegraud, et R. VanRullen. Spikenet : Real-time visual processing with one spike per neuron. *Neurocomputing*, 58-60 :857–864, 2004.
- [Thorpe et Imbert, 1989] S.J. Thorpe et M. Imbert. Biological constraints on connectionist modelling. In *Connectionism in perspective*, éditeurs R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, et L. Steels, pages 63–93. Elsevier, 1989.
- [Tootell *et al.*, 1982] R.B. Tootell, M.S. Silverman, E. Switkes, et R.L. De Valois. Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218(4575) :902–904, Novembre 1982.
- [Torralba et Oliva, 2003] A. Torralba et A. Oliva. Statistics of natural image categories. *Network : Computation in Neural Systems*, 14(3) :391–412, Aout 2003.
- [Torralba, 2003a] A. Torralba. Contextual priming for object detection. *International Journal of Computer Vision*, 53(2) :169–191, Juillet 2003.
- [Torralba, 2003b] A. Torralba. Modeling global scene factors in attention. *Journal of the Optical Society of America. A : Optics, image science, and vision*, 20(7) :1407–1418, Juillet 2003.
- [Torralba, 2005] A. Torralba. Contextual influences on saliency. In Itti et al. [2005].
- [Treisman et Gelade, 1980] A. Treisman et G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1) :97–136, 1980.
- [Treisman et Gormican, 1988] A. Treisman et S. Gormican. Feature analysis in early vision : Evidence from search asymmetries. *Psychological Review*, 95(1) :15–48, 1988.
- [Treisman et Souther, 1985] A. Treisman et J. Souther. Search asymmetry : a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology - General*, 114 :285–310, 1985.
- [Treisman, 1985] A. Treisman. Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*, 31 :156–177, 1985.

- [Treisman, 1988] A. Treisman. Features and objects : The fourteenth bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 40(2) :201–237, 1988.
- [Treisman, 1991] A. Treisman. Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology : Human Perception and Performance*, 17(3) :652–676, 1991.
- [Treisman, 1998] A. Treisman. Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society London : B Biological Science*, 353(1373) :1295–1306, Aout 1998.
- [Troy et Shou, 2002] J.B. Troy et T. Shou. The receptive fields of cat retinal ganglion cells in physiological and pathological states : where we are after half a century of research. *Progress in retinal and eye research*, 21(3) :263–302, Mai 2002.
- [Troyer et Miller, 1997] T.W. Troyer et K.D. Miller. Physiological gain leads to high isi variability in a simple model of a cortical regular spiking cell. *Neural Computation*, 9(5) :971–983, Juillet 1997.
- [Tsodyks *et al.*, 1998] M. Tsodyks, K. Pawelzik, et H. Markram. Neural networks with dynamic synapses. *Neural Computation*, 10(4) :821–835, Mai 1998.
- [Tsotsos *et al.*, 1995] J.K. Tsotsos, S. Culhane, W. Yan Kei Wai, Y. Lai, N. Davis, et F. Nuflo. Modeling visual attention via selective tuning. *Artificial Intelligence*, 78 :507–545, 1995.
- [Tsotsos *et al.*, 2005] J. Tsotsos, Y. Liu, J. Martineztrujillo, M. Pomplun, E. Simine, et K. Zhou. Attending to visual motion. *Computer Vision and Image Understanding*, 100(1-2) :3–40, Octobre 2005.
- [Tsotsos, 1989] J.K. Tsotsos. The complexity of perceptual search tasks. In *Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1571–1577. AAAI, Detroit, USA, Aout 1989.
- [Tsotsos, 1990] J.K. Tsotsos. Analysing vision at the complexity level. *Behavioral and Brain Sciences*, 13 :423–469, 1990.
- [Tsotsos, 1991] J.K. Tsotsos. Is complexity theory appropriate for analysing biological systems? *Behavioral and Brain Sciences*, 14(4) :770–773, 1991.
- [Tsotsos, 1992] J.K. Tsotsos. On the relative complexity of active v.s. passive visual search. *International Journal of Computer Vision*, 7(2) :127–141, 1992.
- [VanRullen *et al.*, 1998] R. VanRullen, J. Gautrais, A. Delorme, et S.J. Thorpe. Face processing using one spike per neurone. *Biosystems*, 48 :229–239, 1998.
- [VanRullen et Thorpe, 2001] R. VanRullen et S.J. Thorpe. Rate coding versus temporal order coding : what the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13(6) :1255–1283, June 2001.
- [VanRullen, 2000] R. VanRullen. *Une première vague de potentiels d'action, une première vague idée de la scène visuelle - role de l'asynchronie dans le traitement rapide de l'information visuelle*. Thèse de Doctorat, Université Paul Sabatier, Toulouse, France, 2000.
- [Verberne *et al.*, 2000] A. Verberne, F. van Harmelen, et A.T. Teije. Anytime diagnostic reasoning using approximate boolean constraint propagation. In *Int. Conf. on Principles of Knowledge Representation and Reasoning (KR)*, pages 323–332, 2000.

- [Vinje et Gallant, 2000] W.E. Vinje et J.L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456) :1273–1276, 2000.
- [Vitay *et al.*, 2005] J. Vitay, N.P. Rougier, et F. Alexandre. A distributed model of spatial visual attention. In *Biomimetic Neural Learning for Intelligent Robots*, Lecture Notes in Artificial Intelligence, pages 54–72. Springer, 2005.
- [Volterra et Steinhäuser, 2004] A. Volterra et C. Steinhäuser. Glial modulation of synaptic transmission in the hippocampus. *Glia*, 47(3) :249–257, 2004.
- [von der Malsburg, 1981] C. von der Malsburg. The correlation theory of brain function. Rapport Technique 81-2, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, 1981.
- [von der Malsburg, 1995] C. von der Malsburg. Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5(4) :520–526, Aout 1995.
- [von der Malsburg, 1999] C. von der Malsburg. The what and why of binding : The modeler’s perspective. *Neuron*, 24(1) :95–104, Septembre 1999.
- [von Neumann, 1958] J. von Neumann. *The Computer and the Brain*. Mrs. Hepsa Ely Silliman Memorial Lectures. Yale Univesity Press, New Haven, UK, 2e édition, 1958.
- [Wässle, 2004] H. Wässle. Parallel processing in the mammalian retina. *Nature Reviews Neuroscience*, 5(10) :747–757, Octobre 2004.
- [Watts, 1994] L. Watts. Event-driven simulation of networks of spiking neurons. In *Advances in Neural Information Processing Systems*, volume 6, pages 927–934, 1994.
- [Westin *et al.*, 1996] C.F. Westin, C.J. Westelius, H. Knutsson, et G. Granlund. Attention control for robot vision. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 726–733. IEEE Computer Society, 1996.
- [Wilson et Cowan, 1972] H.R. Wilson et J.D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1) :1–24, Janvier 1972.
- [Wilson et Cowan, 1973] H.R. Wilson et J.D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics*, 13(2) :55–80, 1973.
- [Wohrer *et al.*, 2006] A. Wohrer, P. Kornprobst, et T. Viéville. Contrast gain control through a feedback in the retina. In *NeuroComp 1ère conférence française de Neurosciences Computationnelles*, pages 154–157, Nancy, France, 2006.
- [Wohrer et Kornprobst, 2008] A. Wohrer et P. Kornprobst. Virtual retina : a biological retina model and simulator, with contrast gain control. *Journal of Computational Neuroscience*, In press, 2008.
- [Wohrer, 2008] A. Wohrer. *Modèle et simulateur à grande échelle d’une rétine biologique, avec controle de gain*. Thèse de Doctorat, Université de Nice-Sophia Antipolis, 2008.
- [Wolfe *et al.*, 1989] J. Wolfe, K.R. Cave, et S.L. Franzel. Guided search : An alternative to the feature integration model for visual search. *Journal of Experimental Psychology : Human Perception and Performance*, 15(3) :419–433, 1989.
- [Wolfe et Horowitz, 2004] J. Wolfe et T.S. Horowitz. What attributes guide the deployment of visual attention and how do they do it ? *Nature Reviews Neuroscience*, 5(6) :495–501, Juin 2004.

- [Wolfe, 1994] J. Wolfe. Guided search 2.0 : A revised model of visual search. *Psychonomic Bulletin and Review*, 1(2) :202–238, 1994.
- [Wolfe, 1998] J. Wolfe. Visual search. In Pashler [1998], pages 13–73.
- [Wolfe, 2000] J. Wolfe. Visual attention. In *Seeing*, éditeur K.K. De Valois, pages 335–386. Academic Press, San Diego, USA, 2e édition, 2000.
- [Wolff *et al.*, 1999] C. Wolff, G. Hartmann, et U. Ruckert. ParSPIKE-a parallel DSP-accelerator for dynamic simulation of large spiking neural networks. In *Int. Conf. on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems (MicroNeuro)*, pages 324–331. IEEE Computer Society, 1999.
- [Young, 1987] R.A. Young. The gaussian derivative model for spatial vision : I. retinal mechanisms. *Spatial Vision*, 2(4) :273–293, 1987.
- [Zhang *et al.*, 1995] Z. Zhang, R. Deriche, O. Faugeras, et Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2) :87–119, Octobre 1995.
- [Zilberstein et Russell, 1996] S. Zilberstein et S. Russell. Optimal composition of real-time systems. *Artificial Intelligence*, pages 181–213, Avril 1996.

Annexe A

Implémentation du réseau de neurones

Le système est écrit en Ada pour pouvoir s’interfacer avec les *drivers* développés pour la gestion des caméras et des robots du laboratoire. D’autre part, le système développé utilise plusieurs *threads* : nous avons utilisé le fait que Ada permet de définir, au niveau syntaxique, des systèmes concurrents.

Le système a été intégralement développé pendant cette thèse. Le code est séparé en une dizaine de *packages* et fait environ 10 000 lignes de code.

A.1 Conception du système

Nous avons utilisé une approche objet et modulaire pour concevoir ce système. La figure A-1 présente les différents composants utilisés. Les modèles neuronaux *SN.LIF*, *SN.SLIF* et *SN.Dirac* sont utilisés pour constituer des cartes neuronales (*D.Spiking Map*). Ces cartes sont des instances du type abstrait *Device*, tout comme *D.Inflow* qui permet de s’interfacer avec différentes sources d’images d’entrées. Les cartes neuronales sont associées à une liste de masques de connexion (*MC.List*) et à des listes d’évènements (*E.List*).

Ce système est réparti sur plusieurs *threads* : chaque carte neuronale 2D (*D.Spiking Map* sur la figure A-1) est contrôlée par une tâche Ada, qui gère sa mise à jour et les communication avec les autres cartes (principalement pour les émissions des potentiels d’action).

Le système utilise une approche hybride, entre les approches synchrone et asynchrone (voir la partie 3.3.3), dont la complexité algorithmique est principalement fonction du nombre de potentiels d’action émis à chaque pas de temps de simulation. En effet, pour les approches synchrones la complexité dépend (1) du coût de la mise à jour des neurones

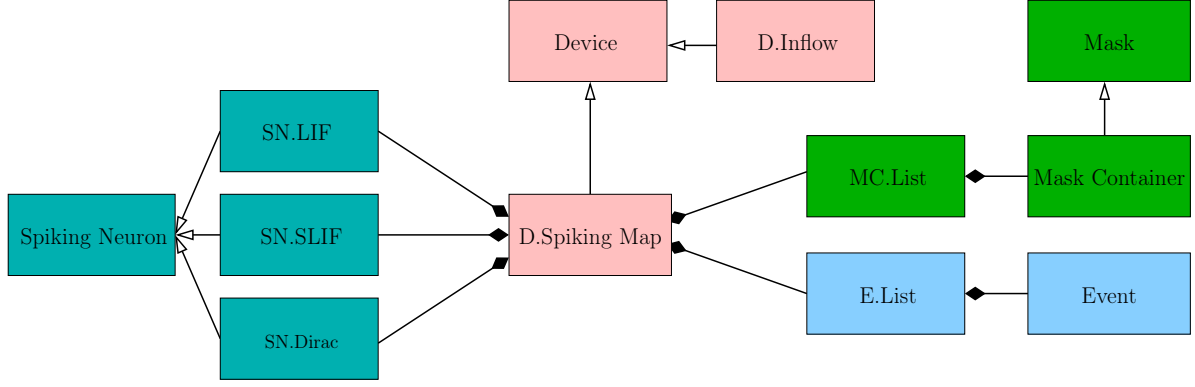


FIGURE A-1 Schéma présentant les liens entre les différents *packages* du système.

et (2) du coût de la propagation des potentiels d'action [Brette *et al.*, 2007], soit :

$$c_u \times \frac{N}{\Delta t} + c_p \times F \times N \times p \quad (\text{A-1})$$

où c_u est le coût de la mise à jour d'un neurone, N le nombre de neurones du système, Δt le pas de temps choisi pour la simulation, c_p le coût de propagation d'un potentiel d'action, F la fréquence moyenne de décharge des neurones du système et p le nombre moyen de connexions sortantes.

L'approche hybride consiste à ne mettre à jour que les neurones actifs, c'est-à-dire ceux qui reçoivent des potentiels d'action. Dans ce cas, la complexité devient :

$$c_u \times \frac{A}{\Delta t} + c_p \times F \times N \times p \quad (\text{A-2})$$

où A est le nombre moyen de neurones actifs. Pour montrer le coût computationnel de l'utilisation de notre système, nous proposons dans la suite de cette annexe de détailler les parties importantes du code, c'est-à-dire celles qui concernent la mise à jour des cartes neuronales et la propagation des potentiels d'actions.

A.2 La propagation des potentiels d'action

Les cartes neuronales sont des instances de **Map**, un type dérivé du type abstrait **Device**. Elles sont composées d'un tableau 2D de neurones (**Unit_Array**), d'une liste d'événements à traiter (**Events**), d'une liste de potentiels d'action à propager (**Output_List**) vers d'autres cartes au travers de masques de connexion (**Mask_List**).

```
type Map (Size_X, Size_Y : Positive) is new Device with record
  Unit   : Unit_Array (1..Size_X, 1..Size_Y); — les unités
  Events : Event_List_Ptr; — liste des événements
```

```

Output_List : Output_Spike_List_Ptr;      — liste des spikes emis
Mask_List   : Mask_List_Ptr;              — masques de connexion
T           : Map_Task (Map'Access);      — la tâche associée
...
end record;
```

Une carte neuronale est contrôlée par une tâche associée (T) de type Map_Task :

```

task type Map_Task (O : access Device'Class) is
  entry Start;
  entry Kill;
  entry Write (Events  : in out Event_List_Ptr;
               Link_Id  : in String);
  ...
end Map_Task;
```

Une tâche Map_Task est définie par différents *rendez-vous* (Start, Kill, Write, ...) qui permettent de synchroniser les tâches entre elles. Lors d'une simulation, le rôle d'une Map_Task consiste à échanger des trains de potentiels d'action avec d'autres cartes puis à mettre à jour la carte contrôlée lorsque tous les potentiels d'action du pas de temps simulé courant ont été échangés.

```

task body Map_Task is
begin
  Main_Loop : loop
    exit when End_Of_Task;
    select
      accept Start do
        ...
      end Start;
    or
      accept Kill do
        ...
      end Kill;
    or
      accept Write (Events  : in out Event_List_Ptr;
                    Link_Id  : in String) do
        — récupère le train de potentiels d'action entrant
        — et les ajoute à la liste d'événements à traiter
        — par la carte
        ...
      end Write;
    or
      delay Duration (Random (G)) * Very_Small_T;
      — envoyer les potentiels d'actions sortant
```



```
— à toutes les cartes en aval
...
Map (To).T.Write (Events_To_Send, Link_Id.all);
...
— si tous les potentiels d'action ont été échangés
— la carte peut être mise à jour
if Exchange_Ok then
    Update;
end if;
end select;
end loop Main_Loop;
```

La structure `select accept X; or delay T; Y;` permet à la tâche d'accepter des appels pour le *rendez-vous* `X` pendant un temps `T` puis d'exécuter `Y`. Pour notre système, ceci nous permet de synchroniser, pour une carte donnée, les informations entrantes et sortantes quelque soit la configuration des connexions entre les cartes. Ceci évite par exemple des situations bloquantes lorsque deux cartes sont connectées de façon réciproque ou s'il existe un chemin de connexion bouclé entre les différentes cartes de neurones.

Les trains de potentiels d'action échangés entre les cartes (`Event_List_Ptr`) sont des listes d'événements, pour lesquelles chaque événement contient le temps d'émission du potentiel d'action, les coordonnées du neurone qui le reçoit et le poids synaptique. Ainsi, lorsqu'un neurone émet un potentiel d'action, on ajoute K événements à la liste d'événement sortant à envoyer avec $K = \sum_{p \in P} X_p Y_p$, où X_p et Y_p dénotent la taille du masque de connexions p . Par exemple, pour une carte qui propage ses potentiels d'actions vers trois autres cartes au travers de trois masques de connexions identiques de taille 15×15 , chaque potentiel d'action émis par cette carte ajoute $3 \times 15 \times 15 = 675$ événements à propager. Ces masques de connexions (*connection masks*, [Wolff *et al.*, 1999; Schaefer *et al.*, 2002]) implémentent une technique proche du *weight sharing*, qui est très utilisée pour l'apprentissage de champs récepteurs [Fukushima, 1980; Rumelhart et McClelland, 1986; LeCun et Bengio, 1998; Riesenhuber et Poggio, 1999; Serre *et al.*, 2005; Masland et Martin, 2007]. Ceci permet de stocker efficacement des connexions régulières entre deux cartes, puisque les connexions utilisent un adressage relatif qui est traduit en coordonnées absolues lorsqu'un neurone décharge. On parle alors d'implémentation *sender-oriented*, méthode qui est très bien adaptée au contexte de réseau de neurones impulsionnels en vision : cela permet de gagner du temps et de réduire l'empreinte en mémoire par rapport à une implémentation *receiver-oriented*, pour laquelle les neurones cibles doivent interroger les synapses à chaque mise à jour pour vérifier l'évolution de leur état.

Une fois que tous les potentiels d'action entrants et sortants ont été échangés, la carte est mise à jour. Le type de mise à jour dépend du modèle synaptique utilisé et la partie suivante en propose une explication détaillée.

A.3 La mise à jour des cartes neuronales

Notre système utilise des intégrateurs à fuite pour lesquels nous avons implémenté trois types de modèles synaptiques : les interactions synaptiques instantanées (*SN.Dirac* sur la figure A-1), les fonctions α (*SN.LIF*) et les modèles en différences d'exponentielles (*SN.SLIF*). Les aspects mathématiques de ces modèles sont détaillés dans l'annexe B, cette partie propose des explications concernant leur implémentation.

Les modèles utilisés héritent du type `Spiking_Neuron`, qui définit un type de neurone abstrait et ses primitives.

```

package Spiking_Neuron is
  type Neuron is abstract tagged limited private;
  type Neuron_Ptr is access all Neuron'Class;

  — Retourne le potentiel du neurone
  function Potential (N : access Neuron'Class) return Real;
  — Retourne l'état du neurone
  function Firing (N : access Neuron'Class) return Boolean;
  — autres fonctions basiques
  ...

private
  type Neuron is tagged limited record
    Potential      : Real;           — potentiel de membrane
    G_Leak         : Real;           — conductance de fuite
    E_Leak         : Real;           — potentiel de repos
    Firing         : Boolean := False; — état du neurone
  end record;
end Spiking_Neuron;

```

Les intégrateurs à fuite utilisant des interactions synaptiques instantanées ont la mise à jour la plus simple et la moins coûteuse en terme computationnel. À chaque nouveau potentiel d'action reçu, la valeur du potentiel membranaire courant est calculée en ajoutant directement le poids synaptique à la valeur du potentiel de membrane mise à jour (ici avec une décroissance exponentielle).

```

function Update (N : in Neuron_Ptr;
                  T : in Real;

```

```

        W : in Real) return Boolean is
begin
  — si le neurone a déjà déchargé à ce pas de temps, on ne fait rien
  if N.Firing and then N.Last_Spike = T then
    return False;
  else
    — sinon on réinitialise firing et on calcule le nouveau potentiel
    N.Firing := False;
    N.Potential := W + N.Potential * Exp (N.G_Leak * (T - N.Last_Input));
    N.Last_Input := T;

    — le neurone decharge si le potentiel est supérieur au seuil
    if N.Potential <= N.Threshold then
      return N.Firing;
    else
      N.Firing := True;
      N.Last_Spike := T;
      N.Potential := N.E_Reset;
    end if;
    return N.Firing;
  end if;
end Update;
```

Au niveau de la carte neuronale, il suffit donc de parcourir la liste d'événements à traiter et de vérifier, pour chaque événement, s'il déclenche l'émission d'un potentiel d'action.

Les modèles synaptiques qui utilisent une fonction α ou une différence d'exponentielles sont plus coûteux en terme computationnel. Pour ces deux modèles, il est nécessaire de découpler la prise en compte des potentiels d'action reçus et la mise à jour du potentiel membranaire. Ainsi, au niveau d'une carte neuronale, il faut parcourir la liste d'événements pour prendre en compte les modifications suscitées par les potentiels d'action puis mettre à jour le potentiel membranaire des neurones qui ont reçu au moins un potentiel d'action. Par exemple pour une fonction α :

```

procedure Add_PSP (N : access Neuron;
        W : in Real;
        T : in Real;
        Spike_Time : in Real) is
begin
  N.Potential := W*(T-Spike_Time)/N.Tau * Exp(1-(T-Spike_Time))
    + N.Potential;
end Add_PSP;
```

```

procedure Update (N : access Neuron;
                  T : in Real) is
begin
  N.Potential := N.G_Leak * (N.Potential - N.E_Leak) + N.Potential;
  N.Firing := False;
end Update;

```

Le coût computationnel de notre implémentation des fonctions α est très élevé, car pour chaque potentiel d'action il faut calculer sa contribution sur le potentiel membranaire et cette contribution s'étale sur plusieurs pas de temps simulé. Il est donc nécessaire de conserver dans la liste d'événements les potentiels d'action des pas de temps passés jusqu'à ce que leur contribution soit négligeable. Le nombre d'événements à traiter à chaque pas de temps peut donc devenir très important et nécessiter de nombreux calculs.

Nous avons implémenté un modèle de synapse en différence d'exponentielles qui permet d'intégrer directement la contribution d'un potentiel d'action dans l'étape de mise à jour du potentiel membranaire (voir l'annexe [B.3](#)).

```

procedure Add_PSP (N : access Neuron;
                  W : in Real;
                  T : in Real) is
  T_Rel : constant Real := T - N.T_Last_Spike;
begin
  N.Kbarre0 := N.Kbarre0 + W * N.Kj * Exp (N.K0 * T_Rel);
  N.Kbarre1 := N.Kbarre1 + W * N.Kj * Exp (N.K1 * T_Rel);
end Add_PSP;

function Update (N : access Neuron;
                  T : in Real) return Boolean is
  T_Rel : constant Real := T - N.T_Last_Spike;
  V_PSP : constant Real := N.Kbarre0 * Exp (-N.K0 * T_Rel)
    - N.Kbarre1 * Exp (-N.K1 * T_Rel);
begin
  N.Potential := N.G_Leak * (N.Potential - N.E_Leak) + N.Potential + V_PSP;
  N.Firing := False;

  if N.Potential >= N.Threshold then
    N.Potential := N.E_Reset;
    N.Kbarre0 := 0.0;
    N.Kbarre1 := 0.0;
    N.Firing := True;
    N.T_Last_Spike := T;
  end if;
  return N.Firing;

```

end Update ;

Annexe B

Modèles synaptiques implémentés dans le système

L'architecture neuronale est composée d'intégrateurs à fuite, avec ou sans conductances synaptiques. Le système implémenté dans le cadre de cette thèse permet d'utiliser différents types de neurones. Cette partie détaille les différents types de modèles d'interaction synaptique qu'il est possible d'utiliser avec notre système.

B.1 Conductance synaptique instantanée

Pour les modèles d'interactions synaptiques instantanées, on définit les potentiels postsynaptiques (PPS) en provenance du neurone j comme :

$$S_j(t) = \sum_f \delta(t - t_j^f) \quad (\text{A-1})$$

où $\delta(x)$ est une distribution de Dirac, avec $\delta(x) = 0$ pour $x \neq 0$ et $\int_{-\infty}^{\infty} \delta(x) dx = 1$. L'influence des PPS sur le potentiel de membrane est donnée par :

$$\text{PSP}_i(t) = \sum_j w_{i,j} S_j(t) \quad (\text{A-2})$$

Ce type d'interaction synaptique est peu coûteux en calculs, puisqu'il s'agit de réaliser une simple somme pour calculer la contribution des PPS reçus pour un neurone donné.

B.2 Fonction α

Pour modéliser de façon plus réaliste le potentiel postsynaptique $PPS_i(t)$ généré par l'arrivée d'un potentiel d'action à $t_j^{(f)}$ sur la synapse j d'un neurone i , on utilise classiquement une fonction α (voir la partie 2.3.1). Dans notre système, nous avons utilisé une fonction α de la forme :

$$\alpha(x) = w_{ij} \frac{x}{\tau_{syn}} e^{1 - \frac{x}{\tau_{syn}}} H(x) \quad (A-3)$$

où w_{ij} est le poids synaptique, τ_{syn} est une constante de relaxation synaptique et $H(x)$ est la fonction échelon. L'évolution du potentiel membranaire du neurone i est alors décrite par :

$$C \frac{dV_i}{dt} = g_l(V_i - E_l) + \sum_{j,f} PPS_i(t - t_j^f) \quad (A-4)$$

$$= g_l(V_i - E_l) + \sum_j w_{ij} \sum_f \frac{t - t_j^f}{\tau_{syn}} e^{1 - \frac{t - t_j^f}{\tau_{syn}}} H(t_j^f, t) \quad (A-5)$$

B.3 Modèle en différence d'exponentielles

Cependant, les fonctions alpha sont coûteuses à simuler car il faut conserver en mémoire l'ensemble des temps t_j^f pour mettre à jour le neurone i . Nous avons donc choisi d'implémenter également un modèle de $PPS_i(t)$ utilisant des différences d'exponentielles [Gerstner et Kistler, 2002, , chapitre 4.1.3.1]. Ce modèle, dont le dérouls temporel est similaire à la fonction α (figure A-1), permet de paramétrer les temps de montée et de descente du potentiel postsynaptique de façon indépendante.

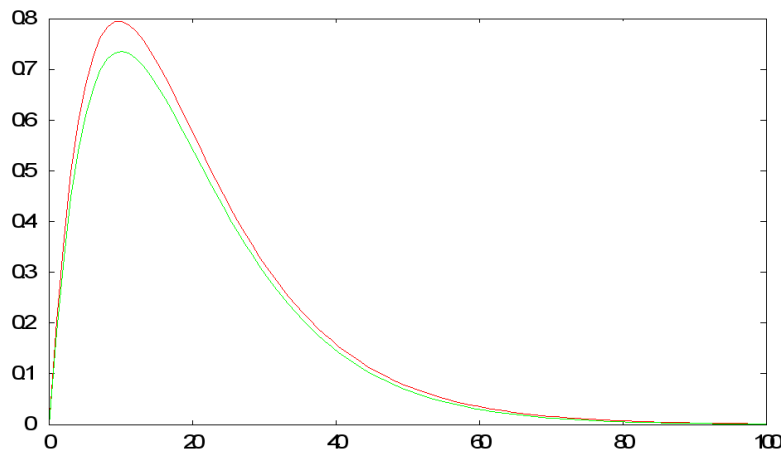


FIGURE A-1 Comparaison de la fonction alpha (en vert ou gris clair) et de la différence d'exponentielles (en rouge ou gris sombre).

Dans le cas où les contributions des PPS sont additives, on peut rapidement calculer les contributions d'une somme de j PPS sur le potentiel de membrane. En effet, l'équation décrivant l'évolution du potentiel de membrane après qu'un train de PPS a atteint le neurone est donnée par

$$PPS(t) = \sum_{j=1}^n \frac{k_0 w_{ij}}{k_1 - k_0} \left[e^{-k_0(t-t_j^f)} - e^{-k_1(t-t_j^f)} \right] \quad (\text{A-6})$$

$$= e^{-k_0 t} \sum_{j=1}^n \frac{k_0 w_{ij}}{k_1 - k_0} e^{k_0 t_j^f} - e^{-k_1 t} \sum_{j=1}^n \frac{k_0 w_{ij}}{k_1 - k_0} e^{k_1 t_j^f} \quad (\text{A-7})$$

On pose

$$K_0(n) = \sum_{j=1}^n \frac{k_0 w_{ij}}{k_1 - k_0} e^{k_0 t_j^f} \quad (\text{A-8})$$

$$K_1(n) = \sum_{j=1}^n \frac{k_0 w_{ij}}{k_1 - k_0} e^{k_1 t_j^f} \quad (\text{A-9})$$

$$(\text{A-10})$$

pour réécrire l'équation (A-7) sous la forme

$$PPS(t) = K_0 e^{-k_0 t} - K_1 e^{-k_1 t} \quad (\text{A-11})$$

Il est possible de calculer, pour chaque pas de temps, la contribution du train de PPS en une seule étape de calcul, en utilisant le résultat de l'équation (A-11). Si un nouveau PPS arrive sur le neurone à l'instant t_n^f , les valeurs de $K_0(n)$ et $K_1(n)$ sont remises à jour avec :

$$K_0(n) = K_0(n-1) + \frac{k_0 w_{ij}}{k_1 - k_0} e^{k_0 t_n^f} \quad (\text{A-12})$$

$$K_1(n) = K_1(n-1) + \frac{k_0 w_{ij}}{k_1 - k_0} e^{k_1 t_n^f} \quad (\text{A-13})$$

$$(\text{A-14})$$

Annexe C

Générateur de potentiels d'action pour les cartes d'entrée

Le système proposé a été conçu pour évaluer l'apport des neurones impulsionnels pour des tâches de vision artificielle et une ré-implémentation pourrait largement améliorer la vitesse d'exécution de ce programme. En particulier, pour implémenter une version plus optimisée des cartes d'entrées, il est possible de modéliser les neurones des cartes d'entrée comme des générateurs de potentiels d'action. L'architecture proposée permet de traiter des séquences d'images et l'implémentation d'un système asynchrone nécessite de pouvoir prévoir le temps d'émission des potentiels d'action (voir l'équation (4.3)) et de les mettre à jour lorsque l'entrée évolue. Soit à partir de l'équation (4.2) :

$$V(t) = \tau KL(1 - e^{-\lambda t}) + V_0 e^{\lambda t} \quad (\text{A-1})$$

On exprime la valeur de \hat{t} en fonction de ϑ :

$$\hat{t} = t_0 + \tau \ln \left(\frac{\vartheta + \tau KL}{V_0 + \tau KL} \right), \text{ si } \frac{\vartheta + \tau KL}{V_0 + \tau KL} > 0 \quad (\text{A-2})$$

Nous avons vérifié la cohérence de ces résultats en comparant les valeurs numériques obtenues à partir de l'équation (A-2) et la simulation d'un intégrateur à fuite. Les résultats de cette comparaison sont présentés sur la figure A-1 et la figure A-2. Sur la première figure le neurone est soumis à un courant constant, sur la deuxième le neurone est soumis à un courant variable prenant 3 valeurs différentes au cours de la simulation.

Nous proposons un canevas pour une implémentation plus adéquate de ces neurones utilisés dans les cartes d'entrée, dont les points importants sont définis ci-après :

1. Un neurone peut émettre un potentiel d'action seulement dans le cas où $\frac{\vartheta + \tau KL}{V_0 + \tau KL} > 0$.
0. Sinon, le potentiel tend à se stabiliser sur une valeur sous-liminaire.

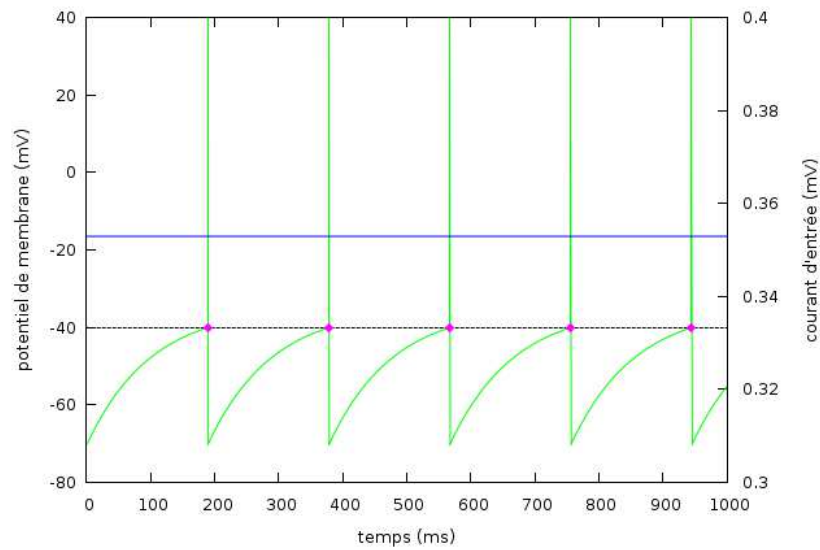


FIGURE A-1 Évolution du potentiel de membrane d'un intégrateur à fuite (en vert ou gris clair) soumis à un courant d'entrée constant (en bleu ou gris foncé). Les temps d'émission des potentiels d'action, obtenus à partir de la solution analytique sont indiqués par les points violets.

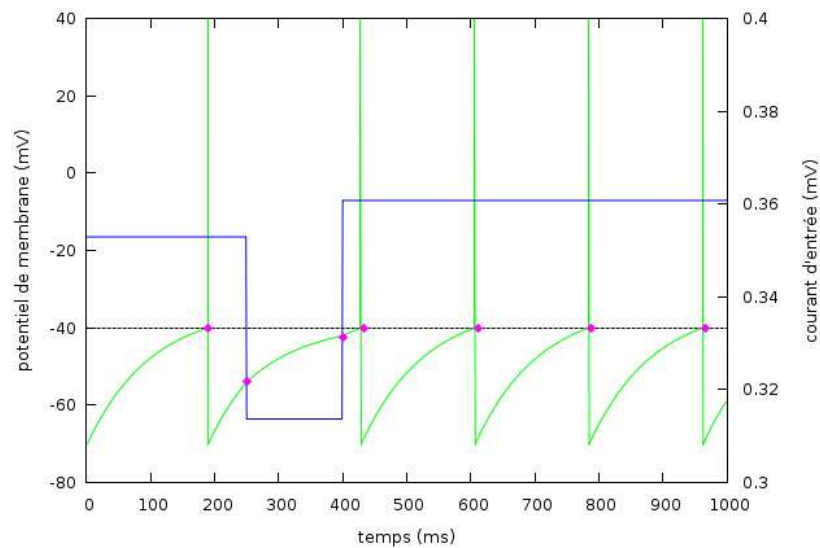


FIGURE A-2 Évolution du potentiel de membrane d'un intégrateur à fuite (en vert ou gris clair) soumis à un courant d'entrée constant par morceaux (en bleu ou gris foncé). Les temps d'émission des potentiels d'action, obtenus à partir de la solution analytique sont indiqués par les points violets.

2. Un neurone sous l'influence d'un courant constant et suffisamment élevé émet des potentiels d'action régulièrement, la latence du premier potentiel d'action est donnée par l'équation (A-2), puis le neurone décharge ensuite avec un intervalle inter-spike (ISI) donné par l'équation (4.3).
3. Il n'est pas nécessaire de faire de calculs si le courant d'entrée ne change pas. C'est le cas si la valeur de I ne change pas quand l'image est mise à jour.
4. Si le courant d'entrée change et est suffisamment élevé, il suffit alors de calculer la valeur du potentiel au moment du changement, valeur qui est donnée par l'équation (A-1), et calculer le temps d'émission du prochain potentiel d'action avec l'équation (A-2).

Le système implémenté simule pour l'instant des intégrateurs à fuite plutôt que ces générateurs de potentiels d'action, car ceci nous permet une plus grande souplesse pour intégrer rapidement des modèles de neurones plus complexes, comme des modèles stochastiques utilisés dans la partie 4.2.3 par exemple. Ces générateurs, beaucoup plus légers à simuler, peuvent être utilisés pour une ré-implémentation à visée applicative et peuvent s'intégrer facilement dans un système asynchrone (voir la partie 3.3.3).

Annexe D

Résultats complémentaires pour l'extraction de saillances

Les deux figures suivantes présentent les images obtenues avec le codage fréquentiel pour toutes les cartes de l'architecture. On peut constater que les informations apparaissant sur chaque carte sont très différentes. Ces informations sont combinées sur la carte de saillances \mathcal{S} .

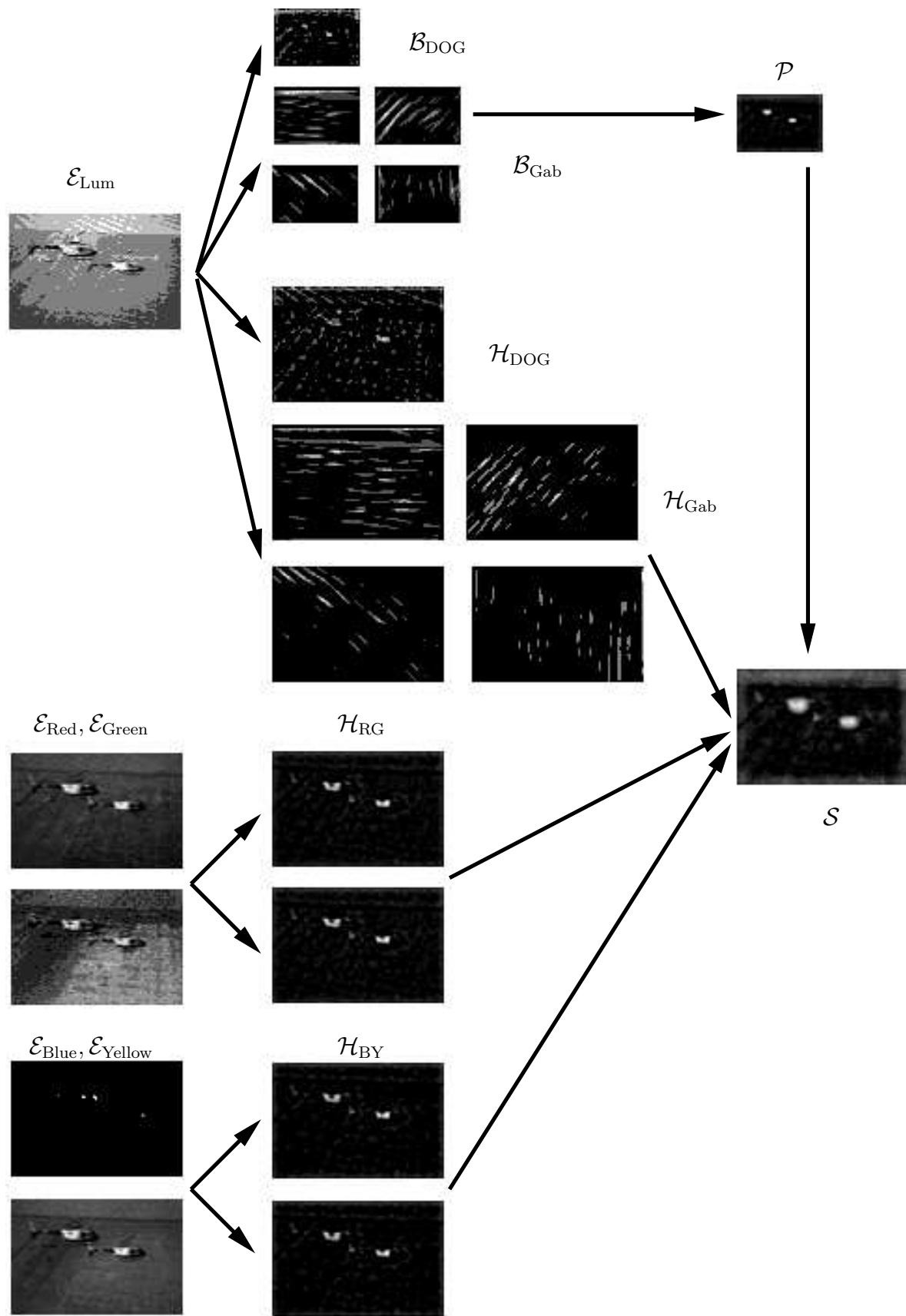


FIGURE A-1 Présentation de toutes les images résultantes de chaque carte neuronale pour le codage fréquentiel, sur l'image représentant des hélicoptères de la figure 5.12.

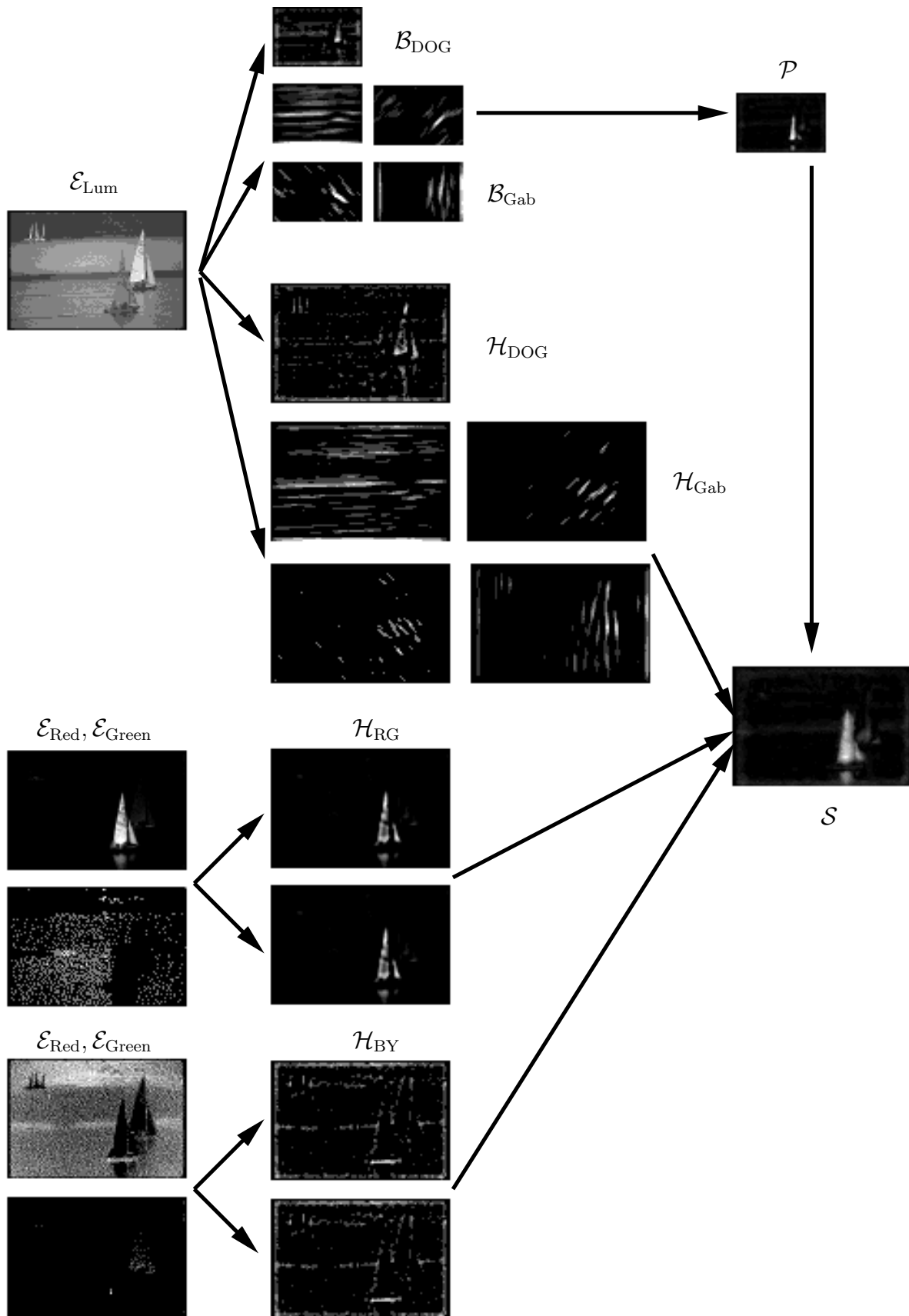


FIGURE A-2 Images résultantes de chaque carte neuronale pour le codage fréquentiel sur l'image montrant des bateaux de la figure 5.12.