



HAL
open science

Mécanismes numériques et distribués de l'anticipation motrice

Jérémy Fix

► **To cite this version:**

Jérémy Fix. Mécanismes numériques et distribués de l'anticipation motrice. Modélisation et simulation. Université Henri Poincaré - Nancy I, 2008. Français. NNT: . tel-00336194v3

HAL Id: tel-00336194

<https://theses.hal.science/tel-00336194v3>

Submitted on 18 Nov 2008

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.

Mécanismes numériques et distribués de l'anticipation motrice

THÈSE

présentée et soutenue publiquement le 30 Octobre 2008

pour l'obtention du

Doctorat de l'université Henri Poincaré – Nancy 1
(spécialité informatique)

par

Jérémy Fix

Composition du jury

- Président :* Sylvain Contassot-Vivier, Professeur, Université Henri-Poincaré Nancy-I
- Rapporteurs :* Arnaud Revel, Maître de conférence habilité, Université de Cergy-Pontoise, ENSEA
Guillaume Masson, Directeur de recherche, CNRS
- Examineurs :* Jean Lorenceau, Directeur de recherche, CNRS
Nicolas Rougier, Chargé de recherche, INRIA
- Directeur :* Frédéric Alexandre, Directeur de recherche, INRIA

Mis en page avec la classe thloria.

Résumé

Cette thèse s'inscrit dans le domaine des neurosciences computationnelles dont le but est de modéliser des fonctions cognitives complexes par le biais de simulations informatiques et numériques en s'inspirant du fonctionnement cérébral. Les modèles et simulations proposées reposent sur le paradigme des champs neuronaux, que nous exploitons pour étudier dans quelle mesure des capacités cognitives complexes peuvent être le résultat émergent de l'interaction de cellules élémentaires simples.

Nous nous intéressons dans cette thèse à la modélisation de l'attention visuelle, avec ou sans mouvement oculaire. Pour guider le développement de nos modèles, nous proposons dans une première partie une revue de données psychologiques et physiologiques sur l'attention visuelle avant de proposer un modèle computationnel de l'attention visuelle sans saccade oculaire. Puis, nous nous intéressons dans une seconde partie à la manière dont on peut intégrer les saccades oculaires dans nos modèles en s'inspirant des données anatomiques et physiologiques sur le contrôle des saccades oculaires chez le primate. Les performances des différents mécanismes proposés sont évalués en simulation en les appliquant à des tâches de recherche visuelle.

Nos travaux de thèse permettent également d'étudier un paradigme de calcul original qui repose sur des calculs distribués, asynchrones, numériques et adaptatifs qui permettent d'envisager le déploiement des mécanismes proposés dans ce cadre sur des supports de calculs parallèles.

Mots-clés: Sciences cognitives, Attention visuelle, Perception visuelle, Mouvements oculaires, Réseaux neuronaux, Connexionnisme

Abstract

This thesis belongs to the computational neuroscience domain in which we aim at understanding complex cognitive functions with computer simulations built on the current knowledge of the brain. The proposed models and simulations are built on the paradigm of dynamic neural fields, that we use in order to study in which way complex cognitive capabilities can be the emergent result of the interaction of elementary units.

In this thesis, we are interested in the modelisation of visual attention, with and without eye movements. To guide the development of these models, we propose in the first part a review of the current psychological and physiological data on visual attention, before proposing a computational model of visual attention without saccadic eye movement. Then, we study in the second part the way we can integrate saccadic eye movements in our models based on the current anatomical and physiological data on the oculomotor control in the primate. The performances of the different proposed mechanisms are evaluated by simulating visual search tasks with saccadic eye movements.

This work also makes us able to study a computational paradigm that relies on distributed, asynchronous, numerical and adaptive computation which permits to consider further developments of the proposed mechanisms on parallel architectures.

Keywords: Cognitive science, visual attention, visual perception, eye movements, neural networks, connexionsm

Remerciements

Ces travaux de thèse et cette expérience inoubliable ne sont pas le résultat d'un voyage solitaire et je tiens à remercier ici l'ensemble des personnes qui y ont contribué. La rédaction n'est pas mon fort mais je vais essayer de remercier chacun d'entre vous comme il se doit. J'ai toujours eu tendance à privilégier le fond sur la forme, et même si c'est un handicap pour la rédaction, je vais profiter de cette page de remerciement pour m'abstraire de ces contraintes.

Mes premiers remerciements vont tout naturellement aux membres du jury pour avoir pu consacrer une partie de leur temps et de leur énergie dans l'évaluation de ces travaux. Je tiens à remercier Sylvain Contassot-Vivier d'avoir accepté de présider ce jury, les rapporteurs Guillaume Masson et Arnaud Revel pour avoir eu la patience de relire le manuscrit et Jean Lorenceau pour avoir joué l'avocat du diable en posant des questions "naïves".

J'ai eu la chance, durant ma thèse et également pendant les deux stages qui ont précédé, d'être encadré, guidé, et alimenté par Frédéric Alexandre et Nicolas Rougier que je tiens à remercier. Même si sa modestie le pousse à indiquer qu'il ne contribue qu'indirectement à l'encadrement des thèses, il n'en reste pas moins que ses qualités scientifiques et personnelles font de Frédéric une pièce essentielle du puzzle. Je tiens également à remercier Nicolas pour nos échanges scientifiques et amicaux de ces dernières années. Il a toujours su prendre suffisamment de hauteur pour m'aider à forger ces travaux, et qui nous manque parfois lorsqu'on a "la tête dans le guidon". Je n'oublierais pas non plus les efforts qu'ils ont fait pour transformer des gribouilles en un document compréhensible par tous, en faisant parfois intervenir des règles de typographie insoupçonnées. J'en profite également pour remercier Julien Vitay puisque les travaux de thèse exposés ici ont leurs racines dans ses propres travaux de thèse et c'est finalement avec lui que j'ai fait mes premiers pas dans le Cortex.

De manière plus générale, je souhaite remercier l'ensemble des membres de l'équipe Cortex, en commençant par Martine Kuhlman qui nous rend la vie administrative si simple, dont on entrevoit la complexité qu'aux dernières étapes de la thèse. Je tiens également à remercier tous les permanents : Hervé Frezza-Buet pour m'avoir intronisé dans le monde fantastique gloubi-boulgazgouiffre de la recherche et pour ses blagues et gadgets, Yann Boniface pour ses tenues bariolées et son humour débridé qui peut paraître un peu étrange quand on débarque dans l'équipe, Laurent Bougrain pour son oreille attentive et ses talents d'archiver, Bernard Giraud et son amour des barbabapa, Dominique Martinez pour son amour de la Lorraine, Thomas Voegtlin pour sa discipline, Axel Hutt, Thierry Viéville qui sait donner une dimension moins formelle à la recherche, et les extracorticaux Bruno Scherrer, Alain Dutech et Amine Boumaza. Je souhaite bon courage à ceux qui sont sur les rails : Hana Belmabrouk : je pourrais me nourrir au moins un repas si je vais en Tunisie : "Assalamau alaykom, bortokal min fazlik, chokran", Lucian Alecu, Mathieu Lefort, Max Ambard qui fait chanter l'olfaction, Mauricio Cerda-Villablanca : "¡debes seguir aprendiendo Francés porque vas a hacer la presentación en Francés!", Thomas Girod, Randa Kassab. Je remercie également les autres "précaires", Noelia Montejo, Eduardo Martin, Greg Rolland et Roger, qui a passé une partie de son existence le ventre à l'air avant l'intervention conjointe d'un médecin Mexicain et Chilien.

Mes derniers remerciements sont adressés à ma famille. Je commencerais pas remercier Nath parce qu'elle su faire preuve de beaucoup de patience, surtout pendant la phase terminale et qu'elle est d'un soutien inestimable. Je tiens également à remercier Michel, qui préfère manipuler

les lettres là où je préfère manipuler les chiffres et qui m'a donné un sacré coup de main pour la relecture, et Agnes qui m'a soutenu au travers d'une alimentation saine. Merci aussi à ma soeurette Nenette, au frangin Benou et à ma beldoche Marthoune et à toute ma famille, mais je crains qu'un manuscrit entier ne suffirait pas pour tous les citer. Mes derniers remerciements sont adressés à mes parents Nini et mon ptit père, c'est finalement grâce à vous, à votre éducation et à vos sacrifices que j'ai eu la chance d'en arriver jusque là, merci.

*A ceux qui me sont chers, sans qui cette thèse n'aurait jamais vu le jour,
A mes parents.*

Table des matières

Glossaire	1
-----------	---

Introduction

Chapitre 1

L'attention visuelle sélective

1.1	Attention visuelle exogène guidée par les propriétés physiques des stimuli	10
1.2	Attention visuelle endogène guidée par les buts	13
1.2.1	Attention guidée par des informations spatiales	14
1.2.2	Attention guidée par les attributs visuels	19
1.3	Bilan des résultats expérimentaux sur l'attention visuelle	23
1.4	Revue des principaux modèles attentionnels	24
1.4.1	Carte de saillance : attention bottom-up, L. Itti	25
1.4.2	Selective Tuning Model : compétition pour la représentation, J.K. Tsotsos	27
1.4.3	The reentry hypothesis, F. Hamker	29

Chapitre 2

Emergence de l'attention visuelle sur un substrat de calculs numériques et distribués

2.1	Une introduction aux neurones biologiques	37
2.2	Le connexionnisme comme cadre de modélisation : Calculs locaux numériques distribués et adaptatifs	41
2.2.1	Calcul local et émergence	42
2.2.2	Calcul numérique	45
2.2.3	Calculs et représentations distribués	47
2.2.4	Adaptatif	50
2.2.5	Bilan du calcul local, numérique, distribué et adaptatif	56
2.3	Les champs neuronaux	57
2.3.1	Quelques exemples illustratifs	57

2.4	Un mécanisme d'attention sélective dans le cadre des champs neuronaux	64
2.4.1	Un mécanisme de sélection	65
2.4.2	La nécessité d'une forme de mémoire spatiale	69
2.4.3	Intégrer des attributs non spatiaux : un mécanisme de recherche visuelle	75

Chapitre 3

Le contrôle cortical et sous cortical des saccades oculaires : représentation et manipulation de l'information spatiale

3.1	Les mouvements oculaires	88
3.1.1	Fixation	88
3.1.2	Vergence	90
3.1.3	Réflexe vestibulo-oculaire	91
3.1.4	Réflexe opto-cinétique et poursuite lente	92
3.1.5	Saccade oculaire	93
3.2	Contrôle sous-cortical des saccades oculaires	96
3.2.1	La formation réticulée	98
3.2.2	Le colliculus supérieur	100
3.3	Les structures corticales impliquées dans le contrôle des saccades oculaires	102
3.3.1	De la rétine au cortex visuel strié et extrastrié	103
3.3.2	Cortex pariétal	112
3.3.3	Cortex frontal	118
3.3.4	Ganglions de la base	127
3.4	Mécanismes de traitement de l'information spatiale : transformation et mise à jour	131
3.4.1	Transformations sensorimotrices : champs de gain	131
3.4.2	Mettre à jour une mémoire de travail spatiale	135

Chapitre 4

Anticiper pour assurer la cohérence d'une mémoire de travail spatiale

4.1	Mécanisme sigma-pi pour les transformations de représentations spatiales	141
4.1.1	Etude analytique du cas 2D continu	142
4.1.2	Illustration dans le cas 1D discret	144
4.1.3	Conclusion	145
4.2	Saccades vers des cibles mémorisées dans un référentiel extra-rétinien	145
4.2.1	Architecture du mécanisme proposé	146
4.2.2	Simulations et résultats	148
4.2.3	Conclusion	153

4.3	Saccades vers des cibles mémorisées dans un référentiel centré-œil : rôle de l'anticipation	154
4.3.1	Introduction	154
4.3.2	Mécanisme proposé	156
4.3.3	Simulations et résultats	157
4.3.4	Conclusion	163
4.4	Mécanisme de déploiement overt de l'attention visuelle spatiale	164
4.4.1	Introduction	164
4.4.2	Mécanisme proposé	164
4.4.3	Simulations et résultats	166
4.4.4	Conclusion	169

Conclusion et perspectives

Annexe A

Paramètres des modèles

A.1	Equations du mécanisme des paragraphes 2.4.3 et 4.4	183
A.1.1	V4	183
A.1.2	Percept	183
A.1.3	Focus	183
A.1.4	Wm	184
A.1.5	Anticipation	184
A.1.6	IT	184
A.1.7	PF	185
A.1.8	Move	185
A.1.9	Switch	185
A.2	Paramètres du mécanisme du paragraphe 2.4.3	185
A.3	Paramètres du mécanisme du paragraphe 4.2	187
A.4	Paramètres du mécanisme du paragraphe 4.3	188
A.5	Paramètres du mécanisme du paragraphe 4.4	189

Annexe B

Décomposition en valeurs singulières

B.1	Introduction	191
B.2	Application au calcul d'un produit de convolution	191
B.2.1	Position du problème	191
B.3	Application aux réseaux de neurones	193

B.3.1	Principe	193
B.3.2	Cas des projections afférentes	193
B.3.3	Cas des projections latérales, évaluation asynchrone par carte	194
B.3.4	Evaluation des performances	194

Annexe C

Distributed Asynchronous Numerical and Adaptative computations

C.1	Introduction	197
C.2	Présentation des différents packages	198
C.2.1	dana.core : la librairie de base, calcul local et asynchrone	198
C.2.2	dana.cnft : définition de l'équation d'évolution des cellules, calcul numérique200	
C.2.3	gpython : une librairie de visualisation	201
C.2.4	dana.learn : une librairie d'apprentissage, calcul adaptatif	201
C.2.5	dana.sigmapi : Unités Sigma-Pi	203
C.2.6	dana.svd : Optimisation des calculs	204
C.2.7	dana.image : Librairie de traitement d'image	205
C.2.8	gpython.world : Un environnement virtuel	208

Annexe D

Codage de la proprioception : monotone ou explicite

D.1	Introduction	211
D.2	Étude empirique	211
D.2.1	Définition du problème	211
D.2.2	Simulation et résultats	213
D.3	Résultat analytique	215

Annexe E

Publications

E.1	A Distributed Computational Model of Spatial Memory Anticipation During a Visual Search Task	218
E.1.1	Introduction	218
E.1.2	Visual search	219
E.1.3	A model of visual search with overt attention	223
E.1.4	Discussion	228
E.2	From physiological principles to computational models of the cortex	235
E.2.1	Motivations	235
E.2.2	The computational approach	238

E.2.3 Discussion	245
Bibliographie	247

Glossaire

AIP : Aire intrapariétale antérieure (*Anterior Intraparietal Area*)
CD : Noyau caudé (*Caudate Nucleus*)
dIPFC : Cortex préfrontal dorsolatéral (*Dorsolateral Prefrontal Cortex*)
EBN : Excitatory Burst Neurons
FIT : Théorie d'intégration des traits (*Feature Integration Theory*)
FEF : Champ oculomoteur frontal (*Frontal Eye Field*)
GPe : Globus pallidus externe
GPi : Globus pallidus interne
IBN : Inhibitory Burst Neurons
IPS : Sillon intrapariétal (*Intraparietal Sulcus*)
LIP : Aire intrapariétale latérale (*Lateral Intraparietal Area*)
LLBN : Long Lead Burst Neurons
LGN : Corps genouillé latéral (*Lateral Geniculate Nucleus*)
MIP : Aire intrapariétale médiale (*Medial Intraparietal Area*)
MLBN : Medium Lead Burst neurons
MST : Aire supérieure temporale médiale (*Medial Superior Temporal Area*)
MT : Aire temporale moyenne (*Middle Temporal Area*)
OKR : Réponse optocinétique (*Optokinetic Response*)
OKN : Nystagmus optocinétique (*Optokinetic Nystagmus*)
OPN : Omnipause Neurons
PEF : Champ oculomoteur pariétal (*Parietal Eye Field*)
PIP : Aire intrapariétale postérieure (*Posterior Intraparietal Area*)
SC : Colliculus Supérieur (*Superior Colliculus*)
SEF : Champ oculomoteur supplémentaire (*Supplementary Eye Field*)
SNr : Substantia Nigra pars Reticulata
SNe : Substantia Nigra pars Compacta
STN : Noyau Sous-Thalamique (*Subthalamic Nucleus*)
STR : Striatum
TE : Lobe temporal antérieur
TEO : Lobe temporal postérieur
VIP : Aire intrapariétale ventrale (*Ventral Intraparietal Area*)
VOR : Réflexe vestibulo-oculaire (*Vestibulo-ocular reflex*)

Introduction

No amount of experimentation can ever prove me right ; a single experiment can prove me wrong.

Albert Einstein

Dans l'étude du système nerveux, nous disposons d'observations à différents niveaux (figure 1) qu'on peut nommer niveau macroscopique, niveau mésoscopique et niveau microscopique qui proviennent de l'ensemble des disciplines scientifiques que sont les sciences cognitives. Le niveau macroscopique est le niveau technologiquement le plus accessible. Il consiste par exemple à mener des expériences de psychologie et à étudier le comportement du sujet en interaction avec son environnement (figure 1a). Par exemple, B.F. Skinner proposa le conditionnement opérant en étudiant le comportement d'un rat, placé dans une cage, dans laquelle est disposée une pédale dont l'appui déclenche la délivrance d'une boulette de nourriture. L'étude du conditionnement met en particulier l'accent sur la manière dont le sujet apprend à interagir avec son environnement. Au niveau mésoscopique, on s'intéresse aux aires corticales, découvertes au début du siècle par l'anatomiste Allemand K. Brodmann, et aux principaux flux d'information entre ces aires (figure 1b). A ce niveau, les expériences consistent par exemple à léser une aire corticale et à en mesurer les conséquences. Par exemple P. Broca a découvert une aire impliquée dans la production du langage en s'intéressant à un sujet, dont une partie cortex était lésée, qui ne pouvait prononcer que le mot "tan". Sur la figure 1b sont représentés quelques aires corticales ainsi que quelques principaux flux d'information. On distingue notamment le cortex visuel (dont V1, V2, V3 et V4 sont des subdivisions) dans le lobe occipital, qui traite l'information visuelle transmise par la rétine. Les aires situées le long de l'axe occipito-pariétal sont plutôt spécialisées dans le traitement de l'information spatiale tandis que les aires situées le long de l'axe occipito-temporal sont plutôt spécialisées dans le traitement des attributs visuels comme la couleur ou la forme. Les aires du cortex moteur transmettent les commandes motrices pour contrôler les membres tandis que le cortex somatosensoriel informe sur leur état. Le cortex préfrontal est impliqué dans des fonctions cognitives complexes comme la planification ou la mémoire à court terme. Certaines structures du cortex sont impliquées dans des boucles comme, par exemple, les boucles entre le cortex et les ganglions de la base. A un niveau plus fondamental, qu'on appellera microscopique, on découvre que ces aires corticales sont constituées de populations de neurones (figure 1c). Ces cellules émettent de brèves impulsions électriques transportées le long de fibres pour atteindre d'autres cellules. Le système nerveux peut ainsi être vu comme un ensemble de cellules élémentaires interconnectées pour former un réseau. C'est le niveau auquel on se situe dans cette thèse : l'interaction de populations de cellules élémentaires. Bien sûr, on pourrait continuer la description du système nerveux à des niveaux plus fondamentaux. On a parlé de populations de neurones,

mais on pourrait également s'intéresser au neurone, ou même encore aux propriétés chimiques de ces cellules. L'hypothèse qu'on fait ici est que l'interaction de populations de neurones est un niveau d'abstraction suffisant pour expliquer certaines capacités cognitives complexes.

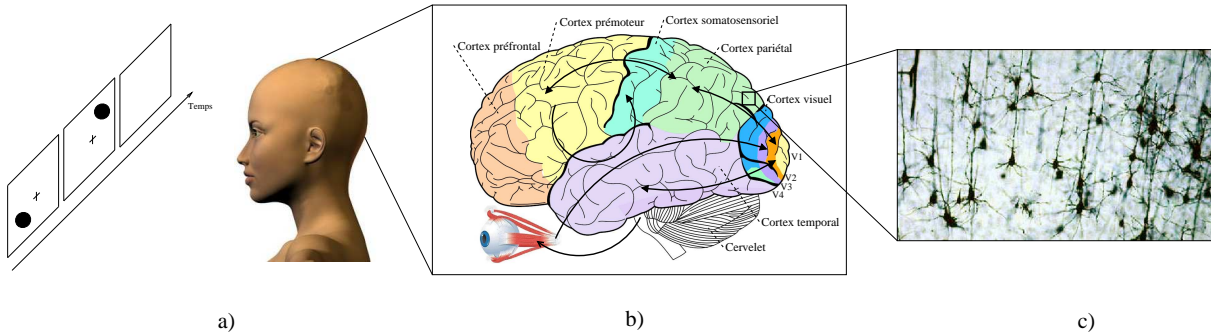


FIG. 1 – Le système nerveux peut être étudié à différents niveaux. De manière non invasive, les expériences de psychologie consistent à faire réaliser des tâches à des sujets tout en mesurant, par exemple, leur temps de réaction. Les études anatomiques nous renseignent sur un premier niveau d'organisation du cerveau. Sur l'illustration centrale sont par exemple représentées quelques grandes aires corticales et quelques flux d'information. A un niveau de granularité plus fin, on découvre que les aires corticales sont constituées de populations de neurones. Se référer au texte pour plus de détails.

La modélisation informatique peut être appréciée selon deux points de vue : soit comme une approche complémentaire à l'expérimentation dans l'étude d'un système physique, soit comme l'étude d'un paradigme de calcul dans le cadre duquel le modèle est défini. La modélisation d'un système consiste à retenir un certain nombre d'hypothèses à partir des connaissances actuelles du système étudié et à construire, par analogie, un second système plus facilement appréhendable. En informatique, par exemple, on cherchera à décrire le système par un ensemble d'équations mathématiques qu'on simule. L'interprétation des résultats de la simulation dépend de l'objectif visé. Pour les neurosciences, on cherche à fournir des prédictions sur le fonctionnement du système nerveux, testables expérimentalement. La principale hypothèse de laquelle nous partons est l'hypothèse connexionniste selon laquelle les fonctions cognitives émergent de l'interaction de cellules élémentaires qui peut être vue comme des calculs locaux massivement distribués. En informatique, on cherche à construire un système opérationnel en partant de l'observation que des fonctions comme la navigation, la planification, la reconnaissance d'objets, la résolution de problèmes pour lesquels l'espace des possibles est grand (par exemple le jeu d'échecs ou le jeu de go ; la navigation peut également être considéré comme un problème pour lequel l'espace des possibles est grand), le langage, etc... sont des fonctions dont on aimerait bien doter les systèmes informatiques mais qu'on réalise encore difficilement par des approches standards. On distingue ainsi les deux domaines de contribution de nos travaux : pour les neurosciences dans la compréhension du fonctionnement du système nerveux et pour l'informatique dans l'étude d'un paradigme de calcul original.

Dans cette thèse, on s'intéresse à la modélisation d'une capacité cognitive en particulier qu'est l'attention visuelle, en montrant que cette fonction peut émerger de calculs locaux massivement

distribués. Nous mettons en particulier l'accent sur la sélectivité de l'attention visuelle qui permet de traiter en priorité les stimuli visuels les plus pertinents pour le comportement. Différentes expériences, sur lesquelles on reviendra dans le chapitre 1, montrent que nous ne traitons pas l'intégralité de l'information visuelle transmise par les photorécepteurs de la rétine mais que le traitement est guidé en faveur d'une partie de cette information. L'attention visuelle peut être guidée spatialement ou par des attributs visuels (comme la couleur, l'orientation, le mouvement, etc...), par des informations exogènes, c'est à dire propres à l'environnement visuel, ou endogènes, c'est à dire définies par des buts internes. Par exemple, un flash lumineux soudain attire votre regard. C'est un exemple d'information exogène qui guide l'attention visuelle. Si vous cherchez vos clefs dans une pièce, vous êtes par exemple amenés à les chercher sur une table, ou bien à guider votre recherche par la couleur de ces clefs : vos connaissances a priori des propriétés de l'objet que vous recherchez permettent de rendre votre recherche plus efficace.

De la présentation des connaissances actuelles sur l'attention visuelle, au niveau psychologique et physiologique, nous dérivons des principes de fonctionnement qui guident le développement du modèle exposé dans le chapitre 2. En particulier, nous nous placerons dans l'hypothèse que l'attention visuelle spatiale est déployée séquentiellement. C'est une hypothèse de travail encore controversée puisque d'autres auteurs proposent que l'attention visuelle est une ressource limitée déployée en parallèle sur plusieurs stimuli, d'autres proposent encore qu'elle est déployée séquentiellement à une fréquence plus élevée que ce que propose l'hypothèse de déploiement séquentiel "standard". Cette hypothèse de déploiement séquentiel implique certaines contraintes qui nous permettent de proposer dans le chapitre 2 un modèle défini dans le cadre des champs neuronaux. Les champs neuronaux sont une théorie de laquelle on dérive un système d'équations différentielles qui régissent l'évolution du mécanisme défini dans ce cadre. Ce chapitre, à la fois d'introduction du paradigme de calcul et de présentation d'un modèle d'attention visuelle, nous permet d'illustrer comment une fonction complexe, comme l'attention visuelle, peut émerger de calculs locaux massivement distribués.

Dans la première partie du manuscrit, nous évoquons une forme de l'attention visuelle qu'on appelle attention visuelle *covert*, c'est à dire l'attention visuelle déployée sans mouvement oculaire. Il est très simple de l'illustrer. Pour cela il suffit de fixer son regard sur un objet tout en essayant de prendre connaissance des objets aux alentours. Garder le regard fixe sur un objet et explorer la scène visuelle sans bouger les yeux n'est néanmoins pas la manière la plus naturelle d'observer une scène visuelle, en particulier parce que la résolution spatiale de la rétine n'est pas uniforme. Elle est en effet maximale dans une région de la rétine, qu'on appelle la fovéa, et qui est de la taille d'un citron qu'on regarde lorsqu'on le tient dans la main, bras tendu. La résolution spatiale de la rétine décroît rapidement avec l'excentricité par rapport à la fovéa. Il est ainsi essentiel, si on souhaite extraire le maximum d'information d'un objet visuel, de pouvoir orienter la fovéa sur cet objet. Les saccades oculaires sont un des mouvements oculaires qui permettent d'orienter la fovéa vers un objet. Ce sont des mouvements potentiellement de grande amplitude, rapides, et qu'on exécute en moyenne à une fréquence de trois mouvements par seconde. L'une des conséquences de ces mouvements est que l'information qui se projette sur la rétine change brutalement toutes les 300 ms. Et pourtant, le monde visuel nous paraît stable. Les circuits neuronaux impliqués dans le contrôle oculomoteur sont de mieux en mieux compris. On présentera dans le chapitre 3 l'état des connaissances actuelles sur le contrôle oculomoteur en mettant l'accent sur la manière dont l'information spatiale est représentée dans ce circuit.

Quel est le lien entre l'attention visuelle traitée dans la première partie du manuscrit et les mouvements oculaires traités dans la seconde partie ? Selon la théorie prémotrice de l'attention, ce sont en grande partie les mêmes aires corticales qui sont impliquées dans le contrôle oculomoteur et le déploiement de l'attention visuelle spatiale. On distingue ainsi le déploiement *covert* de l'attention, que nous avons déjà mentionné, et le déploiement *overt* de l'attention qui correspond au déploiement de l'attention sur une région de l'espace visuelle suivi d'un mouvement oculaire en direction de cette région. Nous étudions alors dans le chapitre 4 comment il est possible d'étendre le modèle de déploiement de l'attention *covert*, présenté dans le chapitre 2, pour aborder l'une des conséquences des saccades oculaires qu'est le rapide et brutal changement d'information envoyée par la rétine. Cela nous amènera à nous intéresser à la manière dont l'information spatiale peut être représentée dans les modèles, et en particulier le référentiel dans lequel ces informations sont représentées. On explore en particulier la possibilité que des informations spatiales soient mémorisées dans un référentiel centré-tête, c'est à dire indépendant de la position des yeux. Par contre, puisque les aires oculomotrices, qui transmettent la commande motrice aux structures sous-corticales qui vont finalement innerver les muscles extraoculaires, travaillent dans un référentiel rétinotopique, il est nécessaire de transformer l'information mémorisée dans un référentiel extra-rétinien dans un référentiel rétinotopique. Ces contraintes nous amèneront ainsi à étudier un mécanisme de transformation d'information spatiale dans le cadre des champs neuronaux. Nous explorons également une seconde architecture dans laquelle les informations spatiales sont mémorisées dans un référentiel rétinotopique. Une mémoire centrée-œil implique qu'elle soit mise à jour avec chaque mouvement oculaire. On propose alors que cette mémoire peut être mise à jour par anticipation. En effet, les saccades oculaires sont des mouvements volontaires et, malgré la controverse qui existe à ce sujet, certains auteurs proposent qu'une copie du mouvement oculaire (copie efférente ou décharge corollaire) est transmise à certaines aires cérébrales lorsqu'un mouvement va être exécuté. A partir de ces idées, on propose un mécanisme qui met à jour une mémoire spatiale centrée-œil à partir d'une copie de la commande motrice qui va être exécutée. Cette seconde partie du manuscrit se termine par l'intégration du mécanisme d'anticipation au mécanisme d'attention visuelle *covert* de la première partie. Ceci nous conduit finalement à proposer un modèle dans lequel l'attention visuelle *covert* et *overt* émergent de calculs locaux et massivement distribués. Le système ainsi obtenu est utilisé pour réaliser une tâche de recherche visuelle avec des saccades oculaires.

Les principales contributions de cette thèse sont la présentation de propriétés fonctionnelles de sélection et de mémorisation qui émergent d'architectures simples définies dans le cadre des champs neuronaux et présentées dans les sections 2.4.1 et 2.4.2 ; la proposition d'un modèle de déploiement de l'attention *covert* en section 2.4.3 ; l'étude d'un modèle de saccades vers des cibles mémorisées dans un référentiel centré-tête, transformées dans un référentiel centré-œil en utilisant un codage de la position actuelle de l'œil, en section 4.2 ; l'étude d'un modèle de saccades vers des cibles mémorisées dans un référentiel centré-œil, mises à jour par anticipation, en section 4.3 ; l'étude d'un modèle de déploiement de l'attention *covert* et *overt* dans la section 4.4.

Chapitre 1

L'attention visuelle sélective

ATTENTION n.f. - Action de fixer son esprit sur qqch. ; concentration de l'activité mentale sur un objet déterminé

*Le petit robert de la langue française,
2006*

Introduction

L'attention est cette faculté qu'a le cerveau de sélectionner une partie de l'information disponible et d'en rejeter une autre. Elle est étudiée depuis longtemps ; pour ce qui concerne l'audition, [Cherry, 1953] réalise des expériences d'écoute dichotique qui consistent à diffuser un message différent dans chaque oreille d'un sujet à l'aide d'un casque. Il a constaté que lorsque le sujet doit reporter le message de l'une des deux oreilles, ce dernier a beaucoup de difficultés à reporter le message diffusé dans l'autre oreille. Ce phénomène se retrouve également dans le problème du *Cocktail Party* : lorsque nous discutons avec un interlocuteur dans une pièce remplie de personnes qui parlent également, nous pouvons ne prêter attention qu'aux paroles de notre interlocuteur tout en ignorant celles de nos voisins. Ce constat conduit [Broadbent, 1958] à proposer la "théorie du filtre sélectif de l'attention" : seule une partie restreinte des sensations d'un stimulus accède au traitement qui permet de l'identifier ; c'est une théorie de sélection précoce (*early selection*) par l'attention selon laquelle une sélection des sensations a lieu pour l'accès à l'identification ou à la perception. Un peu plus tard, [Gray et Wedderburn, 1960] proposent une expérience similaire à celle de [Cherry, 1953] dans laquelle les messages des deux oreilles sont mélangés : on diffuse le message "chat-4-souris" dans l'oreille gauche et "3-mange-5" dans l'oreille droite. On constate alors que les sujets sont capables d'associer les mots en fonction de leur sens, c'est à dire qu'ils peuvent reporter les phrases "3-4-5" et "chat-mange-souris". Cette étude va à l'encontre de la proposition de Broadbent et conduit [Deutsch et Deutsch, 1963] à proposer une théorie de sélection tardive (*late selection*) par l'attention suggérant que la sélection des informations a lieu à un niveau plus intégré que celui de l'extraction des caractéristiques physiques des stimuli.

A la suite des travaux concernant le rôle de l'attention pour le traitement de stimuli auditifs, les études se sont focalisées sur le rôle de l'attention dans la modalité visuelle. En particulier, [Treisman et Gelade, 1980] propose “la théorie d'intégration des traits” (*Feature Integration Theory*, FIT) qui suggère deux étapes entre les stimuli et leur identification ou l'accès à la mémoire à court terme. La première consiste en un filtrage parallèle des caractéristiques physiques des stimuli (couleur, orientation, mouvement) puis en un déploiement séquentiel de l'attention. Le même débat entre sélection précoce et tardive qui a eu lieu pour des stimuli auditifs s'observe également pour les stimuli visuels. La proposition originelle de [Treisman et Gelade, 1980] suggère que la recherche d'une cible définie par une propriété physique est réalisée en parallèle tandis que la recherche d'une conjonction de propriétés est réalisée en série. C'est à dire que lorsqu'une cible visuelle est suffisamment contrastée, elle est immédiatement localisée. Alors que si la cible est moins contrastée, plusieurs stimuli sont analysés, les uns après les autres, avant de trouver la cible. Cette distinction brutale entre recherche de trait et recherche par conjonction a été critiquée par différents auteurs [Duncan et Humphreys, 1989, Wolfe, 1998], conduisant A. Treisman à réviser en partie la FIT [Treisman et Sato, 1990]. Comme le soulignent [Desimone et Duncan, 1995], les modèles de sélection précoce impliquent que l'attention se déploie rapidement d'un objet à l'autre car ils sont traités les uns après les autres (les travaux récents de [Woodman et Luck, 2003, Bricolo et al., 2002] défendent ce point de vue) tandis que les modèles de sélection tardive n'imposent pas cette contrainte puisque plusieurs objets sont reconnus en même temps (les travaux de [Ward et al., 1996] défendent cette seconde thèse). Il ne semble pas possible de trancher en faveur de l'une ou l'autre des propositions ; en fait, il est probable qu'il y ait de la compétition à différents niveaux [Allport, 1993], suggérant que la sélection précoce et la sélection tardive sont deux interprétations extrêmes d'un même phénomène.

L'attention sélective aurait également un rôle pour l'accès à la conscience. Dans des expériences de psychologie, ses effets sont d'autant plus remarquables que le sujet testé est impliqué dans une tâche cognitive. Les expériences de [Simons et Chabris, 1999] montrent en particulier que nous ne sommes pas conscients de l'ensemble des informations qui se projettent sur la rétine : par exemple, lorsqu'un sujet est impliqué dans une tâche cognitive consistant à compter les passes que se font des joueurs de basket, il n'est pas conscient d'un distracteur qu'un observateur passif perçoit sans difficulté. Les expériences de *change blindness* [Rensink et al., 1997] illustrent également cette sélectivité de l'attention. Ces expériences consistent à présenter successivement deux images identiques à l'exception d'une partie, la transition se faisant par un clignotement. Il est remarquable de constater que les sujets ont du mal à détecter la différence entre ces images. Par contre, dès lors qu'on a pris conscience de cette différence, il n'est plus possible de l'ignorer.

Les différents exemples mentionnés précédemment mettent l'accent sur la sélectivité de l'attention visuelle. Comme nous allons le voir dans ce chapitre, l'attention visuelle peut être guidée par différents facteurs. On distingue en général deux familles de facteurs qu'on appelle facteurs exogènes et facteurs endogènes [Jonides, 1981, Corbetta et Shulman, 2002]. Les facteurs exogènes regroupent l'ensemble des propriétés physiques des stimuli dans leur environnement (e.g. contraste, nouveauté) tandis que les facteurs endogènes dépendent de buts internes volontaires (e.g. la couleur d'un objet recherché). Dans la section 1.1, on présente les résultats de différentes expériences de psychologie qui indiquent quels facteurs visuels peuvent guider l'attention. Lorsque l'attention est guidée par des facteurs exogènes, on parle d'attention *bottom-up* ou *stimulus driven*. Dans la section 1.2, on étudie, en exposant les résultats de quelques expériences d'élec-

trophysiologie, quels attributs visuels peuvent guider le déploiement “volontaire” de l’attention visuelle. Cette forme d’attention visuelle est appelée attention *top-down* puisqu’elle est guidée par des connaissances a priori ou des buts internes. On verra notamment qu’une information spatiale ou qu’une information relative aux attributs visuels d’une cible sont des facteurs qui guident l’attention visuelle. Enfin, dans la section 1.4, on présente une revue des modèles proposés pour l’attention visuelle.

1.1 Attention visuelle exogène guidée par les propriétés physiques des stimuli

[Desimone et Duncan, 1995] définissent l'attention exogène ou *bottom-up* comme un mécanisme qui dérive de biais intrinsèques, ou appris, des systèmes perceptifs en faveur de certains types de stimuli. Ces biais s'observent dans le paradigme de recherche de trait proposé dans la *Feature Integration Theory* (FIT) [Treisman et Gelade, 1980]. Les auteurs distinguent deux paradigmes de recherche visuelle : une recherche parallèle et une recherche en série. Selon ces auteurs, la recherche d'une cible définie par un unique trait visuel (trouver un objet rouge parmi des objets bleus) est rapide et ne nécessite pas l'intervention de l'attention. Elle est caractérisée par un temps de recherche de la cible indépendant du nombre de distracteurs¹ (figure 1.1a). Au contraire, une recherche de conjonction de traits dépend linéairement du nombre de distracteurs (figure 1.1b, trouver la barre verticale noire parmi des barres horizontales noires et des barres verticales blanches). Dans ce cas, les auteurs suggèrent que l'attention est déployée de manière séquentielle sur tous les distracteurs jusqu'à trouver la cible.

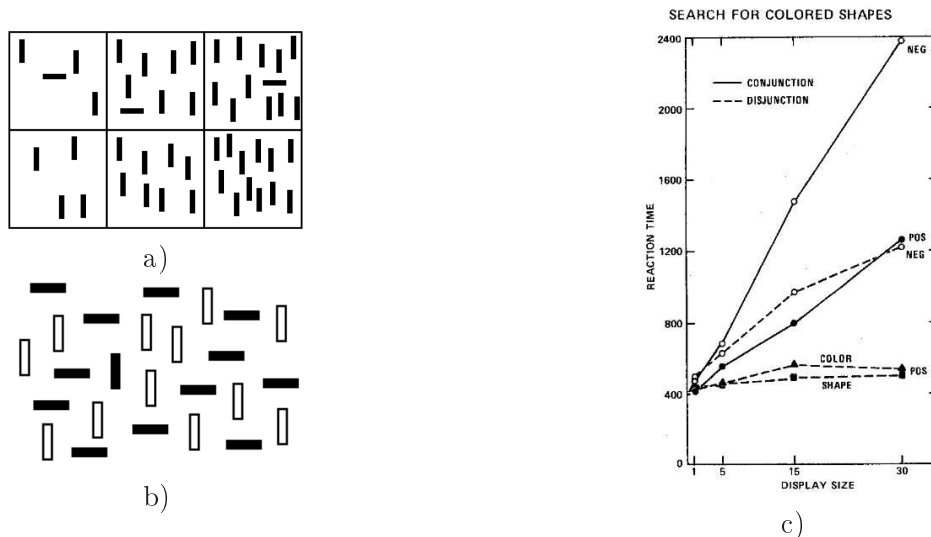


FIG. 1.1 – a) On détecte très rapidement la présence ou l'absence d'une barre horizontale. Le temps mis pour réaliser la tâche est indépendant du nombre de distracteurs. b) Au contraire, le temps mis pour trouver une barre verticale noire dépend du nombre de distracteurs (barres verticales blanches ou horizontales noires). c) Courbes typiques de temps de réponse en fonction du nombre de distracteurs dans des recherches simples ou par conjonction. a), b) adapté de [Wolfe, 2000]. c) adapté de [Treisman et Gelade, 1980]

Cette dichotomie brutale entre recherche parallèle et recherche séquentielle a été critiquée par différents auteurs. En effet, certaines recherches par conjonction peuvent être relativement efficaces [McLeod et al., 1988]. D'autre part, comme le souligne [Wolfe, 1998], la recherche d'une conjonction de traits pourrait profiter de l'extraction des traits préattentifs en préselectionnant un sous ensemble des stimuli qui sont plus probables d'être la cible plutôt qu'en parcourant tous les stimuli. [Duncan et Humphreys, 1989] propose d'introduire la notion de similarité entre la

¹Les distracteurs sont définis comme l'ensemble des stimuli qui ne sont pas la cible

cible et les distracteurs et entre les distracteurs eux-mêmes. Plus la cible est similaire aux distracteurs ou plus les distracteurs sont similaires entre eux, plus la tâche est difficile et nécessite plus de temps. [Wolfe, 1998] préfère parler de recherche efficace et inefficace. L'efficacité d'une recherche visuelle est estimée en calculant la pente de la courbe représentant le temps de réaction en fonction du nombre de distracteurs (figure 1.1c). Plus la pente est importante, moins la recherche est efficace.

Les auteurs s'accordent sur le fait que des traits visuels sont extraits en parallèle, en accord avec la sélectivité des cellules dans les aires visuelles, et que ces derniers peuvent guider le déploiement de l'attention. Différents travaux identifient les caractéristiques visuelles pouvant être considérées comme des traits préattentifs. [Wolfe, 2005] en propose une liste qu'il classe en probable, possible et improbable. Dans la catégorie des traits préattentifs probables, il distingue entre autres la couleur, l'orientation, la taille, la courbure, le mouvement. Ces traits préattentifs sont généralement déterminés en présentant un ensemble de stimuli dont l'un diffère des autres selon l'une de ces dimensions. Si en augmentant le nombre de distracteurs, le temps mis pour trouver la cible est constant, alors on a des indices que ce trait peut être préattentif. Néanmoins, ce constat ne se vérifie pas systématiquement et il existe des configurations pour lesquelles une recherche de conjonction de traits peut conduire à une courbe temps de réaction versus nombre de distracteurs de pente proche de zéro. La figure 1.2 présente l'une de ces configurations : le grand carré blanc est rapidement localisé même si on augmente le nombre de distracteurs. Pour autant, la caractéristique "grand-blanc" a très peu de chance d'être un trait préattentif. Une seconde expérience permettant d'identifier ces traits préattentifs repose sur une asymétrie de la recherche visuelle : il est plus facile de détecter la présence d'un trait que son absence. Par exemple, il est plus simple de détecter une lettre inversée parmi des lettres droites que l'inverse (figure 1.2b).

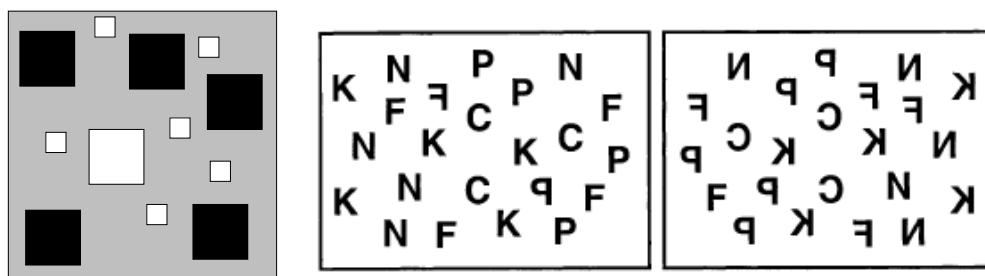


FIG. 1.2 – a) La recherche de la conjonction grand-blanc est efficace sans pour autant que la caractéristique grand-blanc ne puisse être considérée comme un trait préattentif. Adapté de [Wolfe, 2005]. b) Asymétrie dans la recherche visuelle. Il est plus facile de détecter la présence d'un trait que son absence. Adapté de [Wolfe, 2001]

Dans la proposition originelle de [Treisman et Gelade, 1980], l'attention est déployée séquentiellement et permet de sélectionner un stimulus et de lier les traits visuels extraits en parallèle par le filtrage préattentif. Ce second point est connu sous le nom de *binding problem* : si différents attributs sont extraits en parallèle et qu'on présente deux objets, comment assurer le lien entre les attributs d'un même objet. Un déficit dans la liaison des attributs d'un même objet conduirait à la perception de conjonctions virtuelles (*illusory conjunctions*) [Treisman et Schmidt,

1982] (la présentation d'une barre horizontale bleue et d'une barre verticale verte conduirait, si la liaison entre les attributs n'est pas correctement assurée, à la perception d'une barre horizontale verte ou d'une barre verticale bleue). Dans la FIT, le *binding problem* est résolu parce que l'attention est dirigée spatialement sur la région qu'occupe un objet. Cette idée fait référence à la métaphore du *spotlight* attentionnel qui suggère que l'attention est dirigée sur une région de l'espace et favorise le traitement de l'information contenue dans cette région. Si l'attention est dirigée sur une région spatiale, elle favorise le traitement des attributs du stimulus à cette position. L'association des attributs visuels d'une cible est ainsi une conséquence du fait qu'ils appartiennent à la même cible, occupant ainsi la même région de l'espace visuel. Cette idée met en avant un rôle facilitateur de l'attention.

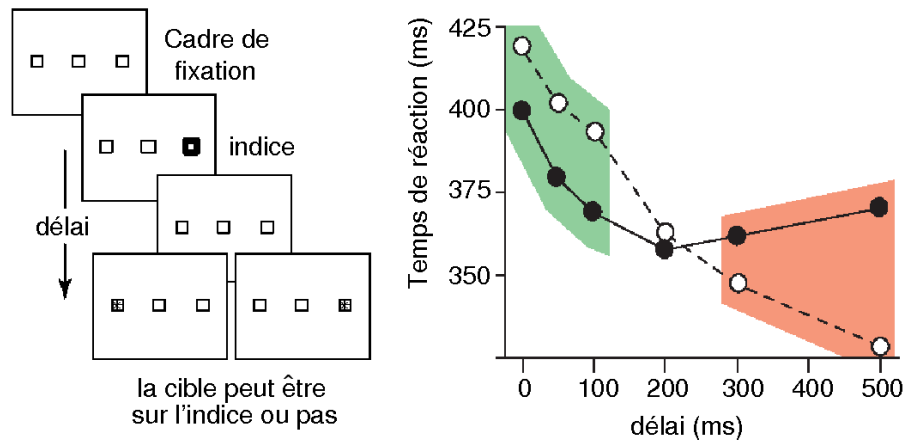


FIG. 1.3 – Gauche : paradigme expérimental permettant d'observer le phénomène d'inhibition de retour. Une position spatiale parmi deux est indiquée puis une cible est présentée à l'une de ces deux positions. On mesure le temps de réaction du sujet vers la cible, en fonction du délai entre l'indice et la cible. Droite : Temps de réaction en fonction du délai lorsque l'indice et la cible sont présentés à la même position (courbe en trait plein) et lorsqu'ils sont présentés à des positions différentes (courbe en pointillés). On observe un effet facilitateur pour un délai court et un effet inhibiteur pour un délai long.

L'un des phénomènes mettant en avant les effets facilitateurs et inhibiteurs de l'attention est l'inhibition de retour (IOR) introduite par [Posner et al., 1980]. L'expérience classique (figure 1.3) consiste à placer un sujet humain devant un dispositif visuel composé de trois points en lui donnant comme consigne de fixer le point central. Un des deux points périphériques est alors illuminé brièvement. Cet indice sert à attirer l'attention du sujet vers cette position spatiale. Au bout d'un temps variable, un des deux points est illuminé (la cible) et le sujet doit exécuter une réponse manuelle associée à la cible. On distingue le cas valide où l'indice et la cible occupent la même position et le cas invalide où l'indice et la cible sont présentés à des positions différentes. On constate que si le délai entre la présentation de l'indice et de la cible est court (≤ 150 ms), le temps de réaction moyen dans le cas valide est plus court, ce qui correspond à un effet facilitateur, et plus long dans le cas invalide ce qui correspond à un effet inhibiteur. Au contraire, si le délai est plus long (≥ 300 ms), le temps de réaction moyen est plus long dans le cas valide que dans le cas invalide. Ces résultats suggèrent que l'attention ne reste pas sur l'indice et qu'il lui faut un certain temps pour revenir sur une région sur laquelle elle a été précédemment déployée. Dans [Rafal et al., 1989], les auteurs distinguent le cas où l'indice est exogène (la position spa-

tiale de l'indice est explicite comme dans l'expérience de Posner et Cohen) du cas où l'indice est endogène (la position spatiale est indiquée par une flèche centrale). Ils montrent en particulier que lorsque la tâche consiste à détecter une cible sans bouger les yeux, l'inhibition de retour ne s'observe que si l'indice est exogène, c'est à dire explicite. Par ailleurs, ils montrent également que si la tâche nécessite de préparer une saccade dans la direction de l'indice, l'inhibition de retour s'observe quelle que soit la nature de l'indice. Ce dernier résultat suggère un lien entre le déploiement de l'attention et la préparation d'une saccade oculaire sur lequel nous reviendrons dans le paragraphe 1.2.1. D'après [Klein, 2000], l'inhibition de retour permettrait de faciliter la recherche visuelle en empêchant l'attention d'être focalisée sur une région de l'espace précédemment examinée.

Le phénomène d'*attentional blink* [Raymond et al., 1992] met également en avant la dimension temporelle de l'allocation de l'attention. Dans cette expérience, on présente rapidement une succession de stimuli à une même position. Lorsque le sujet répond à une première cible T1, ses performances pour répondre à une cible T2 chutent si le délai entre la présentation des deux cibles est compris entre 100ms et 500ms. En dehors de cette fenêtre temporelle, les performances sont meilleures. Cette expérience suggère que lorsque l'attention est portée sur un stimulus qui est traité pour être identifié, il existe un intervalle de temps pendant lequel une cible présentée à la même position spatiale ne peut être identifiée. Ces résultats peuvent être interprétés dans le cadre de l'inhibition de retour. [Kristjánsson et Nakayama, 2002] ont étendu ces résultats en étudiant le cas où les deux cibles peuvent être présentées à différentes positions. Ils reproduisent les effets d'*attentional blink* et montrent par ailleurs qu'il est atténué lorsque les cibles T1 et T2 sont éloignées spatialement.

1.2 Attention visuelle endogène guidée par les buts

En plus des propriétés physiques des stimuli dans leur environnement pouvant guider le déploiement de l'attention, différentes expériences montrent qu'une connaissance a priori biaise le comportement. [Yarbus, 1967] montre par exemple que le parcours oculaire pendant l'observation d'une scène est corrélé à l'objectif poursuivi par cette recherche (figure 1.4, déterminer l'âge des personnes présentes dans la scène, déduire leur activité). Comme nous le verrons par la suite, certains auteurs suggèrent que le parcours oculaire et le déploiement de l'attention sont intimement liés. Néanmoins, nous nous limitons dans ce chapitre au déploiement de l'attention sans mouvement oculaire. L'expérience de Posner présentée dans la section précédente est un exemple illustrant qu'une connaissance a priori a une influence sur le comportement, même sans mouvement oculaire, le sujet devant garder les yeux fixés sur le point de fixation.

Différentes expériences montrent les corrélats neuronaux du biais apporté par une connaissance a priori, véritable facteur endogène permettant de guider l'attention. On distingue généralement deux types de facteur endogène :

- l'attention guidée par des informations spatiales
- l'attention guidée par des attributs visuels

Nous proposons dans les deux sections qui suivent une revue des expériences mettant en exergue

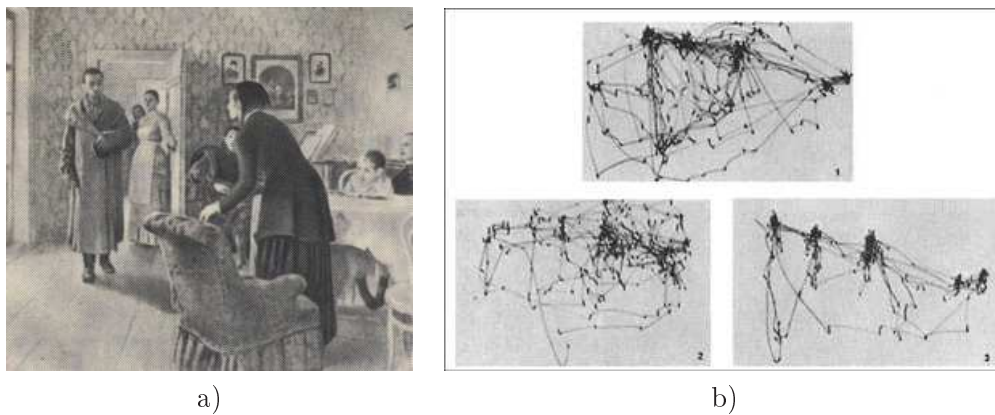


FIG. 1.4 – a) Dans son expérience [Yarbus, 1967] présente une scène qu'un sujet doit examiner avec différentes consignes. Pendant que le sujet réalise la tâche, les mouvements de ses yeux sont enregistrés. b) Mouvements oculaires enregistrés pour trois objectifs différents, dans l'ordre : "Observation sans consigne", "Estimer les conditions matérielles de la famille", "Déterminer l'âge des personnes". Adapté de [Yarbus, 1967]

ces différentes contributions. Il est à noter que certains travaux récents suggèrent une troisième composante pour l'attention, basée sur les objets mais que nous ne développerons pas ici [Blaser et al., 2000]. En effet, l'objet de ce chapitre n'est pas de présenter de manière exhaustive tous les travaux relatifs à l'attention visuelle mais de sélectionner les propriétés essentielles qui guideront la définition du modèle qu'on présentera dans le chapitre 2.

1.2.1 Attention guidée par des informations spatiales

Dans [McAdams et Maunsell, 1999], les auteurs étudient les effets de l'attention spatiale sur la réponse de cellules dans V4². Les auteurs réalisent des expériences *delayed match-to-sample* en enregistrant l'activité de neurones de V4. Dans une expérience *delayed match-to-sample* classique, on présente une cible, suivie d'un masque, suivi d'un affichage constitué de deux stimuli : la cible et un distracteur. Le sujet doit alors répondre (par exemple par une réponse manuelle ou par un mouvement oculaire) en direction de la cible. Les auteurs utilisent ici cette expérience pour étudier la manière dont la réponse de certaines cellules est modifiée selon que le champ récepteur des cellules enregistrées couvre la cible ou le distracteur. Dans leur étude, les auteurs montrent en particulier que lorsque l'attention spatiale est dirigée dans le champ récepteur d'un neurone sensible à l'orientation, sa courbe de réponse en fonction de l'orientation du stimulus présenté est multipliée par un gain. Ce gain multiplicatif augmente en moyenne de 32% la réponse des neurones sans altérer sa sélectivité.

Un simple gain multiplicatif de la réponse prédit que l'effet de l'attention augmente avec le

²V4 est une aire visuelle impliquée dans le traitement de caractéristiques visuelles simples comme la couleur ou l'orientation. Elle est représentée sur la figure 1 ; on reviendra plus en détails sur cette aire dans la section 3.3.1

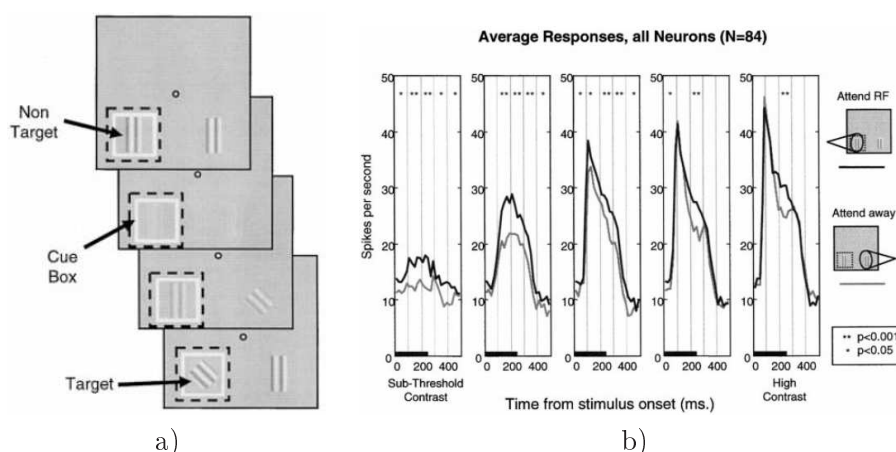


FIG. 1.5 – a) Deux séquences de stimuli sont présentées à deux positions spatiales. La séquence pertinente est indiquée au début de l'expérience. L'animal doit détecter une cible orientée à 45° à cette position spatiale. b) Réponse moyenne de cellules dans V4 à différents contrastes avec l'attention spatiale dirigée dans leur champ récepteur (ligne en trait plein) et l'attention spatiale dirigée en dehors de leur champ récepteur (ligne en pointillés). Adapté de [Reynolds et al., 2000]

contraste des stimuli. Or, plusieurs études montrent que l'effet de l'attention est maximal pour un intervalle de contraste puis décroît au dehors. En particulier [Reynolds et al., 2000] montrent que diriger l'attention sur une région de l'espace augmente la sensibilité de neurones de V4 dont le champ récepteur couvre cette région. Durant toute l'expérience (1.5a), l'animal doit garder son regard sur un point de fixation. L'expérience consiste alors à présenter une succession de stimuli à deux positions spatiales, une dans chaque hémichamp. Les séquences de stimuli sont constituées d'un stimulus sinusoïdal horizontal (*grating*) dont ils font varier le contraste puis d'une cible orientée à 45°. Au début de l'expérience, une des deux régions de l'espace est indiquée par un cadre blanc. Cet indice indique la position spatiale à laquelle la cible (contraste orienté à 45°) doit être détectée : le singe doit relâcher un levier lorsque la cible apparaît à cette position sans tenir compte de la séquence présentée à la position non indiquée. Comme toutes les expériences visant à étudier le rôle de l'attention visuelle, cet artifice doit contraindre le singe à porter son attention sur une région de l'espace. Les auteurs enregistrent l'activité de neurones de V4 dont le champ récepteur couvre soit la séquence pertinente, soit la séquence non pertinente. Ils obtiennent ainsi la réponse de neurones de V4 à des stimuli de différents contrastes, lorsque l'attention est dirigée dans leur champ récepteur et lorsque l'attention est dirigée en dehors de leur champ récepteur. La figure 1.5b résume une partie de leurs résultats. Elle représente de gauche à droite la réponse moyenne de 84 neurones à des stimuli de contraste croissant. La courbe noire correspond à la réponse lorsque l'attention spatiale est portée sur le champ récepteur de la cellule, la courbe grise lorsque l'attention est portée en dehors du champ récepteur. On constate en particulier que la fréquence de décharge spontanée des cellules est accrue, même en l'absence d'une stimulation (cas d'un contraste nul de la figure la plus à gauche). Par ailleurs, l'influence de l'attention spatiale sur la fréquence de décharge dépend du contraste du stimulus présenté dans le champ récepteur de la cellule enregistrée. On observe cette dépendance sur les figures 1.5b mais elle est accentuée lorsqu'on trace le changement de la réponse, avec ou sans attention spatiale, en fonction du contraste du stimulus, représenté par la courbe en tirets sur la figure 1.6. Ces résultats conduisent les auteurs à proposer que les effets de l'attention spatiale peuvent

être modélisés par un gain multiplicatif sur les entrées (*contrast gain model*), dont l'amplitude dépend du contraste du stimulus occupant le champ récepteur de la cellule.

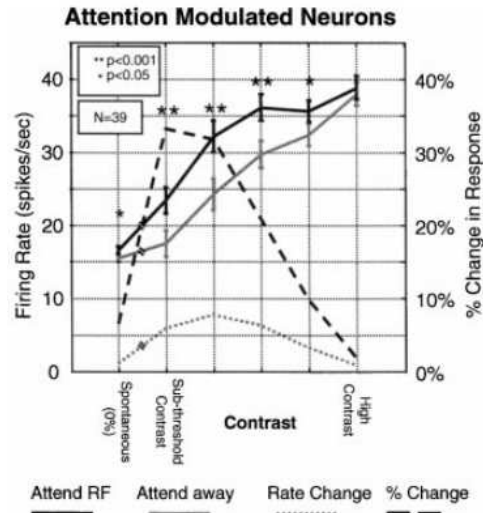


FIG. 1.6 – Réponse de neurones de V4 en fonction du contraste du stimulus dans le cas où l'attention est dirigée dans le champ récepteur et le cas où l'attention est dirigée en dehors (lignes en trait plein). Le pourcentage de changement de la réponse est représenté en pointillés. Adapté de [Reynolds et al., 2000]

En partant de ces observations, on cherche à savoir comment les effets de l'attention visuelle spatiale peuvent être modélisés. En général, les auteurs distinguent trois façons de modéliser les effets de l'attention spatiale sur la réponse d'une cellule en fonction du contraste d'un stimulus qui occupe son champ récepteur ; ses effets peuvent être modélisés par un gain multiplicatif sur :

- les entrées de la cellule (*contrast gain model*)
- la réponse de la cellule (*response gain model*)
- l'activité de la cellule (*activity gain model*)

Le premier modèle (figure 1.7a) établit deux prédictions :

- l'influence de l'attention est plus important pour des contrastes modérés que pour des contrastes petits ou grands
- l'attention ne modifie pas le taux de décharge spontanée (réponse à un stimulus de contraste nul)

Le second modèle (figure 1.7b) prédit que l'influence de l'attention est d'autant plus important que le contraste du stimulus est important, sans modifier le taux de décharge spontanée. Le troisième modèle (figure 1.7c) conduit à des prédictions similaires au modèle précédent à la différence que le taux de décharge spontanée est également augmenté. Dans une étude récente [Williford et Maunsell, 2006], les auteurs soulignent qu'il n'est pas encore possible de trancher en faveur de l'un ou l'autre de ces modèles, tout en soulignant que seul le gain sur l'activité reproduit l'augmentation du taux de décharge spontanée observée dans différentes études [Luck et al., 1997].

Les travaux cités précédemment s'intéressent essentiellement aux effets de l'attention visuelle spatiale sur la réponse d'une cellule dont le champ récepteur est occupé par un stimulus et soulignent que ses effets peuvent être modélisés par un gain multiplicateur même s'il n'est pas encore

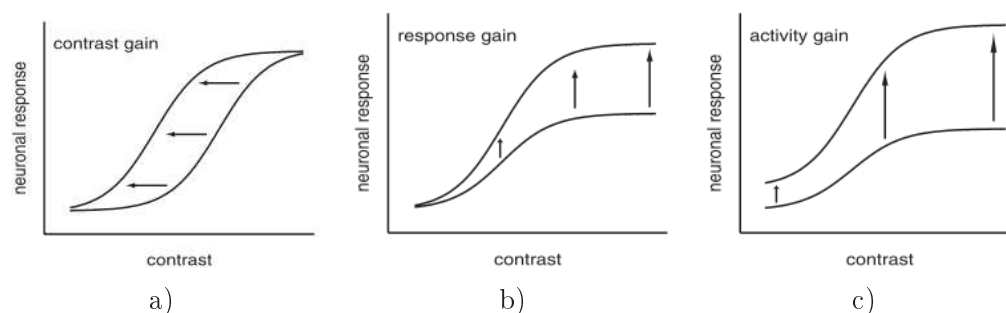


FIG. 1.7 – Différents modèles des effets de l’attention spatiale sur la sensibilité d’un neurone. a) Les effets de l’attention sont modélisés par un gain sur les entrées. b) Les effets de l’attention sont modélisés par un gain sur la réponse (sans moduler le taux de décharge spontanée). c) Les effets de l’attention sont modélisés par un gain sur l’activité. Adapté de [Williford et Maunsell, 2006]

possible de trancher pour un gain sur l’entrée, la réponse ou l’activité d’une cellule.

Nous venons de voir un effet de l’attention visuelle spatiale sur la réponse d’une cellule lorsque son champ récepteur est occupé par un seul stimulus. Un second effet de l’attention visuelle spatiale intéressant est observé lorsque plusieurs stimuli occupent le même champ récepteur. Dans [Desimone et Duncan, 1995], les auteurs s’intéressent à l’activité de cellules dans V4 et IT³ lorsque deux cibles sont présentes dans leur champ récepteur. L’une de ces cibles évoque une réponse forte (good stimulus), l’autre une réponse faible (poor stimulus) lorsqu’elles sont présentées seules. Ils montrent en particulier que si l’attention spatiale est dirigée en dehors du champ récepteur d’une cellule de V4 ou IT, sa réponse est intermédiaire entre la réponse forte et la réponse faible. Lorsque l’attention spatiale est dirigée vers son champ récepteur, la réponse est biaisée en faveur de la cible dont l’attribut est pertinent pour le comportement. Nous verrons plus en détail dans la prochaine section cette seconde forme de l’attention, guidée par les attributs visuels. Ces expériences ont conduit les auteurs à proposer le paradigme de compétition biaisée (*biased competition*) qui suggère que lorsque plusieurs stimuli occupent le champ récepteur d’une cellule, et si l’attention spatiale est dirigée vers ce champ récepteur, alors les stimuli entrent en compétition pour être représentés par la cellule. Une des raisons qui pourraient expliquer cette compétition est l’augmentation de la taille des champs récepteurs le long de la voie ventrale (0.1° pour V1, 3° pour V4, 6° pour TEO, 25° pour TE) qui implique, que sans compétition, il y aurait ambiguïté quant à l’information représentée par ces cellules.

Jusqu’à maintenant, nous avons montré les effets de l’attention visuelle spatiale sans en discuter l’origine. La “théorie prémotrice de l’attention” [Rizzolatti et al., 1987, Craighero et Rizzolatti, 2005] fournit une hypothèse quant à l’origine des signaux qui causent les effets de l’attention visuelle spatiale. Selon ces auteurs, le déploiement de l’attention spatiale résulte de l’activation des circuits également impliqués dans la programmation d’une activité motrice comme un mou-

³IT : le cortex inférotemporal est impliqué dans le traitement de caractéristiques visuelles complexes comme des formes ou des visages. Il est représenté sur la figure 1 ; on reviendra plus en détails sur cette aire dans la section 3.3.1

vement oculaire : les circuits du contrôle oculomoteur et du déploiement de l'attention spatiale partageraient un grand ensemble de structures [Corbetta et al., 1998]. Les résultats de différentes expériences renforcent cette théorie. Dans les expériences de [Hoffman et Subramaniam, 1995], les sujets doivent exécuter une saccade vers un stimulus tout en détectant une cible présentée à la même position ou à une position différente juste avant le mouvement. Les auteurs montrent alors que la détection de la cible est d'autant plus efficace qu'elle occupe la position cible de la saccade. Pour étudier le lien entre le contrôle oculomoteur et le déploiement de l'attention spatiale, on peut également s'intéresser aux conséquences de la stimulation ou de l'inactivation de structures dont on montre qu'elles sont également impliquées dans le contrôle oculomoteur. En effet, si le déploiement de l'attention spatiale est assuré par des structures également impliquées dans le contrôle oculomoteur, on devrait constater des déficits attentionnels en perturbant ces dernières. C'est le constat que font différents auteurs en stimulant ou en inactivant temporairement deux principales structures dont on connaît l'implication dans le contrôle oculomoteur : FEF⁴ et SC⁵ [Moore et Fallah, 2001, Moore et Fallah, 2004, Wardak et al., 2006, McPeck et Keller, 2004].

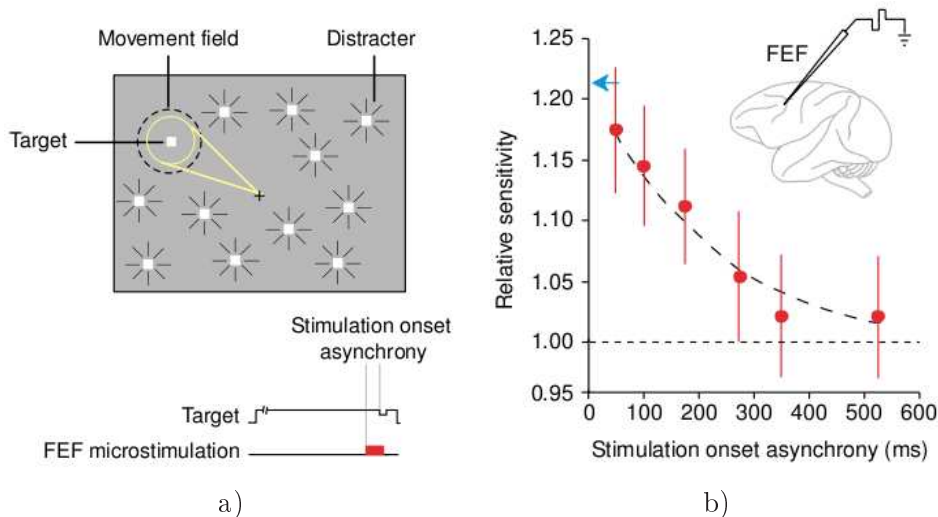


FIG. 1.8 – a) Le singe doit détecter une variation d'intensité d'une cible périphérique, en gardant le regard fixé sur la croix, tout en ignorant un distracteur flashé séquentiellement à des positions aléatoires. Pendant la tâche, un site de FEF est microstimulé avec un délai précédant la variation d'intensité de la cible variable. b) Sensibilité relative en fonction du délai entre la microstimulation et la variation d'intensité de la cible. Plus le délai est court, plus les performances sont équivalentes au cas où aucun distracteur n'est présenté, caractérisé par une sensibilité relative d'environ 1.20 (flèche sur l'axe des ordonnées). Cet effet n'est observé que lorsque la cible occupe le champ de mouvement du site stimulé dans FEF. Adapté de [Moore et Fallah, 2004]

Dans [Moore et Fallah, 2001, Moore et Fallah, 2004], les auteurs étudient les conséquences comportementales et physiologiques d'une microstimulation de FEF (figure 1.8). En stimulant sélectivement FEF avec un courant suffisamment fort, il est possible de susciter un mouvement

⁴FEF : *frontal eye field* est l'une des principales aires oculomotrices. Elle est située dans le lobe frontal. On reparlera plus en détails de cette aire en section 3.3.3

⁵SC : *superior colliculus* est l'une des principales aires oculomotrices. Elle est située au sommet du mésencéphale. On reparlera plus en détails de cette aire en section 3.2.2

oculaire en direction du champ de mouvement du site stimulé. Après avoir identifié le champ de mouvement d'un site de FEF, les auteurs stimulent ce site avec un courant plus faible afin de ne pas évoquer de mouvement oculaire. Ils constatent alors que cette microstimulation favorise la détection d'une cible positionnée dans le champ récepteur aligné avec le champ de mouvement du site stimulé. Ils montrent également dans une autre étude [Moore et Armstrong, 2003] qu'une microstimulation de FEF produit des effets dans V4 similaires aux effets de l'attention spatiale mentionnés précédemment [Reynolds et Chelazzi, 2004]. Dans [Wardak et al., 2006], les auteurs montrent qu'en inactivant temporairement un site de FEF, non seulement la faculté de produire des saccades visuellement guidées est altérée mais la recherche d'une cible visuelle parmi des distracteurs, sans mouvement oculaire, l'est également.

Des expériences d'activation et d'inactivation réversible de SC ont également été réalisées [Cavanaugh et Wurtz, 2004, Müller et al., 2005, McPeck et Keller, 2004] et ont conduit à des résultats similaires. Ces résultats viennent renforcer la proposition de la théorie prémotrice de l'attention selon laquelle le contrôle oculomoteur et le déploiement de l'attention spatiale sont intimement liés.

Dans cette section, nous avons vu deux effets de l'attention visuelle spatiale sur la réponse de certains neurones. Nous retiendrons que l'attention spatiale favorise le traitement d'un signal si celui-ci n'est pas suffisamment représenté et qu'elle biaise également le traitement de l'information visuelle pour en favoriser une partie au détriment d'une autre. Enfin, nous avons également vu que ces effets attentionnels pourraient être expliqués par des afférences des structures oculomotrices, en accord avec la théorie prémotrice de l'attention. Dans la prochaine section, nous discutons de l'influence d'une connaissance a priori d'attributs visuels pertinents (e.g. couleur, orientation) sur le guidage de l'attention.

1.2.2 Attention guidée par les attributs visuels

Plusieurs études montrent que la connaissance a priori d'une caractéristique d'une cible à rechercher influe à la fois sur le temps de réponse du sujet pour la tâche mais également sur le niveau physiologique. Nous traiterons dans ce paragraphe des attributs non spatiaux comme l'orientation, la couleur, la direction du mouvement. Pour rechercher de tels corrélats neuronaux, il est naturel de commencer par étudier les cellules le long de la voie ventrale, du cortex visuel au cortex inférotemporal. Comme nous le verrons plus en détail dans la section 3.3.1, les structures corticales le long de cette voie sont spécialisées dans le traitement d'attributs visuels tels que la couleur, l'orientation ou des formes plus ou moins complexes comme par exemple les visages.

Dans [Moran et Desimone, 1985], les auteurs montrent l'influence de l'attention sur la réponse de neurones dans V4 et IT. Dans leur expérience, ils présentent deux stimuli qui peuvent ou non occuper simultanément le champ récepteur d'une cellule qu'ils enregistrent. L'un des stimuli évoque une forte réponse de la part de la cellule ("bon stimulus"), l'autre évoque une faible réponse ("mauvais stimulus"). Ils observent alors les propriétés suivantes :

- lorsque les deux stimuli occupent le champ récepteur de la cellule enregistrée, sa réponse est principalement déterminée par la réponse au stimulus pertinent présenté seul ;

– lorsque seul le bon stimulus occupe le champ récepteur de la cellule, sa réponse est la même, que l'attention soit portée sur le bon stimulus ou sur le mauvais. Ils concluent ainsi que lorsque deux stimuli occupent le champ récepteur d'un neurone, sa réponse est déterminée par les propriétés du stimulus pertinent, comme si l'influence du distracteur était supprimée.

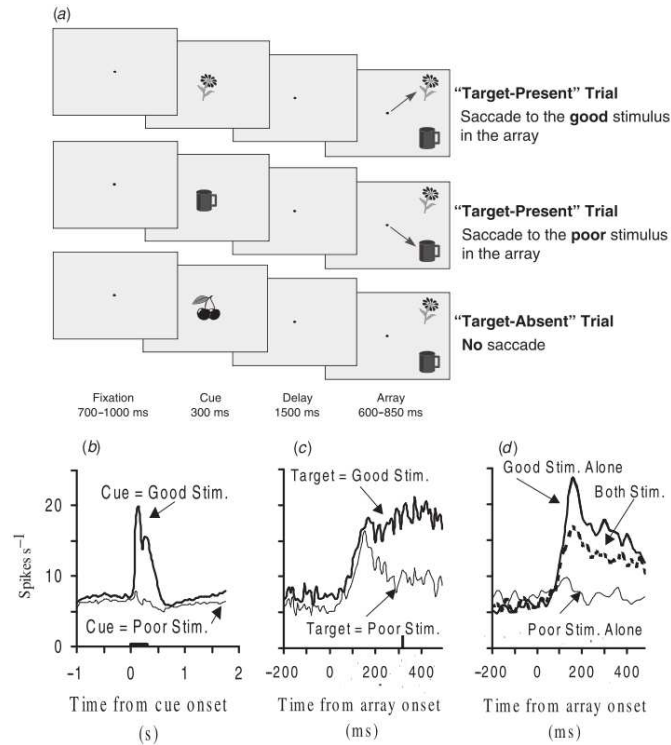


FIG. 1.9 – a) Trois paradigmes d'une tâche *delayed match-to-sample* sont considérés : lorsque la cible est présente et qu'elle suscite une réponse forte de la part de la cellule enregistrée, lorsque la cible est présente et qu'elle suscite une réponse faible et lorsque la cible n'est pas présente. b) Réponse à la présentation de l'indice et pendant le délai. Elle est plus forte pour le stimulus préféré (*good stim*) que pour le stimulus non préféré (*poor stim*), même pendant le délai. c) À la présentation de l'ensemble des stimuli, la réponse est d'abord identique quelle que soit la cible puis diverge significativement en direction de la réponse au stimulus recherché présenté seul. d) Lorsque la cible est absente, la réponse de la cellule est intermédiaire entre la réponse aux stimuli préféré et non préféré présentés seuls. Adapté de [Chelazzi et al., 1998]

Cette compétition en faveur du stimulus attendu est également mise en évidence dans des travaux ultérieurs [Chelazzi et al., 1998, Reynolds et al., 1999]. [Chelazzi et al., 1998] soulignent en particulier l'évolution temporelle de l'activité d'un neurone de IT dont le champ récepteur est occupé par deux stimuli (bon et mauvais) pendant une tâche *delayed match-to-sample*. Cette expérience consiste à présenter un objet cible que l'animal doit mémoriser et retrouver dans un ensemble de plusieurs objets présentés après un délai. Les auteurs distinguent trois paradigmes suivant que la cible est présente ou non et qu'elle évoque une réponse forte ou faible lorsqu'elle est présentée seule (figure 1.9a). Comme [Moran et Desimone, 1985], ils constatent qu'à la présentation des objets, si la cible (définie par des attributs vers lesquels l'attention est portée) est

présente, la réponse des neurones est forte (le bon stimulus est présent dans le champ récepteur) puis diverge significativement comme si seul le stimulus attendu était présent (figure 1.9c). En particulier, si la cible est le stimulus qui évoque la réponse la plus faible, la réponse est effectivement faible même si le stimulus évoquant une réponse forte est encore présent dans le champ récepteur de la cellule. Ils montrent également que si la cible est absente, la réponse de la cellule est intermédiaire entre la réponse faible et la réponse forte (figure 1.9d). Ces résultats expérimentaux sont à la base des principes du paradigme de “compétition biaisée” proposé par [Desimone et Duncan, 1995]. Selon ce paradigme, lorsque plusieurs stimuli occupent le champ récepteur d’une cellule, ils entrent en compétition pour être représentés par cette cellule. Cette compétition est biaisée en faveur du stimulus qui possède les attributs attendus. Par contre, s’ils n’occupent pas simultanément le champ récepteur de la cellule, l’attention n’a plus d’effet sur la réponse. Selon les auteurs, cette compétition permet de représenter l’information pertinente en supprimant l’information non pertinente, atténuant ainsi l’ambiguïté de l’information représentée par les cellules le long de la voie ventrale dont la taille des champs récepteurs augmente (de 3° dans V4 à 25° dans TE).

Les résultats précédents illustrent un effet suppressif local de l’attention guidée par les attributs, lorsque plusieurs stimuli occupent un même champ récepteur : lorsque plusieurs stimuli, dont l’un est la cible, sont présentés dans le champ récepteur d’une cellule, la réponse de la cellule est uniquement déterminée par la réponse à la cible présentée seule, comme si l’influence du distracteur était supprimée. D’autres études montrent une influence globale [Motter, 1994a, Motter, 1994b, Treue et Martínez-Trujillo, 1999]. Dans [Motter, 1994a], les auteurs présentent un indice de couleur au centre du champ visuel puis un ensemble de stimuli (paradigmes *match* et *non-match* de la figure 1.10a). Les stimuli sont disposés de telle manière qu’ils n’occupent pas le champ récepteur d’un même neurone de V4. Après un délai, toutes les cibles sont éteintes sauf deux : l’une est de la couleur de l’indice, l’autre non. L’animal a pour tâche de discriminer l’orientation de la barre dont la couleur est celle de l’indice. Les auteurs considèrent deux cas : dans le premier, le stimulus qui occupe le champ récepteur d’une cellule enregistrée est de la couleur de l’indice (*match*). Dans le second, il n’a pas la couleur de l’indice (*non-match*). Les auteurs montrent que la réponse des neurones est plus forte lorsque le stimulus a la couleur de l’indice que lorsqu’il ne l’a pas. Cette différence de réponse est observée que le stimulus ait les attributs préférés de la cellule (figure 1.10b, haut) ou non préférés (figure 1.10b, bas). Dans [Motter, 1994b], les auteurs utilisent une version modifiée des paradigmes précédents dans laquelle ils changent la couleur de l’indice après la présentation de l’ensemble des stimuli (paradigme *cue switch* de la figure 1.10a). Ils montrent alors que la réponse des neurones de V4 change dynamiquement avec l’indice (figure 1.10c) avec une latence variant de 150 à 300ms selon les cas (*non-match* vers *match* ou *match* vers *non-match*).

Cet effet non spatial de modulation par des attributs visuels est également observé dans MT⁶ [Treue et Martínez-Trujillo, 1999]. Dans cette étude, on présente à un singe deux ensembles de points en mouvement pendant que l’activité de cellules dans MT est enregistrée. L’un de ces ensembles de points se trouve dans le champ récepteur de la cellule, l’autre dans l’hémichamp opposé. L’expérience consiste à déplacer ces deux ensembles de points soit dans la même direction, soit dans une direction opposée tout en “donnant” l’instruction à l’animal de discriminer la

⁶MT (ou V5) est une aire visuelle dont les cellules sont sélectives au mouvement visuel. On reviendra plus en détails sur cette aire dans la section 3.3.1

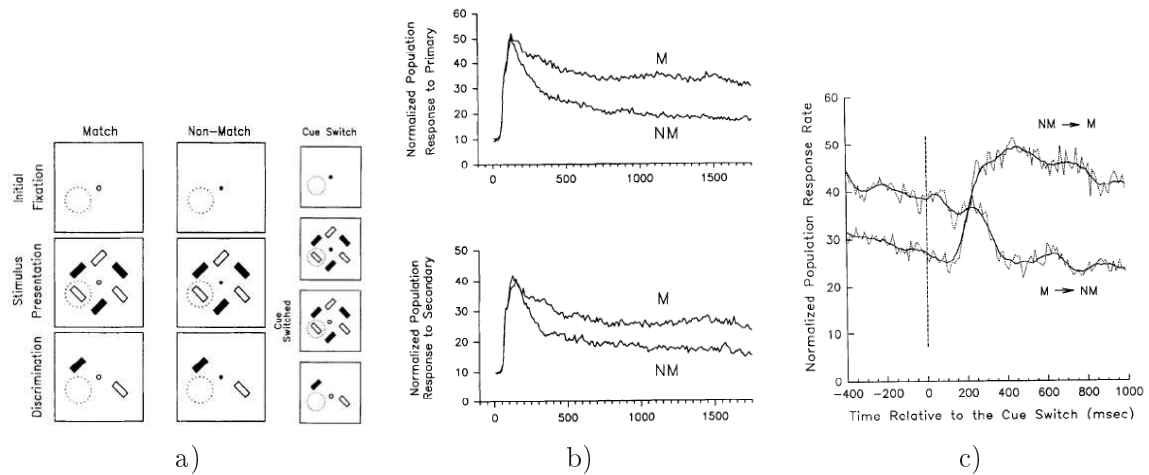


FIG. 1.10 – a) Trois des paradigmes utilisés par [Motter, 1994a, Motter, 1994b]. Pour le sujet, seul le dernier affichage est pertinent (en bas des figures) : le singe doit discriminer l'orientation du stimulus qui a la couleur du point de fixation. Cette tâche permet uniquement de forcer l'animal à se mémoriser la couleur du point de fixation. Avant de réaliser cette tâche de discrimination, on présente au singe un ensemble de cibles. Pendant cette phase de présentation des cibles, le point de fixation peut ou non changer de couleur. Il ne change pas de couleur pour les paradigmes *match* et *non-match* mais change de couleur dans le paradigme *cue-switch*. Pendant que le singe réalise la tâche, on enregistre l'activité d'une cellule dont le champ récepteur est représenté par le cercle en pointillés. b) Que le stimulus qui occupe le champ récepteur de la cellule ait ou non les attributs préférés de la cellule (différence entre les deux graphiques), sa réponse est significativement plus forte lorsque ce stimulus a la couleur du point de fixation (courbe M) que lorsqu'il ne l'a pas (courbe NM). c) Si, pendant la présentation des stimuli, la couleur du point de fixation est changée, on observe que la réponse de la cellule enregistrée change dynamiquement. On peut interpréter ce changement comme une modification de la pertinence du stimulus qui occupe le champ récepteur de la cellule enregistrée. Adapté de [Motter, 1994a, Motter, 1994b]

direction du mouvement de l'ensemble en dehors du champ récepteur. Les auteurs observent que la réponse des cellules est plus forte lorsque l'ensemble de points en dehors du champ récepteur se dirigent dans la même direction que l'ensemble de points dans le champ récepteur, que dans la direction opposée. Ils montrent également que lorsque plusieurs stimuli occupent le même champ récepteur d'une cellule de MT, sa réponse est biaisée en faveur du stimulus sur lequel l'attention est portée, en accord avec le paradigme de "compétition biaisée" proposée par R. Desimone et J. Duncan. Ces résultats conduisent [Treue et Martínez-Trujillo, 1999] à proposer le *feature similarity gain model* selon lequel l'attention sur les attributs a un effet global qui reflète la similarité entre les attributs pertinents pour la tâche dans laquelle le sujet est engagé et la sélectivité des neurones. Cet effet peut être modélisé par un gain multiplicatif sur les entrées de la cellule qui n'altère pas la forme de la sélectivité des cellules.

Les études présentées précédemment illustrent qu'une connaissance a priori d'un attribut pertinent pour le comportement a un effet sur l'activité de cellules dans V4, IT et MT. Cette influence est également observée dans d'autres aires comme V2 ou V1 [Reynolds et al., 1999, Posner et Gilbert, 1999]. Cet effet est non spatial, dépendant des attributs et d'autant plus important que plusieurs stimuli (possédant ou non cet attribut) partagent le champ récepteur d'une même cellule.

1.3 Bilan des résultats expérimentaux sur l'attention visuelle

Dans les sections précédentes nous avons présenté différents travaux expérimentaux soulignant les corrélats psychologiques et physiologiques de l'attention visuelle. Nous retiendrons de ces résultats qu'il y a différents facteurs qui peuvent guider l'attention. D'une part, les caractéristiques physiques des stimuli comme leur apparition soudaine ou leur contraste, favorisent le traitement de ces stimuli. Cette propriété, présentée dans la section 1.1, est remarquablement bien mise en avant par les expériences de recherche visuelle. Intuitivement, on peut comprendre qu'un stimulus qui apparaît soudainement dans le champ visuel soit porteur d'une information qui doit être traitée en priorité (signalant par exemple un danger). D'autre part, les buts internes du sujet guident également l'attention. Nous avons distingué deux formes de buts internes : des buts spatiaux et des buts non spatiaux dont l'influence a été présentée respectivement dans les sections 1.2.1 et 1.2.2. L'un des effets de l'attention visuelle spatiale est d'accentuer la représentation d'un stimulus, dont l'amplitude est par exemple trop faible pour susciter une réponse forte de la part de la cellule. Nous avons vu que cet effet pouvait être modélisé par un gain multiplicatif sur les entrées d'une cellule. Nous avons également vu l'influence de l'attention sur l'activité de cellules dont le champ récepteur est occupé par plusieurs stimuli. En particulier, l'attention guidée par les attributs visuels permet de favoriser le traitement de certains stimuli au détriment des distracteurs.

Dans la prochaine partie, nous présentons trois modèles attentionnels. Ces trois modèles apportent trois contributions distinctes qui nous paraissent intéressantes, à savoir :

- comment représenter l'information visuelle pour guider l'attention *bottom-up*
- le rôle de la compétition pour la représentation
- comment intégrer une connaissance a priori dans un modèle attentionnel

1.4 Revue des principaux modèles attentionnels

Différentes approches ont été proposées pour modéliser l'attention visuelle dont on trouvera une revue détaillée dans [Frintrop, 2005a]. Historiquement, la proposition de [Koch et Ullman, 1985] est la première, et repose sur les principes de la *Feature Integration Theory* (FIT) [Treisman et Gelade, 1980]. Les auteurs proposent d'extraire en parallèle différents traits visuels, les résultats étant représentés par un ensemble de cartes topographiques 2D. Ces cartes sont ensuite combinées dans une carte de saillance à partir de laquelle on détermine quelle région de l'espace visuel doit être traitée en priorité. Un algorithme *winner-take-all* (WTA) est appliqué sur cette carte de saillance pour détecter la région la plus saillante qui capture l'attention. Cette première proposition a été par la suite entendue dans différents travaux [Itti et al., 1998, Milanese, 1993]; nous détaillerons en particulier dans le paragraphe 1.4.1 le système *Neuromorphic Vision Toolkit* (NVT)⁷ [Itti et al., 1998] qui est l'un des systèmes attentionnels les plus connus à l'heure actuelle. Il n'intègre que depuis très récemment un biais représentant des connaissances a priori, c'est la raison pour laquelle nous mettrons essentiellement l'accent sur la manière dont ce système extrait les régions de l'espace visuel les plus pertinentes uniquement sur la base des propriétés intrinsèques de la scène visuelle. Dans la proposition originelle de la FIT, si la cible n'est pas détectée du premier coup, tous les stimuli sont sélectionnés séquentiellement. [Wolfe et al., 1989, Wolfe, 1994, Wolfe et Gancarz, 1996] a été le premier à critiquer cet aspect et propose dans le *guided search model* qu'une connaissance a priori peut guider le déploiement de l'attention.

La métaphore la plus souvent utilisée pour l'attention est la *spotlight metaphor* qui identifie l'attention à un faisceau lumineux sélectionnant une région contiguë de l'espace visuel. La proposition de [Ahmad, 1992] est littéralement une implémentation de cette métaphore. Les *Shifter circuits* [Anderson et VanEssen, 1987, Olshausen et al., 1993, Mozer et Sitton, 1998, Hudson et al., 1997, Heinke et Humphreys, 2005] proposent également qu'une région contiguë de l'espace est sélectionnée et aiguillée vers un système de reconnaissance afin de réaliser des systèmes de reconnaissance invariante. Cette idée qu'une partie de l'information disponible est sélectionnée et transmise aux niveaux de traitement supérieurs se retrouve également dans le *Feature Gate Model* [Cave et Bichot, 1999] et dans le *Selective Tuning Model* [Tsotsos, 1995] que nous détaillerons dans le paragraphe 1.4.2. Ces modèles reposent sur une structure hiérarchique de niveaux de traitement avec un mécanisme de sélection qui détermine si l'information est autorisée à passer au niveau de traitement supérieur. Ils mettent ainsi l'accent sur la compétition à un niveau de traitement, en vue d'accéder au niveau de traitement suivant.

On peut distinguer une troisième catégorie de modèles qui reposent sur des hypothèses biologiques plus fortes et visent à reproduire des données expérimentales. En particulier, [Lanyon et Denham, 2004] propose une séparation entre une voie spécialisée dans la reconnaissance des stimuli et une voie spécialisée dans le traitement des informations spatiales. Leur modèle reproduit des résultats expérimentaux sur l'influence de l'attention dans V4 et IT. [Deco et al., 2005] implémente les principes de la "compétition biaisée" de R. Desimone et J. Duncan, présentée dans le paragraphe 1.2.2. Enfin, dans [Hamker, 2004, Hamker, 2005c], l'auteur propose un modèle qui repose sur une séparation entre la voie ventrale et la voie dorsale comme les modèles

⁷Le code source ainsi que des vidéos illustrant le fonctionnement du système sont disponibles sur le site de l'équipe de Laurent Itti <http://ilab.usc.edu/bu/>

précédents mais exploite également la proposition de la “théorie prémotrice de l’attention” de G. Rizzolatti selon laquelle le déploiement de l’attention spatiale et l’exécution d’un mouvement oculaire sont intimement liés. Selon l’hypothèse de la réentrée (*reentry hypothesis*) que l’auteur propose, des connaissances a priori (spatiales ou non spatiales) influencent les traitements visuels de bas niveau pour biaiser les traitements suivants en faveur des stimuli pertinents. Nous détaillerons cette dernière proposition dans le paragraphe 1.4.3.

1.4.1 Carte de saillance : attention bottom-up, L. Itti

Le modèle proposé par [Koch et Ullman, 1985] (figure 1.11haut) repose sur les principes de la *Feature Integration Theory*(FIT) [Treisman et Gelade, 1980]. Différents attributs sont extraits en parallèle et représentés dans des cartes spatiales. Ces cartes spatiales convergent sur une carte de saillance qui constitue une représentation unifiée de la pertinence comportementale de chaque position de l’espace visuel. Un algorithme *winner-take-all* est ensuite appliqué pour déterminer sur quelle position spatiale l’attention visuelle est déployée. Cette architecture constitue la base du modèle développé par [Itti et al., 1998, Itti et Koch, 2001] (figure 1.11bas). Différents canaux (mouvement, orientation, luminance, couleur) sont extraits à différentes échelles puis, au sein d’un même canal, les différents niveaux sont combinés pour construire des cartes de saillance intermédiaires qui sont ensuite elles-mêmes combinées pour représenter la carte de saillance principale. La combinaison de ces cartes intermédiaires peut être pondérée par un biais *top-down* définissant des attributs cibles. La région la plus saillante, déterminée par un algorithme *winner-take-all*, constitue la région sur laquelle l’attention est déployée. Une carte spatiale mémorise les positions sur lesquelles l’attention a été déployée et inhibe la carte de saillance afin que l’attention se déploie par ordre de saillance décroissant.

Ce modèle intègre principalement l’attention bottom-up et définit la saillance comme un contraste local dans différents canaux. Différentes études montrent que la saillance ainsi définie est corrélée aux mouvements oculaires réalisés par des sujets humains observant la même scène [Parkhurst et al., 2002].

Ce modèle a connu plusieurs extensions [Walther et al., 2002, Navalpakkam et Itti, 2005]. Dans [Walther et al., 2002], les auteurs lui adjoignent le modèle de reconnaissance hiérarchique développé par l’équipe de Tomaso Poggio [Riesenhuber et Poggio, 1999] : seules les informations de la région la plus saillante atteignent le système de reconnaissance. Dans [Navalpakkam et Itti, 2005], les auteurs adjoignent au modèle originel une mémoire à long terme représentant des connaissances a priori qui peuvent fournir une information complémentaire et ainsi biaiser la carte de saillance. Par exemple, si la scène est un bord de mer et que la tâche consiste à rechercher une personne, il est plus probable de la trouver marchant sur la plage que volant dans le ciel. Enfin, une version légèrement modifiée, mais qui repose sur les mêmes principes que [Itti et al., 1998], est présentée et amplement détaillée dans [Frintrop, 2005b].

L’une des principales hypothèses de ce modèle est qu’il existe une carte topographique représentant la pertinence comportementale de chaque position du champ visuel. Or, plusieurs

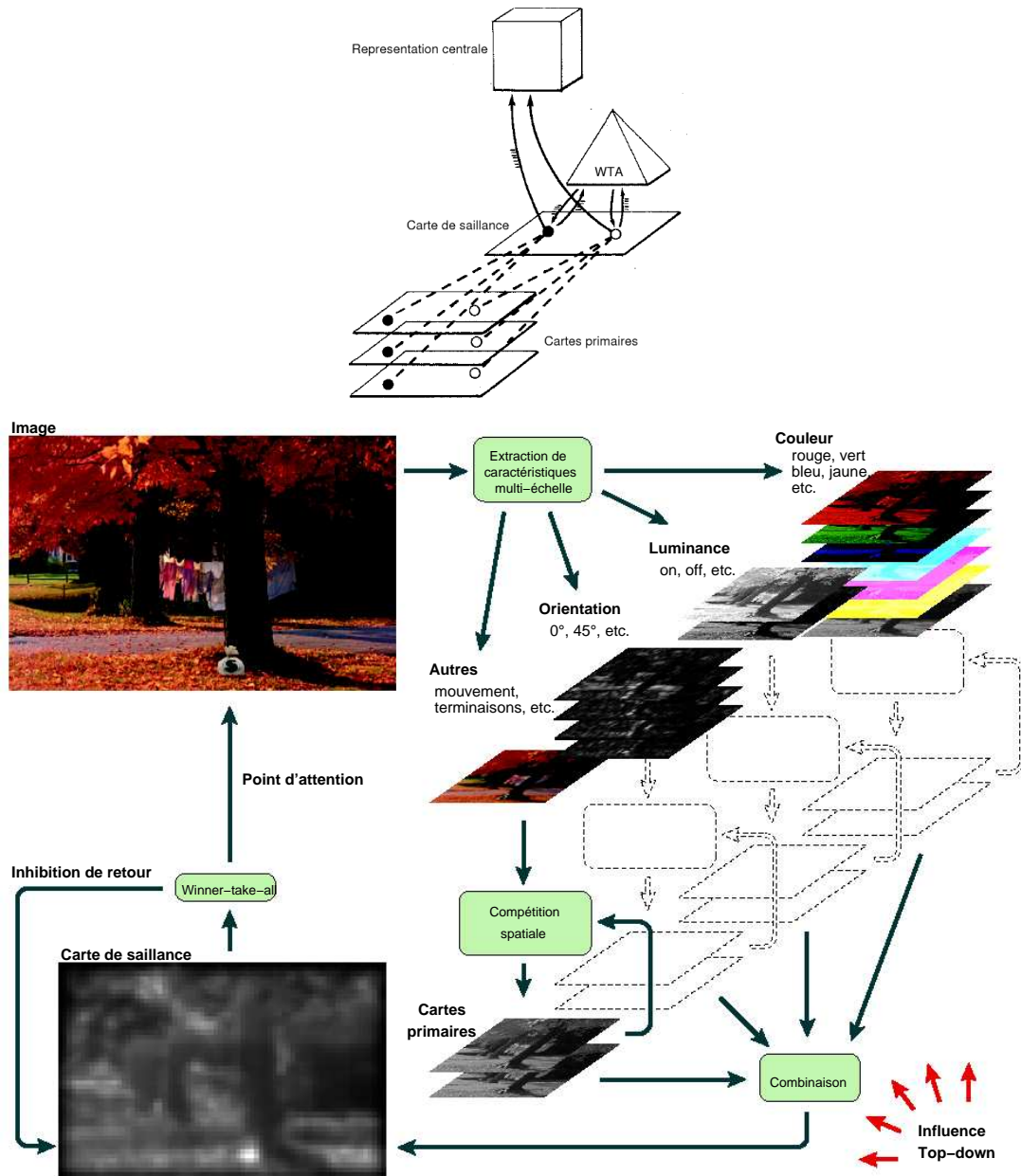


FIG. 1.11 – Haut : Schéma de principe proposé par [Koch et Ullman, 1985]. Des caractéristiques visuelles, extraites en parallèle et représentées dans les cartes primaires convergent sur la carte de saillance. Une compétition est engagée et la région la plus saillante capture l'attention. b) Bas : Modèle proposé par [Itti et Koch, 2001]. Plusieurs caractéristiques visuelles (couleur, orientation, luminance, mouvement) extraites en parallèle sont combinées pour former la carte de saillance. La combinaison de ces cartes primaires peut être biaisée par des connaissances a priori (influence *top-down*). Le mécanisme d'inhibition de retour mémorise les positions spatiales sur lesquelles l'attention a été déployée pour biaiser la compétition en faveur d'une région non sélectionnée précédemment.

études montrent que la saillance serait beaucoup plus distribuée : différentes aires exhibent des propriétés qui en font des candidats potentiels pour représenter la saillance. [Gottlieb et al., 1998] montrent en particulier que les neurones de l'aire LIP dans le pariétal postérieur ont une activité corrélée à la pertinence comportementale des stimuli, qu'ils apparaissent subitement ou qu'ils deviennent pertinents pour la tâche. [Thompson et Bichot, 2005] mènent une série d'études desquelles ils concluent qu'une population de neurones de FEF exhibe toutes les propriétés caractéristiques d'une carte de saillance, et que FEF, connu pour son implication dans la production d'une saccade oculaire, pourrait également être impliqué dans le déploiement de l'attention spatiale sans mouvement oculaire. C'est une idée dont s'inspire en particulier le modèle de F. Hamker que nous détaillerons dans le paragraphe 1.4.3. Enfin, dans les études de [Kustov et Robinson, 1996, Robinson et Petersen, 1992, Li, 2002], les auteurs considèrent que les colliculi supérieurs, le noyau pulvinar du thalamus ou encore l'aire visuelle primaire pourraient être des candidats pour une carte de saillance. La multiplicité des aires dont les neurones ont des activités dépendantes de la pertinence comportementale des stimuli conduit à une vision beaucoup plus distribuée de la saillance que la proposition d'une représentation unique. Néanmoins, ce système s'avère très efficace pour prédire quelles régions de l'espace visuel sont plus susceptibles d'être observées par un sujet.

1.4.2 Selective Tuning Model : compétition pour la représentation, J.K. Tsotsos

Le *Selective Tuning Model* est un modèle qui dérive d'une théorie mathématique développée par ses auteurs [Tsotsos, 1995] reposant sur une formalisation d'une tâche de recherche visuelle et sur la théorie de la complexité. D'après les auteurs, le système visuel des primates est une solution possible au problème général des systèmes visuels, mais n'est certainement pas la seule.

Une différence notable par rapport aux modèles à base de carte de saillance est qu'il n'y pas de représentation centralisée de la saillance. Au contraire, elle est déterminée de manière itérative le long d'une structure hiérarchique pyramidale (figure 1.12). L'information visuelle présentée au niveau le plus bas de la pyramide, la couche d'entrée, est propagée jusqu'au niveau le plus haut. Le long de la pyramide, les champs récepteurs étant de plus en plus grands, l'information visuelle d'une position spatiale se propage selon un cône divergent (non représenté sur la figure). Une compétition *winner-take-all* est engagée sur tout le champ visuel, dans la couche de plus haut niveau, afin de déterminer le gagnant global, c'est à dire l'unité dont l'activité est maximale. Cette compétition peut être biaisée par des connaissances a priori si elles sont disponibles, mais ce n'est pas nécessaire. Le vainqueur active une compétition entre ses entrées directes afin de déterminer la région la plus activée dans son champ récepteur. Toutes les projections n'ayant pas contribué à la victoire du gagnant global sont inhibées. L'inhibition de ces projections conduit à un raffinement itératif de l'activité du gagnant. Ce processus de sélection du gagnant est répété récursivement jusqu'à la couche d'entrée. Ainsi, à partir d'un gagnant global, la cause de sa victoire est localisée dans le champ visuel. Le chemin obtenu du gagnant global à la cause locale dans le champ visuel est appelée *pass zone* tandis que les projections inhibées récursivement forment un anneau inhibiteur. D'après les auteurs, en fonction de la tâche considérée (détection, localisation, reconnaissance), les deux passes (montante puis descendante) ne sont pas toujours nécessaires. Par exemple, pour détecter si un stimulus est présent dans la scène, la première passe montante peut suffire s'il est le gagnant de la compétition. En revanche, pour le localiser, il est

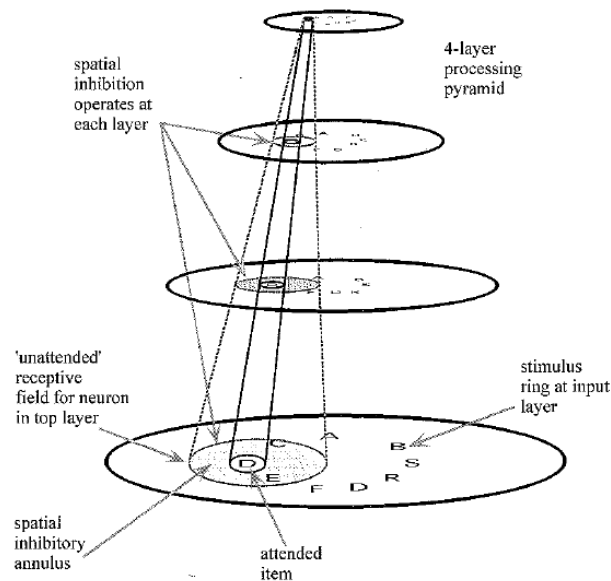


FIG. 1.12 – Schéma de principe du *Selective Tuning Model* [Tsotsos et al., 1995, Tsotsos, 1995]. L'information visuelle présentée au niveau le plus bas de la pyramide est propagée jusqu'au niveau le plus haut. A ce niveau, une compétition sélectionne la région la plus saillante. Le gagnant déclenche récursivement une compétition parmi ses entrées directes qui conduit à la sélection du stimulus le plus saillant dans son champ récepteur et à l'inhibition du voisinage immédiat.

nécessaire de propager le gagnant jusqu'à la couche sensorielle.

Le cadre théorique duquel dérive ce modèle conduit les auteurs à proposer un certain nombre de contraintes qu'un système de vision doit respecter, qu'il soit informatique ou biologique. Pour le système visuel des primates, certaines de ces contraintes constituent des prédictions, d'autres ont été vérifiées expérimentalement. Par exemple, les auteurs prédisent que lorsqu'une région est sélectionnée par l'attention spatiale, son voisinage immédiat est inhibé ; cette propriété est vérifiée expérimentalement par [Cutzu et Tsotsos, 2003]. Les auteurs proposent également qu'il n'y pas de représentation centralisée sur laquelle agit l'attention comme le suggèrent les modèles construits sur une carte de saillance. Comme nous l'avons vu dans le paragraphe précédent, différentes aires dans le cortex exhibent des propriétés attentionnelles qui conduisent les auteurs à les considérer comme des cartes de saillance. Les auteurs suggèrent également que la latence de l'influence de l'attention augmente en se rapprochant des aires visuelles inférieures. Plusieurs travaux, revus dans [Tsotsos et al., 2001], confirment cette prédiction.

La principale application de ce modèle concerne le traitement du mouvement [Tsotsos et al., 2005]. Certaines extensions sont proposées, comme la reconnaissance d'objets [Dolson, 1997] ou l'adjonction d'un système gérant le parcours d'une scène visuelle à l'aide de saccades oculaires [Zaharescu et al., 2004]. L'un des aspects essentiels de ce modèle est la modification itérative, tout au long de la hiérarchie, de la représentation des informations. Ces modifications de la représentation s'opèrent au travers de compétitions successives à tous les niveaux du système.

1.4.3 The reentry hypothesis, F. Hamker

Le modèle proposé par F. Hamker [Hamker, 2005a] repose sur des hypothèses biologiques plus affirmées que les modèles présentés dans les paragraphes précédents, et propose que l'attention est une propriété émergente plutôt qu'une fonctionnalité reposant sur un circuit ad hoc. Parmi ses principales caractéristiques, le modèle est constitué d'une distinction claire entre une voie spécialisée pour la reconnaissance des stimuli et une voie spécialisée dans le traitement des informations spatiales, qu'on peut assimiler à la voie ventrale et la voie dorsale que nous avons brièvement évoquées dans l'introduction et sur lesquelles nous reviendrons au chapitre 3. Il vise également à modéliser les influences de l'attention endogène et exogène. La modélisation de l'attention endogène est très similaire à [Itti et al., 1998] (voir le paragraphe 1.4.1) : des pyramides gaussiennes sont construites à partir de caractéristiques visuelles (couleur, orientation, intensité) extraites en parallèle. Différents niveaux de ces pyramides sont combinés afin de former des cartes de saillance topographiques pour chacune de ces caractéristiques (cartes V4 de la figure 1.13). Cette information alimente deux voies. La première (de V4 vers IT) combine l'observation a priori avec une connaissance sur la cible à rechercher, le résultat étant envoyé en retour sur les cartes V4 afin de favoriser les positions spatiales occupées par des stimuli qui partagent les attributs de la cible. La seconde voie (V4 vers PP ;FEF) traite l'information spatiale et sélectionne l'une des cibles potentielles. Cette information spatiale est envoyée en retour sur les cartes V4 pour favoriser dans cette représentation la position spatiale sélectionnée. Nous présentons le fonctionnement de ce modèle de manière séquentielle mais il est essentiel de comprendre que la sélection d'une cible peut se faire aussi bien sur la base d'une information non-spatiale que sur la base d'une information spatiale, via la représentation intermédiaire V4. Le retour de l'information spatiale sur V4, qui représente un biais en faveur d'une position spatiale sélectionnée et que l'auteur intitule *the reentry hypothesis* [Hamker, 2005c], est inspiré de la "théorie prémotrice de l'attention" de G. Rizzolatti. La symétrie entre la sélection par des informations spatiales et non spatiales permet par exemple d'extraire les attributs visuels d'une cible sélectionnée spatialement sans qu'un mécanisme ne soit dédié à cette fonctionnalité (comme par exemple dans le modèle de [Itti et al., 1998]).

La combinaison d'une observation (par exemple, l'information visuelle projetée sur V4 ou l'information envoyée de V4 à IT) et d'une connaissance a priori (par exemple, les attributs visuels d'une cible recherchée ou une position spatiale sélectionnée) permet de favoriser la représentation d'une information au dépend d'une autre. Ce principe est illustré sur la figure 1.14 : étant donnée une population codant l'information disponible dans son champ récepteur (figure 1.14a), une connaissance a priori biaise cette représentation (figure 1.14b) pour favoriser la représentation d'une partie de l'information. Ce principe de rehaussement du contraste d'une partie de l'information montante, grâce à une information a priori (fournie par exemple par une mémoire) n'est pas une idée nouvelle. En effet, dans les années 1980, Grossberg et Carpenter avaient déjà proposé ce principe dans leur modèle ART (*Adaptive Resonance Theory*), duquel ils ont dérivé par la suite différentes versions ART-2, ARTMAP, Fuzzy-ART, etc... [Grossberg, 1987, Carpenter et Grossberg, 1987, Carpenter et Grossberg, 2003]. Le modèle ART est appliqué à la classification de formes et son architecture est constituée de deux couches : une couche de comparaison et une couche de reconnaissance. La couche de comparaison est excitée par l'entrée visuelle ainsi que

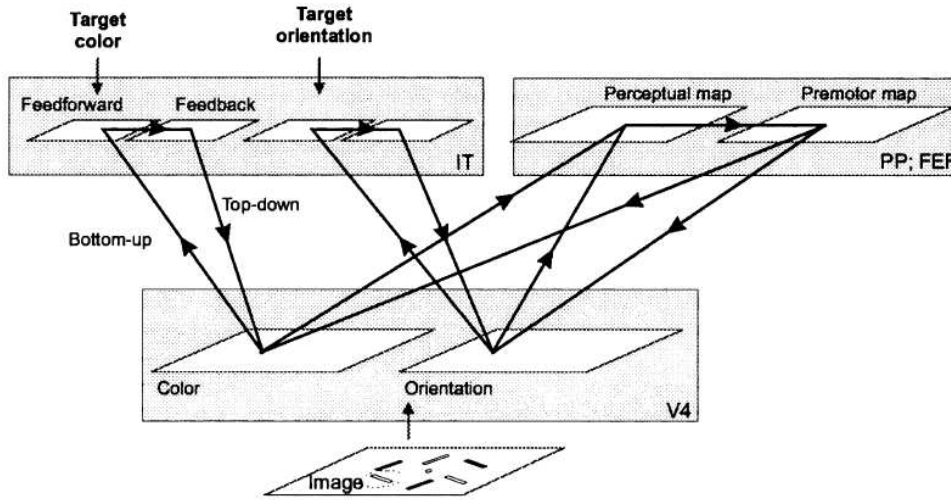


FIG. 1.13 – Schéma de principe du modèle proposé par [Hamker, 2004]. Différentes caractéristiques visuelles sont extraites en parallèle et représentées dans des cartes topographiques notées V4. Ces cartes alimentent deux voies de traitement, l'une dédiée au traitement des informations non spatiales (V4 vers IT), l'autre dédiée au traitement des informations spatiales (V4 vers PP ;FEF). La voie non spatiale est influencée par une connaissance a priori sur une cible à rechercher et vient en retour biaiser la représentation dans V4. La voie spatiale sélectionne une région et fournit à V4 un biais en faveur de cette position.

par la couche de reconnaissance. La couche de reconnaissance, quant à elle, est une mémoire des classes des formes, qui est excitée par la couche de comparaison. Un dernier élément, essentiel dans le modèle ART, est un terme de vigilance qui détermine dans quelle mesure une nouvelle classe d'objets doit être apprise. L'un des concepts introduits par les auteurs est que la couche de reconnaissance, dans laquelle la classe la plus proche de l'objet présenté est sélectionnée par compétition, module l'information montante qui atteint la couche de comparaison, permettant ainsi de raffiner l'information représentée dans la couche de comparaison ainsi que les prototypes de classes appris dans la couche de reconnaissance. On dispose ainsi d'un circuit récurrent qui permet de supprimer des informations non pertinentes pour favoriser la représentation des informations pertinentes.

Ce principe de modulation des entrées par une connaissance a priori est similaire au *contrast gain model* proposé par J. Reynolds pour modéliser les effets de l'attention spatiale sur l'activité de neurones de V4 (paragraphe 1.2.1). Afin d'illustrer comment l'attention module l'activité d'un neurone de V4 dans ce modèle, considérons l'équation d'évolution des neurones V4 proposée par [Hamker, 2005b]. Pour chaque position spatiale \mathbf{x} , pour chaque dimension \mathbf{d} (couleur, orientation) et pour chaque valeur i dans cette dimension (une couleur particulière, une orientation particulière), l'activité $r_{d,i,x}^{V4}(t)$ du neurone V4 associé répond à l'équation différentielle 1.1.

$$\tau \cdot \frac{d}{dt} r_{d,i,x}^{V4}(t) = I_{d,i,x}^{\uparrow} + I_{d,i,x}^N + I_{d,i,x}^A - I_{d,i,x}^{inh} \quad (1.1)$$

Dans cette équation différentielle, $I_{d,i,x}^{\uparrow}$ représente l'information d'entrée, dans le champ récepteur de la cellule. Nous nous limiterons dans ce paragraphe à la description du terme $I_{d,i,x}^A$ qui

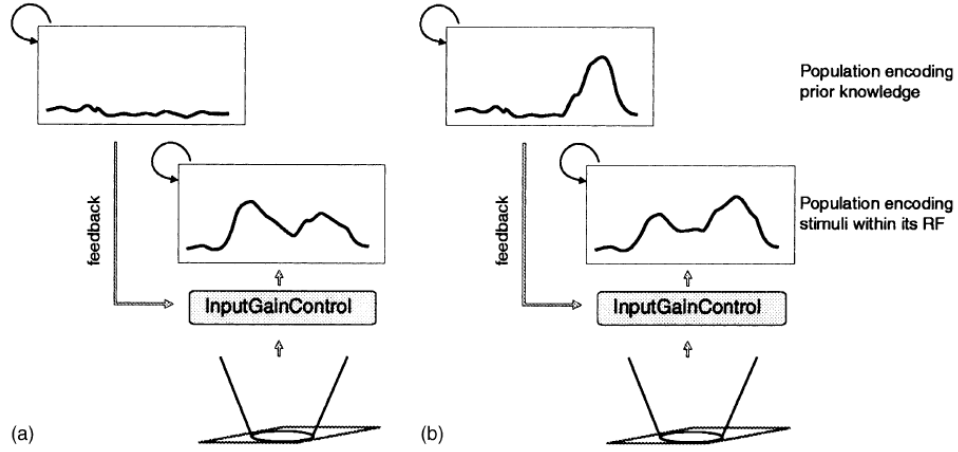


FIG. 1.14 – Illustration de l’influence d’un gain multiplicatif représentant des connaissances a priori. a) Lorsqu’une entrée, ici constituée de deux stimuli, est présentée à une population sans biais représentant des connaissances a priori, la représentation de la population est uniquement guidée par l’information d’entrée. b) Lorsqu’une connaissance a priori est disponible (représentée ici par une bulle d’activité dans la population la plus en haut), elle est utilisée pour biaiser la manière dont la population intègre ses entrées en agissant comme un gain multiplicatif. Adapté de [Hamker, 2004].

représente l’influence de l’attention spatiale et non spatiale, les termes $I_{d,i,x}^N$ et $-I_{d,i,x}^{inh}$ servant à introduire de la compétition entre les cellules et à normaliser leur réponse. Le terme “attentionnel” proposé par l’auteur est constitué d’un membre représentant l’attention non spatiale, guidée par les attributs, et d’un membre représentant l’attention spatiale (équation 1.2).

$$\begin{aligned}
 I_{d,i,x}^A &= f(I_{d,i,x}^\uparrow \cdot \sigma(\alpha - r_{d,i,x}^{V4}) \cdot \max_j w_{i,j}^{ITt,V4} \cdot r_{d,j}^{IT}) \\
 &+ f(I_{d,i,x}^\uparrow \cdot \sigma(\alpha - r_{d,i,x}^{V4}) \cdot w^{FEFm,V4} \cdot r_x^{FEFm}) \\
 f() &= \max_{d,i,x} \\
 \sigma(a) &= \max(a, 0)
 \end{aligned} \tag{1.2}$$

Nous ne décrivons que les idées sous-jacentes à cette équation, sans trop entrer dans les détails. La première idée est que l’attention sur les attributs visuels et l’attention spatiale sont combinées de manière additive. La deuxième idée est que l’attention module l’activité entrante de la cellule (ceci est représenté par la présence du terme $I_{d,i,x}^\uparrow$). La troisième idée est que l’influence de l’attention dépend de l’activité de la cellule : plus la cellule est active, moins l’attention a d’effet (ceci est représenté par le terme $\sigma(\alpha - r_{d,i,x}^{V4})$ où α est un terme à définir et représente l’activité à partir de laquelle la modulation *top-down* n’a plus d’influence). Enfin, l’attention sur les attributs dépend de l’activité des cellules de IT, qui appartiennent à la voie ventrale, et qui sont modulées par la connaissance a priori d’une cible recherchée. L’attention spatiale dépend quant à elle de l’activité des cellules de FEFm, qui appartiennent à la voie dorsale, et qui représentent une cible sélectionnée spatialement (cette population est appelée *premotor* sur la figure 1.13).

Étant donnée cette équation, l’auteur reproduit dans [Hamker, 2005b] les résultats concernant :

- les effets de l'attention spatiale sur la réponse de neurones de V4 observé par [Reynolds et al., 2000] (paragraphe 1.2.1)
- l'effet global de l'attention sur les attributs observé par [Motter, 1994a, Motter, 1994b] (paragraphe 1.2.2)
- l'effet de compétition locale observé par [Desimone et Duncan, 1995, Chelazzi et al., 1998] (paragraphe 1.2.2)

Nous avons mis l'accent sur une population de cellules particulière mais elle s'intègre à un modèle plus global. On trouvera une description détaillée de ce modèle ainsi qu'une application à des tâches de recherche visuelle dans [Hamker, 2005a]. Ce modèle est très intéressant puisqu'il intègre les différentes modulations attentionnelles dont nous avons parlées dans la première partie de ce chapitre et également parce qu'il illustre un principe que nous présenterons plus en détail dans le prochain chapitre, à savoir qu'un phénomène complexe comme l'attention visuelle peut émerger de l'interaction de cellules élémentaires simples.

Conclusion

Dans ce chapitre, nous avons présenté l'attention visuelle en mettant l'accent sur son rôle dans la sélection de l'information visuelle pertinente. Nous avons vu qu'il existe différents facteurs qui modulent la pertinence d'une information visuelle. D'une part, les propriétés intrinsèques d'un stimulus dans son environnement peuvent favoriser son traitement. Par exemple, l'apparition soudaine d'un stimulus ou son contraste dans un environnement sont des propriétés indépendantes de l'observateur, qui peuvent attirer son attention. Cette composante *bottom-up* de l'attention est facilement observée dans des tâches de recherche visuelle pour lesquelles on mesure le temps de réaction du sujet. D'autre part, nous avons également vu, qu'en plus de ces facteurs exogènes, les buts internes peuvent biaiser la sélection attentionnelle. En particulier, une information spatiale ou une connaissance a priori des propriétés visuelles d'un stimulus peuvent guider l'attention. Des mesures électrophysiologiques permettent de souligner cette contribution. De manière intuitive, on comprend que le biais apporté par une connaissance a priori permet de rendre le déploiement de l'attention visuelle plus efficace, puisqu'il favorise le traitement de certaines informations visuelles (régions spatiales à examiner ou attributs visuels à rechercher) en fonction d'un objectif donné.

Après cette première partie relatant quelques uns des plus importants travaux expérimentaux sur l'attention visuelle, nous avons présenté trois modèles attentionnels. Ces différents modèles ne sont pas à considérer de manière exclusive, chacun apportant une contribution complémentaire intéressante. En particulier, les modèles à base de carte de saillance mettent l'accent sur la manière dont l'attention *bottom-up* peut être dirigée, en s'intéressant à la manière dont l'information visuelle peut être traitée pour favoriser le traitement des régions de l'espace visuel les plus contrastées. Néanmoins, différentes caractéristiques de ces modèles peuvent être débattues. D'une part, ces modèles disposent en général d'un système dédié qui permet d'extraire les attributs visuels d'une région sélectionnée spatialement qui nécessite de décoder explicitement la région spatiale sélectionnée par l'attention. D'autre part, la sélection spatiale s'opère grâce à un mécanisme Winner-Take-All qui peut être vu comme un observateur extérieur qui estime la région la plus activée dans la carte de saillance. Nous verrons dans les prochains chapitres com-

ment ce mécanisme de sélection peut résulter d'une compétition plus dynamique et distribuée. Enfin, l'hypothèse de l'unicité de la carte de saillance est à l'heure actuelle très controversée puisque différents travaux suggèrent que des régions du thalamus [Koch et Ullman, 1985, Robinson et Petersen, 1992], du cortex visuel [Zhaoping, 2002, Mazer et Gallant, 2003], du cortex pariétal [Gottlieb et al., 1998, Gottlieb, 2007] ou encore du cortex frontal [Thompson et Bichot, 2005] pourraient contenir une carte de saillance.

Le modèle de *Selective Tuning* met, quant à lui, l'accent sur la compétition pour la représentation. C'est une idée que nous avons rencontrée lorsque nous avons présenté la compétition biaisée. La nécessité de cette forme de compétition s'apprécie plus facilement quand on considère les représentations visuelles le long de l'axe occipito-temporal. Les cellules le long de cet axe ont des champs récepteurs qui couvrent une région de l'espace de plus en plus grande, de 0.1° pour V1 jusqu'à 25° dans le cortex inférotemporal. Si aucun mécanisme ne permettait de biaiser la manière dont les cellules intègrent leurs entrées, l'information traitée par les cellules dans IT serait ambiguë. Le *Selective Tuning Model* utilise à tous les niveaux un mécanisme de Winner-Take-All et comme nous l'avons mentionné pour les modèles à base de carte de saillance, qui utilisent également ce mécanisme, nous verrons dans le prochain chapitre comment un mécanisme de sélection, dont le résultat émergent est semblable à un Winner-Take-All, peut reposer sur une compétition dynamique et distribuée.

Enfin, nous avons présenté l'hypothèse de la réentrée (*reentry hypothesis*) proposée par F. Hamker. Selon cette hypothèse, l'attention *top-down*, qui représente une modulation par des connaissances a priori, peut être modélisée par des projections de structures de plus haut niveau (comme par exemple une mémoire d'attributs visuels pertinents) sur des structures de plus bas niveau (comme par exemple des aires visuelles). Cette architecture est d'autant plus intéressante que le phénomène d'attention visuelle est la conséquence de l'interaction des cellules qui constituent l'architecture. Le constituant de base du modèle est la cellule et la fonction intéressante est observée au niveau global, comme résultant de l'interaction de ces cellules élémentaires. Dans le chapitre 2, nous développons plus en détail ce phénomène d'émergence de fonctions complexes par l'interaction de cellules élémentaires simples. Puis nous étudions comment l'attention visuelle peut être considérée comme émergente dans une architecture que nous proposons.

Chapitre 2

Emergence de l'attention visuelle sur un substrat de calculs numériques et distribués

A material model is the representation of a complex system by a system which is assumed to have some properties similar to those selected for study in the original complex system.[...] It may enable the carrying out of experiments under more favorable conditions than would be available in the original system.

A. Rosenblueth, N. Wiener
The Role of Models in Science, 1945

Introduction

Nous visons deux objectifs dans cette thèse qui conduisent à deux lectures différentes de nos travaux. Ces deux objectifs s'inscrivent respectivement dans le domaine des neurosciences et dans le domaine de l'informatique. Pour les neurosciences, on cherche à construire un système, analogue au système nerveux, pour aider à sa compréhension. Pour ce faire, on retient un ensemble limité d'hypothèses sur le fonctionnement du système physique, desquels on extrait des principes fonctionnels qui permettent de guider le développement d'un modèle. En informatique, l'objectif est différent. Il consiste à explorer de manière empirique les possibilités offertes par un paradigme de calcul original.

Nous avons déjà évoqué en introduction que le cerveau est constitué de neurones. Ces quelques centaines de milliards de neurones communiquent principalement par l'émission d'un signal électrique, transporté par des fibres. Nous présenterons plus en détails dans la section 2.1 quelques

propriétés essentielles d'un neurone. Nos travaux s'inscrivent dans la thèse connexionniste selon laquelle les fonctions cognitives complexes, comme la mémoire, la navigation ou l'attention visuelle, sont le résultat de l'interaction de ces cellules élémentaires. Nous présenterons plus en détails la thèse connexionniste dans la section 2.2 en soulignant quatre principes fondamentaux que sont le calcul local, numérique, distribué et adaptatif. La présentation de ces quatre principes nous permettra d'explicitier ce que nous entendons par l'émergence de phénomènes complexes d'un réseau d'unités qui réalisent des calculs simples. Nous présenterons ensuite, dans la section 2.3, la théorie des champs neuronaux, de laquelle on dérive un formalisme pour définir un modèle par analogie au système nerveux et dans lequel on retrouve les quatre principes mentionnés précédemment.

Dans la dernière section 2.4, nous montrons comment une fonction, analogue à l'attention visuelle, peut émerger d'un substrat de calculs locaux, numériques et distribués. Nous ferons l'impasse sur l'adaptabilité de la structure, non pas que cette propriété ne nous paraisse pas essentielle, mais parce que nous pensons que la construction d'un modèle dans le formalisme des champs neuronaux est une première contribution ; l'intégration de l'adaptabilité de la structure du modèle est une étape supplémentaire. La définition du modèle repose sur les principes fonctionnels de l'attention visuelle présentés dans le chapitre précédent. Pour des raisons didactiques, la présentation du modèle se fera en plusieurs étapes, en présentant successivement les différents composants fonctionnels qui composent le système, à savoir :

- un mécanisme de sélection
- auquel est adjoint un mécanisme de mémoire spatiale
- auquel est adjoint un mécanisme pour traiter les attributs non spatiaux (couleur, orientation) de l'information visuelle

2.1 Une introduction aux neurones biologiques

Le cerveau est constitué d'environ 10^5 neurones par mm^3 de tissu cérébral. Ces cellules sont spécialisées pour la génération de signaux électriques, en réponse à des signaux électriques ou chimiques, qu'ils transmettent à d'autres cellules. Les principaux constituants d'un neurone sont représentés sur la figure 2.1a. On distingue notamment l'arbre dendritique grâce auquel le neurone reçoit les signaux envoyés par les autres neurones, qu'on appelle neurones afférents, ainsi que l'axone grâce auquel le neurone contacte les autres neurones, qu'on appelle neurones efférents. Le signal électrique pertinent pour le système nerveux est la différence de potentiel entre l'intérieur du neurone et le médium extracellulaire avoisinant [Dayan et Abbott, 2001]. Lorsque la cellule est au repos, son potentiel membranaire est de l'ordre de $-70mV$. La variation de ce potentiel membranaire dépend essentiellement de la modification des concentrations d'ions dans la cellule, qui sont contrôlées par des canaux ioniques, eux mêmes sous l'influence des entrées de la cellule. Si des ions chargés négativement entrent dans la cellule ou si des ions chargés positivement sortent de la cellule, le potentiel membranaire diminue. On parle alors d'hyperpolarisation. Au contraire, lorsque le potentiel membranaire augmente, on parle de dépolarisation. De petites variations du potentiel membranaire sont rapidement atténuées le long de l'axone. Lorsque le potentiel membranaire est suffisamment dépolarisé, il évolue brutalement (figure 2.1b). Cette croissance brutale du potentiel membranaire, d'une amplitude de l'ordre de $100 mV$ et d'une durée de l'ordre d'une milliseconde, est appelée potentiel d'action (*spike*). Ces potentiels d'action transitent le long des axones pour contacter les synapses des dendrites des neurones efférents. Au contraire des petites variations du potentiel membranaire, les spikes sont transmis efficacement le long de l'axone, en particulier grâce aux nœuds de Ranvier qui favorisent leur propagation. C'est principalement par l'émission de trains de spikes que les neurones communiquent au sein du réseau neuronal.

La figure 2.2a représente un train de spikes émis par un neurone du cortex inférotemporal d'un singe, pendant qu'il regarde une séquence vidéo. Les spikes ne sont pas émis à intervalle de temps régulier, mais le temps d'émission des spikes dépend de l'amplitude et de la nature des entrées que reçoit la cellule. En effet, certaines synapses, qui sont les points de contacts entre les axones des neurones afférents et l'arbre dendritique du neurone, sont dites excitatrices, d'autres inhibitrices. Leur influence respective sur l'évolution du potentiel membranaire n'est pas la même : les premières favorisent la dépolarisation du potentiel membranaire, les secondes favorisent son hyperpolarisation. Une manière commode pour représenter la sortie d'un neurone est de représenter l'évolution de sa fréquence de décharge au cours du temps. Il existe différentes méthodes pour évaluer la fréquence de décharge d'un neurone⁸ mais l'idée reste toujours de calculer le nombre de spikes émis pendant une certaine durée et de diviser ces deux quantités. La figure 2.2b illustre l'évolution de la fréquence de décharge du neurone dont l'enregistrement est représenté sur la figure 2.2a.

Que représentent ces trains de spikes ? Qu'encodent-ils ? Ne sont-ils que du bruit ou peut-on leur associer une sémantique ? Différents travaux d'électrophysiologie soulignent que les neurones ont tendance à décharger préférentiellement dans certaines conditions. Pour être plus explicite, prenons l'exemple d'un neurone du cortex moteur qu'on enregistre pendant qu'un singe exécute des mouvements du bras. [Georgopoulos et al., 1982] montre que certains neurones du cortex

⁸On trouvera différentes procédures pour approximer la fréquence de décharge d'un neurone dans [Dayan et Abbott, 2001]

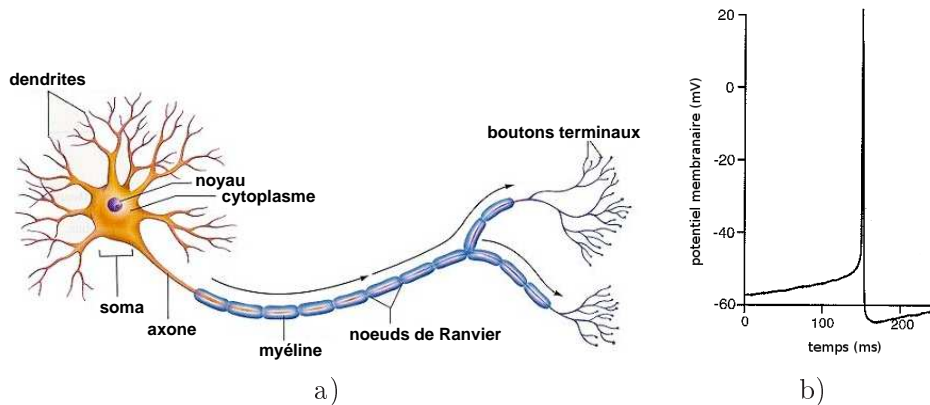


FIG. 2.1 – a) Un neurone intègre les potentiels d'action qu'il reçoit par ses dendrites, et transmet les potentiels d'action qu'il émet via son unique axone. La myéline qui entoure l'axone, ainsi que les noeuds de Ranvier, favorisent le transport des potentiels d'actions jusqu'aux boutons terminaux qui contactent l'arbre dendritique de neurones efférents par l'intermédiaire des contacts synaptiques. b) Représentation de l'évolution du potentiel membranaire, et de l'émission d'un potentiel d'action, ou *spike*. Au repos, le potentiel de membrane d'un neurone, défini comme la différence de potentiel entre l'intérieur de la cellule et le médium extracellulaire avoisinant, est de l'ordre de -70 mV. Sous l'influence des potentiels d'action des neurones afférents, le potentiel membranaire croît, jusqu'à atteindre une valeur critique. Dès lors, le potentiel membranaire croît rapidement, correspondant à une variation d'amplitude de l'ordre de 100 mV, pendant environ 1 ms. Cette évolution rapide du potentiel membranaire correspond à l'émission d'un spike, que l'axone propage efficacement.

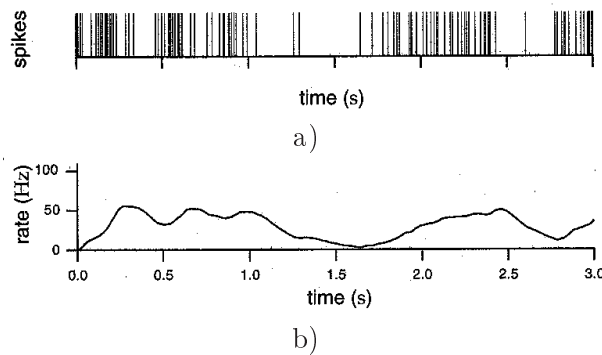


FIG. 2.2 – a) Enregistrement d'un train de spikes émit par un neurone du singe. Les spikes correspondent à des croissances rapides et brutales du potentiel membranaire du neurone, qui se propagent efficacement le long de l'axone. b) Une façon commode de représenter les trains de spikes est de calculer la fréquence de décharge d'un neurone, évaluée par le nombre de spikes émis par seconde. Adapté de [Dayan et Abbott, 2001]

moteur déchargent des spikes à une fréquence plus élevée lorsque le mouvement du bras est dirigé dans une direction qu'on appelle direction privilégiée. En répétant l'expérience pour différentes directions de mouvement, on peut alors tracer la fréquence de décharge de la cellule en fonction de la direction du mouvement. La figure 2.3a représente, pour chaque direction du mouvement du bras, l'activité de la cellule enregistrée pour cinq essais. La figure 2.3b est une illustration plus synthétique de ces résultats. Elle représente la fréquence de décharge de la cellule en fonction de la direction du mouvement du bras. On observe plusieurs choses :

- la fréquence de décharge de la cellule est maximale pour une direction qu'on appelle direction privilégiée ou direction préférée,
- la fréquence de décharge décroît lorsqu'on s'éloigne de la direction préférée. L'évolution de la fréquence de décharge peut être approximée par un cosinus centré sur la direction préférée.

En faisant varier la direction du mouvement, on obtient ainsi la courbe de réponse ou *tuning curve* de la cellule. Un deuxième exemple d'estimation de la courbe de réponse d'une cellule est illustré sur les figures 2.3c,d. Dans cet exemple, adapté de [Dayan et Abbott, 2001], on enregistre l'activité d'une cellule du cortex visuel du singe en fonction de l'orientation d'une barre présentée dans une région de l'espace visuel. Comme nous le verrons plus en détail dans le chapitre 3, les cellules du cortex visuel répondent de manière privilégiée à des stimuli présentés dans une région de l'espace visuel qu'on appelle champ récepteur. La figure 2.3d représente l'activité de la cellule en fonction de l'orientation de la barre présentée dans son champ récepteur. Comme pour le neurone du cortex moteur illustré précédemment, ce neurone du cortex visuel décharge maximale pour une orientation donnée, qui définit l'orientation privilégiée de la cellule, puis décroît lorsque l'orientation de la barre s'éloigne de la direction préférée du neurone.

De cette brève présentation des neurones biologiques, nous retiendrons que ce sont des cellules spécialisées dans la production d'un signal électrique, propagé dans un réseau formé des axones et des synapses. Nous avons également vu que certains neurones répondent préférentiellement pour certains attributs qu'on appelle attributs privilégiés ou attributs préférés. Le code neuronal, c'est à dire savoir associer une sémantique aux trains de spikes que les neurones s'envoient, est un problème encore ouvert. Nous avons néanmoins vu que pour certaines cellules (illustré ici dans le cortex moteur et le cortex visuel), on peut lier la décharge de la cellule et un attribut encodé (la direction d'un mouvement, l'orientation d'une barre). Néanmoins, la décharge d'une seule cellule est ambiguë et c'est seulement en considérant une population de cellules, dont les attributs préférés échantillonnent l'espace de l'attribut encodé, qu'on peut en avoir une bonne estimation. Ceci constitue un exemple de codage par population, sur lequel nous reviendrons dans la section 2.2.3. Dans la prochaine section, nous présentons quatre principes fondamentaux de la thèse connexionniste, dans laquelle s'inscrit nos travaux, que sont le calcul local, numérique, distribué et adaptatif. Ces quatre principes sont en relation étroite avec la biologie en considérant qu'un neurone est une cellule principalement indépendante, qu'il transmet un message sous la forme de trains de spikes aux cellules efférentes, qu'il fait partie d'un réseau grâce à ses projections afférentes et efférentes et que cette structure s'adapte avec l'expérience. Ce dernier point, l'adaptation de la structure, est évoqué dans la prochaine section pour des soucis d'exhaustivité. Néanmoins, il n'a été que peu abordé pendant nos travaux, non pas que nous considérons que ce ne soit pas une propriété essentielle, mais plutôt parce que nous la voyons comme une étape supplémentaire.

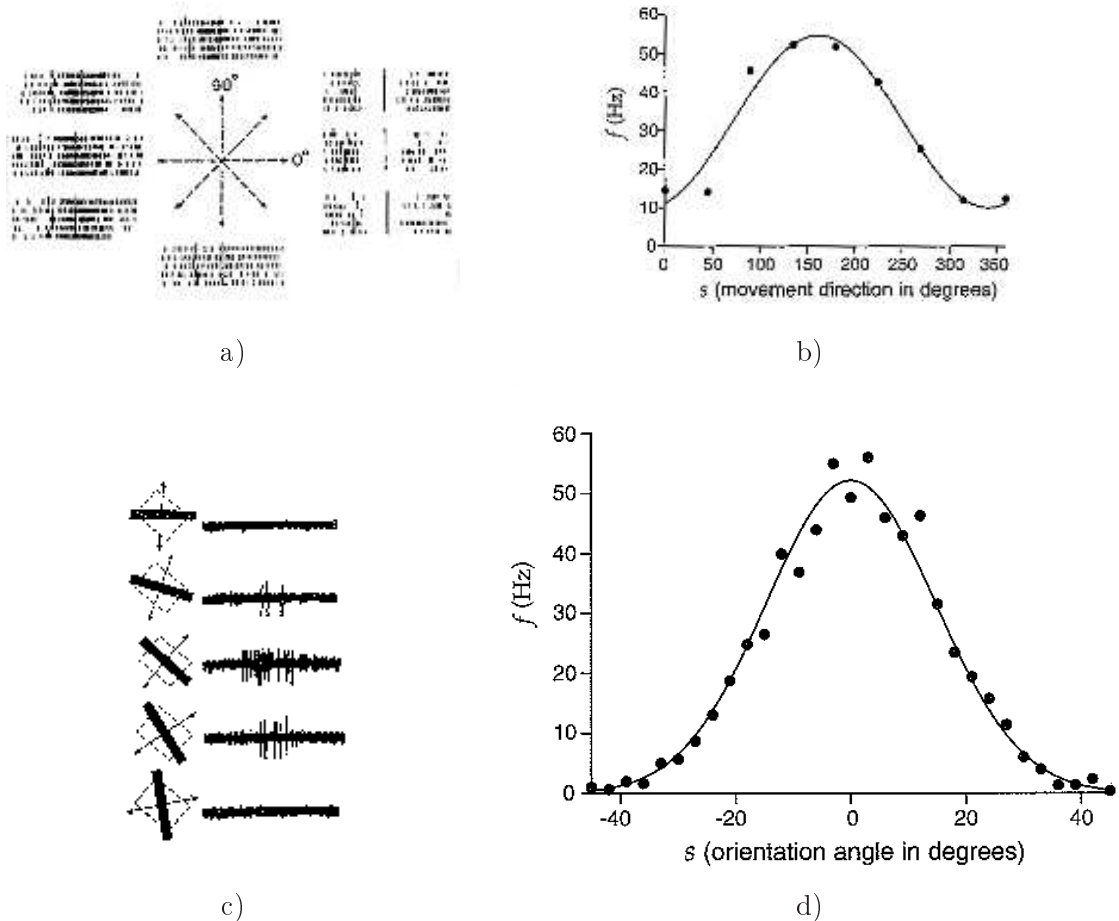


FIG. 2.3 – Courbe de réponse et sélectivité d'un neurone du cortex moteur et d'un neurone du cortex visuel. a) Enregistrements de l'activité d'un neurone du cortex moteur pendant qu'un singe exécute un mouvement du bras dans différentes directions. Les graphiques au bout de chaque flèche représentent les spikes enregistrés pour la cellule, en répétant cinq fois l'expérience. b) Quand on trace la fréquence de décharge de la cellule en fonction de la direction du mouvement du bras, on constate qu'elle est bien approximée par un cosinus centré sur une direction qui définit la direction privilégiée du neurone. c) Enregistrements de l'activité d'un neurone du cortex visuel primaire d'un singe, quand on présente une barre dans différentes orientations. d) Quand on trace l'activité de la cellule en fonction de la direction de la barre, on observe qu'elle est centrée sur une valeur qui définit la direction préférée de la cellule, et qu'elle décroît lorsqu'on s'éloigne de la direction préférée. a,b adaptées de [Georgopoulos et al., 1982] ; c,d adaptées de [Dayan et Abbott, 2001]

2.2 Le connexionnisme comme cadre de modélisation : Calculs locaux numériques distribués et adaptatifs

Comme nous l'avons évoqué dans l'introduction, nous visons deux objectifs. D'une part, proposer des modèles, en analogie avec le fonctionnement du système nerveux, pour aider à sa compréhension. D'autre part, nous cherchons à étudier les possibilités offertes par un paradigme de calcul original. Dans cette section, nous explicitons les quatre principes qui nous paraissent essentiels pour atteindre ces objectifs que sont :

- le calcul local et l'émergence
- le calcul numérique
- le calcul distribué
- l'adaptabilité du système

Selon la thèse connexionniste, les fonctions cognitives complexes, sont le résultat de l'interaction de cellules élémentaires, qui effectuent des calculs locaux simples. Pour bien comprendre les principes fondateurs du connexionnisme, il est intéressant de regarder d'un point de vue historique comment les différentes écoles de pensées se sont construites et se sont combattues. Nous proposons de restreindre ce bref aperçu historique aux idées apparues avec le mouvement cybernétique, au milieu du XX^{ième} siècle. La cybernétique est un mouvement dont le nom est devenu populaire grâce à l'ouvrage de Norbert Wiener publié en 1948 intitulé *Cybernetics or Control and Communication in the Animal and the Machine* dont le titre souligne l'ambition des cybernéticiens d'unir la pensée animale et artificielle. Entre 1942 et 1953 se déroulent les conférences MACY qui regroupaient des neurophysiologistes comme Arturo Rosenblueth, Warren McCulloch, des mathématiciens tels que John Von Neumann, Norbert Wiener, Walter Pitts, Alan Turing et même des ingénieurs, des anthropologues et des psychanalystes. Le point commun de tous ces acteurs était leur intérêt pour les systèmes auto-régulés et le concept de rétroaction qui est l'un des piliers de la cybernétique. Une contribution majeure du mouvement cybernétique est due à Warren McCulloch qui introduit avec Walter Pitts en 1943 une unité élémentaire de calcul qui porte leur nom [McCulloch et Pitts, 1943]. La vision de McCulloch et Pitts était que le cerveau réalise des opérations qui sont le fruit de l'interaction d'unités élémentaires dont le calcul peut se comprendre comme des règles appliquées sur des états logiques.

On retiendra également des conférences MACY l'interaction d'acteurs de différentes disciplines en vue d'étudier la pensée humaine, animale ou artificielle. Cette rencontre entre plusieurs disciplines en vue de comprendre les mécanismes de la pensée humaine, animale ou artificielle donne naissance aux sciences cognitives. Le cadre d'étude des sciences cognitives donne naissance à au moins deux écoles de pensées : le cognitivisme et le connexionnisme.

Selon l'approche cognitiviste, le cerveau est un système de traitement de l'information qui représente et manipule des symboles comme le ferait un ordinateur. L'un des principaux reproches faits au cognitivisme est qu'il présuppose l'existence de symboles sur lesquels le système agit, sans remettre en question la manière dont ces symboles sont construits et acquièrent du sens. Le cognitivisme a eu ses heures de gloire notamment au travers de l'intelligence artificielle "classique" ou "symbolique". En particulier, des systèmes experts, qui reposent sur la manipulation de bases de connaissances à l'aide d'un moteur d'inférence, ont été développés pour analyser

des composants chimiques (système *DENDRAL*, 1965) ou aider au diagnostic médical (système *MYCIN*, 1972). Malheureusement ces systèmes deviennent vite très lourds à mettre en œuvre en partie parce qu'il devient difficile d'assurer la cohérence du système lorsque le nombre de règles devient important. On retiendra trois aspects du cognitivisme :

- les symboles forment des représentations localisées ;
- les règles manipulant les symboles sont appliquées séquentiellement ;
- le système appliquant les règles est centralisé.

Le connexionnisme est une approche différente du cognitivisme. Certains considèrent que le connexionnisme est une approche sub-symbolique là où le cognitivisme est une approche symbolique ; chaque école travaillant à un niveau d'abstraction différent [Varela, 1996]. Néanmoins, il y a des différences fondamentales entre les principes cognitivistes et connexionnistes ; [Medler, 1998] distingue trois principes essentiels pour le connexionnisme :

- le cerveau est constitué d'unités élémentaires ;
- le cerveau a une architecture massivement parallèle et distribuée aussi bien pour les opérations qu'il réalise que pour les représentations qu'il construit ;
- les unités élémentaires sont reliées par des projections pondérées modifiables avec l'expérience.

Ces trois principes mettent en avant la distribution des calculs et des représentations ainsi que la modification de la structure du système par l'apprentissage. Les premiers succès du connexionnisme remontent aux travaux de Rosenblatt qui publia en 1958 un article introduisant le perceptron : un système constitué de deux couches d'unités, capable d'apprendre à associer des entrées et des sorties [Rosenblatt, 1958]. Mais des critiques sévères ont été adressées envers le perceptron, en particulier par Minsky et Papert [Minsky et Papert, 1969]. Ils mettent en avant les limites du perceptron concernant les fonctions qu'il est capable de calculer en démontrant qu'une fonction aussi simple que le ou-exclusif n'est pas calculable avec cette architecture. Cela conduit à un désintéressement du domaine avant que des résultats importants ne lui donne un second souffle, avec en particulier l'introduction des mémoires associatives [Hopfield, 1982], du perceptron multi-couches (MLP, *multilayer perceptron*) ainsi que d'un algorithme efficace pour faire apprendre une telle architecture (la rétropropagation du gradient de l'erreur, *backpropagation*) [Rumelhart et al., 1987].

Les prochains paragraphes sont consacrés à l'illustration des quatre aspects du paradigme de calcul que nous avons utilisé pour développer les mécanismes présentés dans cette thèse :

- le calcul local et l'émergence
- le calcul numérique
- les calculs et les représentations distribués
- le calcul adaptatif

2.2.1 Calcul local et émergence

En 1875, le psychologue George Henry Lewes introduit le terme d'émergence pour qualifier des systèmes et des processus incompréhensibles du point de vue mécanistique. Une propriété émergente est une propriété qui ne peut pas se comprendre en se reposant uniquement sur l'étude des éléments d'un système considérés isolément. C'est une propriété qui ne s'observe qu'au

niveau global du système et pas au niveau local.

Les automates cellulaires sont un exemple célèbre en informatique pour introduire le principe de calcul local et d'émergence. Les automates cellulaires ont été inventés par Stanislaw Ulam et John von Neumann dans les années 1950, en vue de définir un système artificiel capable de s'auto-reproduire. Un exemple très connu d'automate cellulaire est le jeu de la vie introduit par John Conway en 1970. Dans sa définition la plus générale, un automate cellulaire est un système composé de cellules qui se trouvent dans un état parmi un ensemble fini d'états possibles. L'état d'une cellule évolue au cours du temps grâce à l'application d'une règle locale faisant intervenir l'état des cellules dans un certain voisinage. Il existe différentes formes d'automates cellulaires en fonction des règles de mise à jour, du nombre de cellules du système, de la façon dont les états sont mis à jour (de manière simultanée ou non) et de la définition du voisinage qui intervient dans la définition des règles d'évolution. Puisque le propos de ce paragraphe est de souligner l'émergence d'un comportement global causé par l'interaction locale de cellules, définie par des règles simples, on se limite à présenter un exemple d'automate cellulaire à une dimension, deux états (0 ou 1), synchrone, avec un voisinage très localisé limité aux deux voisins les plus proches. Dans ce cas, il y a $2^{2^3} = 256$ règles d'évolution possible dont on trouvera une classification dans [Wolfram, 1984]. L'une de ces règles est illustrée sur la figure 2.4a. Imaginons que les états 0 et 1 soient représentés par les couleurs blanc et noir, l'illustration de la règle 30 en haut de la figure 2.4a (30 est la transcription décimale du nombre binaire 00011110 qui caractérise les états atteints par la cellule étant donné son état et l'état de son voisinage) n'est rien d'autre que l'ensemble des configurations possibles et les états obtenus par application de la règle. Par exemple, si toutes les cellules sont dans l'état 1 (toutes les cellules sont noires) ou si toutes les cellules sont dans l'état 0 (toutes les cellules sont blanches), la cellule au centre passe ou reste à l'état 0. Si la cellule est dans l'état 0 et que seule une des deux cellules voisines est dans l'état 1, la cellule passe dans l'état 1.

La figure 2.4a illustre l'évolution du système étant donné un état initial dans lequel seule une cellule est dans l'état 1 (noire), toutes les autres étant dans l'état 0. Le temps évolue de haut en bas, chaque case sur une ligne représentant une cellule. Si on se limite à la colonne centrale, c'est à dire aux différents états qu'occupent au cours du temps la cellule initialement dans l'état 1, il semblerait qu'il n'y ait aucune régularité, même si on simulait cet automate pour des milliers d'itérations (figure 2.4b). Son évolution est en fait chaotique et cette règle est utilisée par certains logiciels de calculs numériques dans leur générateur de nombres aléatoires [Wolfram, 2002], ce qui souligne la complexité du comportement émergent des interactions locales des cellules définies par une règle très simple.

Dans la communauté des réseaux de neurones artificiels, la cellule la plus simple et la première proposée historiquement est le neurone "formel" de McCulloch et Pitts [McCulloch et Pitts, 1943] (figure 2.5). Une unité a un état qui peut être 0 ou 1 (l'équivalent du caractère *tout ou rien* des spikes émis par les neurones) et des connexions à partir desquelles elle intègre ses entrées. Les auteurs distinguent deux types d'entrée : les entrées excitatrices et les entrées inhibitrices. Si un neurone afférent inhibiteur décharge, la cellule reste dans l'état 0. Si, au contraire, aucun neurone afférent inhibiteur ne décharge, l'état de la cellule est déterminé en comparant le nombre de cellules afférentes excitatrices à un seuil. Si ce nombre dépasse le seuil, le neurone passe dans

- ce signal est transmis aux cellules efférentes

On retrouve également des principes des automates cellulaires dans lesquels les unités définissent leur nouvel état à partir d'une règle de mise à jour dépendant de l'état des unités auxquelles l'unité est connectée. Néanmoins, une différence notable avec les automates cellulaires est que la "règle" d'évolution de l'état de l'unité dans un réseau de neurones artificiels peut être modifiée par des règles d'apprentissage, une propriété que nous abordons au paragraphe 2.2.4. Dans le prochain paragraphe, nous abordons d'un point de vue pratique, la manière dont on simule nos modèles.

2.2.2 Calcul numérique

Une différence fondamentale entre l'intelligence artificielle classique qui repose sur les principes cognitivistes et l'approche connexionniste est que la première repose sur la définition a priori de symboles tandis que ces symboles n'existent pas dans les réseaux connexionnistes. Ils peuvent éventuellement être vus comme une abstraction ou une interprétation des activités des réseaux. Tous les mécanismes que nous allons présenter dans ce manuscrit reposent sur la définition d'une activité qu'une cellule transmet aux cellules auxquelles elle est connectée. Dans le cas du neurone de McCulloch et Pitts, la mise à jour de l'activité dépend de la différence entre les activités reçues par la cellule et son seuil d'activation. Dans le cadre du paradigme de calcul des champs neuronaux que nous allons présenter au paragraphe 2.3, l'évolution de l'activité d'une cellule est régie par une équation différentielle du premier ordre qu'on peut résumer sous la forme de l'équation 2.1.

$$\begin{aligned} \forall i \in [1 : N], \frac{du_i}{dt}(t) &= f_i(\{u_j(t)\}_{j \in [1:N]}, t) + I_i(t) \\ u_i(0) &= u_{0,i} \end{aligned} \tag{2.1}$$

où $u_i(t)$ est l'activité de la cellule i au temps t , f_i une fonction qui définit la dynamique de la cellule et ses interactions avec les autres cellules et $I_i(t)$ une entrée extérieure au réseau. Comme nous le verrons un peu plus tard, les réseaux construits à partir d'unités dont l'activité est régie par ce type d'équation sont en général trop compliqués pour être étudiés analytiquement. Par ailleurs, l'étude analytique de ces systèmes est d'autant plus compliquée que le système est dans l'idéal immergé dans un environnement dynamique, dans lequel des événements inattendus peuvent survenir ; il est ainsi nécessaire de simuler numériquement ces systèmes. D'un point de vue purement technique, la simulation de cette équation différentielle requiert de la discrétiser temporellement. Pour ce faire, il existe différentes méthodes, les plus utilisées étant les méthodes de Runge-Kutta que nous allons brièvement exposer. Ces méthodes sont exposées pour expliciter comment un système, dont l'évolution est régie par une équation différentielle, peut être simulé. Dans cette thèse, néanmoins, nous nous sommes contraints à utiliser la méthode de Runge Kutta la plus simple, qui correspond à la méthode d'Euler, un choix que nous discuterons en conclusion de cette section.

Pour simplifier les explications, considérons l'équation différentielle 2.2

$$\begin{aligned} \frac{dy}{dt}(t) &= f(t, y(t)) \\ y(t_0) &= y_0 \end{aligned} \tag{2.2}$$

Les méthodes de Runge-Kutta [Press et al., 2007] reposent sur une estimation de la fonction $y(t)$ à des temps discrets. Pour des raisons de simplicité, nous noterons $y_0, y_1 \dots y_n$ les valeurs de $y(t)$ estimées aux instants $t_0, t_1, \dots t_n$, avec $t_{i+1} - t_i = h$ le pas de discrétisation. Il existe différentes méthodes de Runge-Kutta pour discrétiser cette équation :

- les méthodes explicites reposent uniquement sur des valeurs déjà calculées de y et un pas de discrétisation h constant
- les méthodes implicites définissent des équations pour lesquelles le terme à déterminer dépend de sa propre valeur (par exemple y_1 est défini comme une fonction de y_1)
- les méthodes adaptatives pour lesquelles le pas de discrétisation h est modifié pendant la simulation

A titre d'illustration, nous n'évoquons qu'une méthode explicite, la méthode de Runge-Kutta d'ordre 4, qui est définie par les équations 2.3 :

$$\begin{aligned}y_{n+1} &= y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\k_1 &= f(t_n, y_n) \\k_2 &= f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_1\right) \\k_3 &= f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_2\right) \\k_4 &= f(t_n + h, y_n + hk_3)\end{aligned}\tag{2.3}$$

Le principe est que la valeur de $y(t_{n+1})$ est calculée à partir de la valeur de $y(t_n)$ et d'une estimation de la pente de $y(t)$ multipliée par le pas de discrétisation. L'ordre de la méthode considérée caractérise l'erreur réalisée par l'approximation. La méthode présentée précédemment est la méthode de Runge-Kutta d'ordre 4, ce qui signifie que l'erreur commise à chaque pas de temps est de l'ordre de h^5 . Nous avons choisi de présenter cette méthode car c'est généralement celle intitulée "méthode de Runge-Kutta". La méthode d'Euler est la méthode Runge-Kutta d'ordre 1.

Tous les mécanismes présentés dans cette thèse ont été simulés en utilisant la méthode d'Euler bien qu'il soit assez facile de passer à une autre méthode d'évaluation. Les mécanismes présentés dans cette thèse reposent sur un paradigme de calcul proposé par la théorie des champs neuronaux, sur laquelle nous reviendrons dans la section 2.3. L'équation d'évolution du système, proposée dans le cadre de cette théorie, est à temps continu. Une approche pour simuler un système dont l'évolution est régie par une équation différentielle est de considérer le schéma de discrétisation pour lequel l'erreur commise par la simulation est la plus petite possible, de telle sorte qu'on simule le plus précisément possible l'équation différentielle. Néanmoins, plus la simulation du système à temps discret est précise vis à vis du système à temps continu, plus sa mise en oeuvre est coûteuse en ressources de calculs. Puisque l'un des objectifs de ces travaux est d'étudier les phénomènes émergents, nous avons utilisé le schéma d'évaluation le plus simple, à savoir la méthode d'Euler. En simulant nos systèmes avec cette méthode, nous commettons certes une erreur de simulation par rapport au système à temps continu, mais nous conservons néanmoins le principe de l'interaction de cellules qui effectuent un calcul simple. Et comme nous le verrons dans la suite de ce chapitre, nous conservons la possibilité d'observer des phénomènes émergents.

2.2.3 Calculs et représentations distribués

Il y a principalement deux notions liées à l'aspect distribué du paradigme dans lequel nous plaçons : les calculs distribués et les représentations distribuées. Nous allons détailler dans cette partie ces deux notions.

Calculs distribués : synchrone versus asynchrone

Les unités élémentaires interconnectées pour former les réseaux que nous allons présenter par la suite peuvent être considérées comme des unités de calcul élémentaires indépendantes par rapport au calcul qu'elles exécutent et au moment de l'exécuter. Nous avons déjà abordé dans le paragraphe précédent l'aspect numérique des opérations réalisées par les cellules. D'une certaine manière, chaque cellule "sait" quel calcul exécuter puisqu'il est défini par ses projections afférentes, sa fonction d'intégration et sa fonction d'activation qui lui sont propres. Elle est également indépendante par rapport au moment d'exécuter son calcul : il n'existe pas de superviseur indiquant aux cellules quand elles doivent être mises à jour. Cette notion de temps discret de mise à jour des cellules peut d'ailleurs être considérée comme un artifice principalement introduit car le temps est discrétisé pour évaluer numériquement les équations qui régissent l'évolution de l'état des cellules.

On peut considérer deux méthodes d'évaluation extrêmes : l'évaluation synchrone et l'évaluation asynchrone. Nous commençons par présenter brièvement ces deux schémas d'évaluation avant de justifier pourquoi il nous paraît plus intéressant d'utiliser une évaluation asynchrone.

La méthode classique pour implémenter une évaluation synchrone consiste, à chaque pas de temps, à évaluer toutes les unités et à stocker le résultat dans un *buffer*. Puis, un signal global indique à toutes les cellules que leur nouvelle activité est l'activité qu'elles ont calculée. De cette manière, l'état des cellules à l'instant $t + \Delta t$ est calculé uniquement à partir de l'état des autres cellules à l'instant t . Pour l'évaluation asynchrone, bien qu'il existe différents schémas pour l'utiliser [Cornforth et al., 2005], le principe repose sur l'évaluation de chaque unité élémentaire, considérée comme indépendante, et la diffusion immédiate du nouvel état calculé. Au contraire d'une évaluation synchrone, il n'est pas nécessaire de stocker le résultat des calculs dans un *buffer* intermédiaire ni de synchroniser les cellules à intervalle régulier.

On trouve deux principaux avantages à l'évaluation asynchrone. D'une part, elle est du point de vue du schéma d'évaluation plus réaliste pour modéliser le système nerveux constitué d'unités élémentaires (les neurones) largement indépendantes. D'autre part, l'évaluation asynchrone a l'avantage d'être intrinsèquement parallèle au contraire d'une évaluation synchrone qui, par définition, nécessite que les cellules se synchronisent à intervalle régulier. Ce parallélisme intrinsèque constitue un atout majeur d'un point de vue informatique pour la distribution des calculs.

Si on souhaite étudier un système dynamique régi par un ensemble d'équations différentielles,

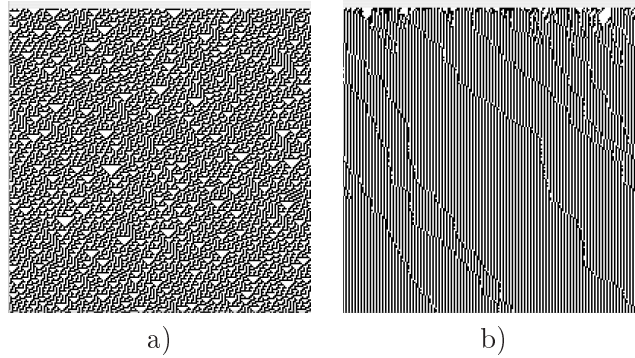


FIG. 2.6 – a) Évolution de la règle 30 en évaluation synchrone b) Lorsqu'on utilise un schéma d'évaluation asynchrone, le comportement est complètement différent.

l'introduction de l'asynchronisme complique grandement la tâche. Par exemple, dans le cadre des automates cellulaires, plusieurs théorèmes ont été prouvés dans le cas d'une évaluation synchrone mais il est montré que le comportement de ces systèmes peut être radicalement différent lorsqu'on utilise une évaluation asynchrone [Fates, 2008]. La figure 2.6 illustre par exemple l'évolution de la règle 30 dans le cas d'une évaluation synchrone et dans le cas d'une évaluation asynchrone. On trouvera également une autre illustration de l'intérêt de l'évaluation asynchrone dans [Vitay, 2006], en particulier concernant le mécanisme de sélection présenté au paragraphe 2.4.1 dont le fonctionnement dépend de l'introduction d'une forme d'asynchronisme. Ce mécanisme sert également d'illustration dans les travaux de [Rougier et Hutt, 2008], qui constitue une étude plus détaillée de l'influence d'une forme d'asynchronisme dans son évaluation.

Représentation localisée versus distribuée

Dans ce paragraphe, on se propose d'illustrer les avantages d'une représentation distribuée, comparée à une représentation localisée. Les critères de comparaison sont le pouvoir d'encodage de la population et sa robustesse à la perte de cellules. Il y a au moins trois paramètres critiques pour un codage par population : la taille de la population N , les attributs préférés des cellules x_i et leur sélectivité σ . Pour des raisons de simplicité, on considère que les attributs préférés des cellules sont uniformément répartis sur l'espace à encoder. A titre illustratif, imaginons que l'espace à encoder soit l'ensemble des réels $[1.0;2.0]$ (on pourrait imaginer un espace d'entrée plus abstrait comme les propriétés visuelles d'objets physiques), de telle sorte que les attributs préférés sont définis par $x_i = 1.0 + i\frac{1.0}{N}$ et fixons la taille de la population de manière arbitraire à 100 cellules. La sélectivité d'une cellule correspond au domaine de l'espace d'entrée à encoder sur lequel la cellule va répondre, relativement à son attribut préféré. Par exemple, considérons que l'activité A_i de la cellule i est définie par une gaussienne, centrée sur son attribut préféré x_i et de variance σ :

$$A_i = e^{-\frac{(x_i-x)^2}{\sigma^2}}$$

On peut voir le codage localisé comme un cas limite de l'encodage précédent, quand la variance tend vers 0, c'est à dire $A_i = \delta(x - x_i)$ où δ est le symbole de Kronecker ($\delta(x) = 1$ si $x = 0$, 0 sinon). Au contraire, on parlera de codage distribué quand la variance est non nulle. La

figure 2.7a représente la réponse de la population encodant la valeur $x = 1.5$ pour une variance proche de 0. La figure 2.7b représente la réponse de la population encodant la même valeur mais cette fois-ci avec une variance plus grande, égale à 0.1.

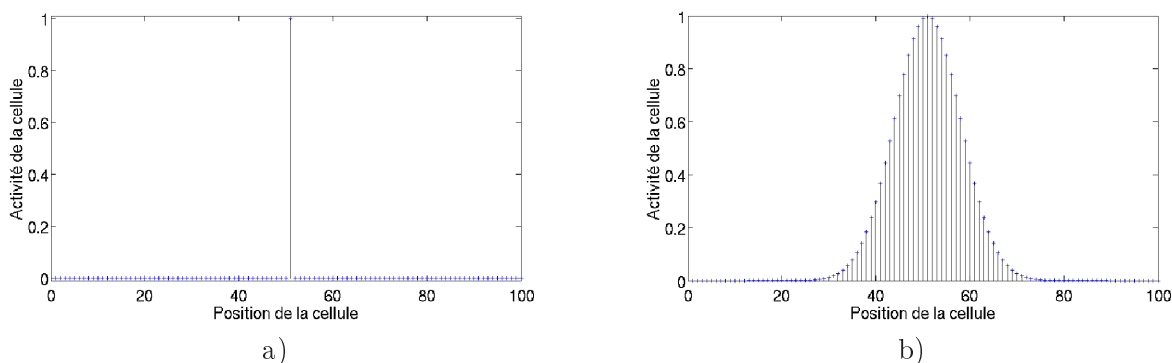


FIG. 2.7 – a) Réponse de la population localisée encodant la valeur 1.5 avec une très petite variance $\sigma = 10^{-11}$. Dans un codage localisé, au plus une cellule est active lorsque la population encode une valeur. Ici, seule la cellule 50 est active. b) Réponse de la population distribuée encodant la valeur 1.5 avec une variance plus grande $\sigma = 0.1$. La variance est définie dans l'espace d'entrée. Puisque l'espace encodé est $[1.0; 2.0]$, elle correspond avec une variance de $0.1N = 10$ cellules dans l'espace neuronal.

Nos deux critères d'évaluation des deux encodages sont le pouvoir de représentation et la robustesse à la perte de cellules. Le pouvoir de représentation peut être défini par l'erreur d'encodage du domaine à représenter. La robustesse à la perte de cellules peut être estimée par l'erreur d'encodage de la population, en fonction d'un taux de perte de cellules. Puisque dans un codage localisé, chaque cellule n'est sélective qu'à son attribut préféré, l'erreur d'encodage de la population sur le domaine n'est nulle que si la taille de la population tend vers l'infini de telle sorte que le pas de discrétisation tend vers 0. Au contraire, dans un codage distribué, puisque les cellules sont sélectives sur un domaine de l'espace d'entrée, elles auront une activité non nulle tant que la valeur à encoder est dans leur domaine de sélectivité. Il est ainsi possible de représenter l'espace continu $[1.0; 2.0]$ avec un nombre fini de cellules. Par ailleurs, le codage localisé est très sensible à la perte de cellules puisque la suppression d'une cellule entraîne l'incapacité à représenter la valeur correspondant à son attribut préféré. Au contraire, dans un codage distribué, et si la sélectivité est suffisamment grande, les domaines de sélectivité des cellules se recoupent de telle sorte que plusieurs cellules sont actives lorsque la population encode une valeur. Ainsi, dans un codage distribué, la suppression d'une cellule n'implique pas une augmentation de l'erreur de représentation aussi drastique que pour un codage localisé : les cellules se partagent la représentation de la valeur.

Pour résumer, pour un domaine à représenter de taille fixe, augmenter la taille de la population augmente son pouvoir de représentation puisque le pas de discrétisation de l'espace d'entrée diminue. Par ailleurs, si les cellules sont trop sélectives, d'une part certaines valeurs de l'espace d'entrée ne sont pas représentées et d'autre part la population est très sensible à la perte d'unités. En diminuant la sélectivité des cellules (elles répondent ainsi sur un domaine plus large

de l'espace d'entrée), on augmente le pouvoir de représentation de la population et on diminue la sensibilité à la perte de cellules puisque plusieurs cellules sont actives lorsque la population encode une valeur. Néanmoins, si la sélectivité est trop large (à la limite, les cellules répondent toutes de la même façon quelle que soit l'entrée), le pouvoir de représentation de la population chute. La sélectivité des cellules est ainsi un compromis entre le pouvoir de représentation de la population et la robustesse à la perte de cellules. La taille de la population ainsi que la répartition des attributs sont également des paramètres critiques d'un codage par population mais nous n'avons illustré ici que la sélectivité des cellules.

Dans l'étude du système nerveux, les premiers résultats concernant le codage par population remontent à [Georgopoulos et al., 1982]. Dans cette étude, les auteurs enregistrent l'activité de neurones du cortex moteur d'un singe, pendant que celui-ci exécute des mouvements du bras dans différentes directions (figure 2.3a,b). En traçant la réponse des cellules en fonction de la direction du mouvement, on constate deux choses :

- les cellules déchargent maximale pour un mouvement dans une direction donnée, qu'on appelle direction préférée,
- les cellules déchargent également lorsque le mouvement n'est pas dans la direction préférée ; l'amplitude de la réponse dépend de l'écart angulaire entre la direction préférée et la direction du mouvement par une fonction en cosinus.

Ces résultats montrent que dans le cortex moteur, le codage de la direction d'un mouvement est un codage par population. On retrouve également un codage par population dans les couches profondes du colliculus supérieur [Lee et al., 1988]. Les neurones de ces couches sont impliqués dans le codage d'une cible d'un mouvement oculaire. Le codage par population semble être une propriété essentielle du cortex. On trouvera de plus amples informations sur le codage par population et ses intérêts dans [Salinas et Abbott, 1994, Pouget et al., 2000, Averbeck et al., 2006, vanHemmen et Schwartz, 2008].

2.2.4 Adaptatif

Différentes règles d'apprentissage ont été proposées pour les réseaux de neurones artificiels. Certaines dépendent de l'estimation d'une erreur quant à ce que devrait calculer le réseau ; cet apprentissage est alors appelé supervisé. D'autres règles d'apprentissages ne nécessitent pas l'estimation d'une erreur ; cet apprentissage est appelé non-supervisé et peut être utilisé pour extraire les régularités d'un espace d'entrée. Enfin, l'apprentissage par renforcement peut être considéré comme un intermédiaire entre l'apprentissage supervisé et l'apprentissage non-supervisé. Il dépend d'une récompense qui indique l'adéquation d'une action engagée par un agent dans un environnement. Cette récompense n'est pas aussi précise que l'erreur fournie dans le cas d'un apprentissage supervisé mais sert néanmoins d'indicateur pour guider le comportement de l'agent.

Apprentissage supervisé

Pour réaliser un apprentissage supervisé, on dispose en général d'un ensemble de données d'entrée, des résultats associés et d'une fonction de coût. L'apprentissage supervisé repose alors sur trois étapes :

- présenter une entrée et évaluer le réseau
- comparer la sortie obtenue et la sortie désirée pour estimer une erreur
- propager l’erreur dans le réseau pour mettre à jour la structure du réseau (e.g. les poids des projections ou le nombre de cellules)

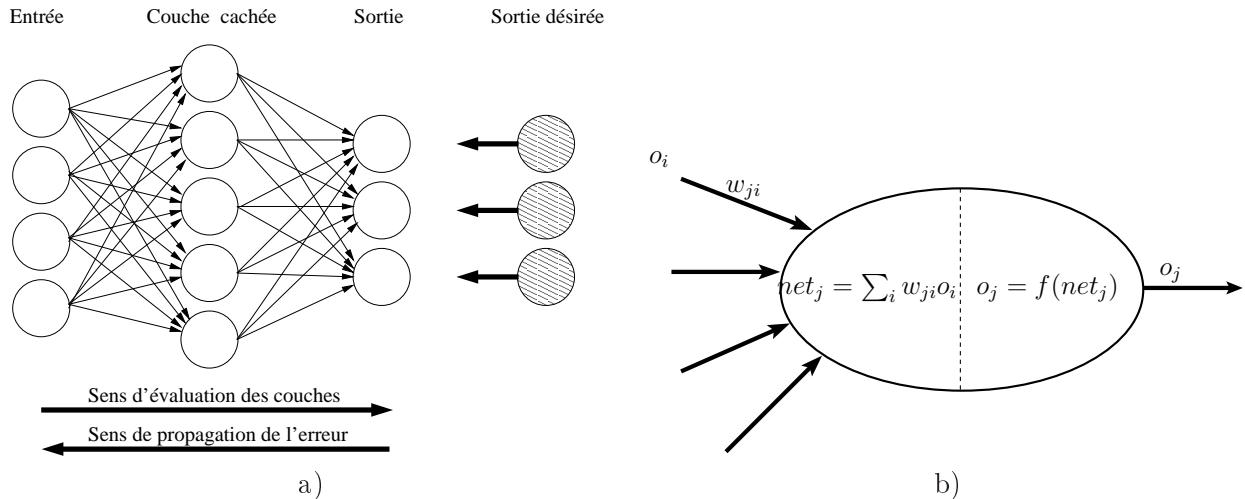


FIG. 2.8 – a) Un perceptron multi-couche (MLP) est défini par une couche d’entrée, une couche de sortie et une ou plusieurs couches cachées. Le réseau est *feedforward* (à distinguer des réseaux récurrents), c’est à dire que les projections vont toutes dans le même sens : de la couche d’entrée vers la couche de sortie. L’algorithme de rétropropagation du gradient de l’erreur nécessite l’évaluation d’une erreur, estimant la différence entre la sortie obtenue et la sortie désirée, qui est rétropropagée de la couche de sortie vers la couche d’entrée. b) Chaque unité du MLP intègre ses entrées par une somme pondérée. Le résultat de cette somme est transformé en appliquant une fonction de transfert dérivable (c’est un critère nécessaire pour pouvoir appliquer l’algorithme d’apprentissage), dont le résultat est transmis aux unités suivantes.

L’apprentissage supervisé repose sur la définition d’une erreur quant à ce que devrait calculer le réseau. Si le réseau est entraîné sur une fonction connue à l’avance, l’estimation de cette erreur est immédiate. On peut également imaginer des systèmes pour lesquels l’estimation d’une erreur est indirecte comme le serait l’apprentissage des mouvements d’un membre, guidés par la vision. L’algorithme le plus célèbre reposant sur l’estimation d’une erreur est l’algorithme de rétropropagation du gradient de l’erreur (*error backpropagation*). Cet algorithme d’apprentissage a été introduit pour doter les perceptrons multi-couches de facultés d’apprentissage afin d’obtenir une architecture dépassant les limites du perceptron [Rumelhart et al., 1987]. Cet algorithme d’apprentissage repose sur une loi de modification des poids synaptiques. Plus précisément, soit un perceptron multi-couches constitués de K couches et de $\{N_k\}_{k \in [1:K]}$ cellules pour chaque couche (figure 2.8a). L’apprentissage supervisé suppose qu’on dispose de P couples $\{I_p, T_p\}_{p \in [1:P]}$ entrée/sortie. Pour chaque entrée I_p (I_p est un vecteur de \mathbb{R}^{N_1} qui représente l’état de toutes les cellules de la couche d’entrée), on évalue l’état de sortie du réseau. Puisque le perceptron multi-couches est un réseau *feedforward*, l’évaluation se fait de la première à la dernière couche (figure 2.8a). On note $net_j = \sum_i w_{ji} o_i$ l’entrée du neurone j où w_{ji} est le poids entre le neurone pré-synaptique i et le neurone post-synaptique j (les notations utilisées sont résumées sur la figure 2.8b). Pour que l’algorithme puisse s’appliquer, la fonction d’activation f , telle que $o_j =$

$f(net_j)$ doit être dérivable (par exemple une sigmoïde). L'algorithme peut se résumer par les trois équations 2.4.

$$\begin{aligned}\Delta w_{ji} &= \eta \delta_j o_i \\ \delta_j &= (T_{p,j} - o_j) f'(net_j) \text{ pour un neurone de la couche de sortie} \\ \delta_j &= f'(net_j) \sum_k \delta_k w_{kj} \text{ pour les neurones des autres couches}\end{aligned}\tag{2.4}$$

C'est à dire que la mise à jour du poids w_{ji} entre le neurone pré-synaptique i et le neurone post-synaptique j dépend du taux d'apprentissage η , d'un erreur δ_j et de l'activité du neurone pré-synaptique o_i . La définition du terme d'erreur δ_j dépend de la couche à laquelle appartient le neurone post-synaptique. Si le neurone appartient à la couche de sortie, son erreur dépend de la différence entre l'activité désirée $T_{p,j}$ et l'activité obtenue o_j , ainsi que de la dérivée de la fonction d'activation. Si le neurone n'appartient pas à la couche de sortie, son terme d'erreur dépend de l'erreur de ses neurones efférents. On peut l'interpréter comme une estimation de la contribution du neurone à l'erreur globale. En résumé, cet algorithme repose essentiellement sur un terme d'erreur déterminé à partir d'une sortie désirée, ce terme d'erreur étant ensuite propagé de la couche de sortie vers la couche d'entrée.

Une deuxième illustration d'apprentissage supervisé concerne les réseaux RBF (*Radial Basis Functions*) [Moody et Darken, 1989, Ghosh et Nag, 2000] qui sont des réseaux à trois couches (une couche d'entrée, une couche cachée et une couche de sortie, comme l'architecture présentée sur la figure 2.8a). Pour être plus exact, l'apprentissage de ces réseaux passe par deux phases : une non-supervisée et une supervisée. L'intérêt de ces réseaux est que non seulement les poids entre les couches sont appris, mais en plus, les fonctions de transfert (qui sont à base radiale, comme par exemple une gaussienne) des neurones de la couche cachée sont également appris. La première phase d'apprentissage consiste à déterminer, à l'aide de méthodes appropriées⁹, les paramètres des fonctions de transfert des neurones de la couche cachée. Par exemple, si on choisit des fonctions de transfert gaussiennes, cela revient à déterminer le centre et la variance de ces gaussiennes. La deuxième phase de l'apprentissage est quant à elle très semblable à l'algorithme de rétropropagation et consiste à propager un terme d'erreur de la couche de sortie vers la couche d'entrée pour déterminer les poids du réseau.

A travers les deux exemples précédents, on constate que l'apprentissage supervisé repose toujours sur la définition explicite d'un terme d'erreur, utilisé pour mettre à jour la structure du réseau. Dans le prochain paragraphe, nous présentons une autre approche, l'apprentissage non-supervisé et qui n'a pas besoin de la définition explicite d'un terme d'erreur.

Apprentissage non-supervisé

Les algorithmes d'apprentissage non-supervisé sont utilisés lorsque les sorties désirées ne sont pas connues ou ne sont pas exprimables. On ne détaillera que deux exemples d'apprentissage non-supervisé : les règles dérivées de la règle de Hebb et les cartes auto-organisatrices.

⁹Les méthodes dites de quantification vectorielle ou de clustering en sont des exemples

La règle d'apprentissage de Hebb [Hebb, 1949] stipule que la mise à jour d'un poids connectant deux cellules dépend de la coactivation de la cellule pré-synaptique et de la cellule post-synaptique. Elle s'exprime simplement par l'équation 2.5.

$$\frac{dw_{ji}}{dt} = \gamma \nu_i \nu_j \quad (2.5)$$

où w_{ji} est le poids de la connexion entre le neurone i et le neurone j et ν_i, ν_j leurs activités respectives. Telle qu'elle est proposée, cette règle est divergente puisque si les deux unités pré- et post-synaptiques sont tout le temps actives en même temps, la règle implique que le poids croît indéfiniment. Il existe différentes variantes de la règle Hebb qui permettent de compenser cet inconvénient comme la règle anti-Hebb qui introduit un terme qui s'oppose à la croissance infinie du poids ou la règle de Oja qui introduit de la compétition entre les projections afférentes d'une cellule de telle sorte que la somme des poids de ces projections soit bornée.

[Gerstner et Kistler, 2002] proposent une généralisation élégante d'une règle d'apprentissage qui repose sur deux principes du postulat de Hebb : la *localité* et la *coopération*. Ils définissent de manière générale une règle d'apprentissage comme la définition d'une fonction F telle que :

$$\frac{dw_{ji}}{dt} = F(w_{ji}, \nu_i, \nu_j)$$

où w_{ji} est le poids de la connexion reliant le neurone pré-synaptique i au neurone post-synaptique j , ν_i, ν_j respectivement les activités des neurones pré- et post-synaptique. Si la fonction F est suffisamment dérivable, on peut alors considérer un développement de Taylor en $\nu_i = \nu_j = 0$:

$$\begin{aligned} \frac{dw_{ji}}{dt} &= c_0(w_{ji}) + c_1^{post}(w_{ji})\nu_i + c_1^{pre}(w_{ji})\nu_j \\ &+ c_2^{post}(w_{ji})\nu_i^2 + c_2^{pre}(w_{ji})\nu_j^2 + c_2^{corr}(w_{ji})\nu_i\nu_j + \mathcal{O}(\nu^3) \end{aligned} \quad (2.6)$$

Ce développement est réalisé jusqu'à l'ordre 2 mais on pourrait imaginer le poursuivre pour inclure des termes d'ordre supérieur. Une règle d'apprentissage est ainsi définie par la donnée des coefficients du développement limité, qui sont des fonctions qui dépendent du poids. En particulier, pour certaines valeurs bien choisies de ces paramètres, on retrouve des règles d'apprentissage classiques :

- la règle de Hebb est obtenue avec : $c_2^{corr}(w_{ji}) = \gamma$, tous les autres coefficients étant nuls
- la règle de Oja [Oja, 1982], qui introduit de la compétition entre les poids d'une même cellule est obtenue avec : $c_2^{post}(w_{ji}) = -\gamma w_{ji}, c_2^{corr} = \gamma$

Les cartes auto-organisatrices de Kohonen sont un deuxième exemple célèbre d'apprentissage non-supervisé [Kohonen, 1982, Kohonen, 1995]. Elles permettent d'extraire la structure d'un corpus de données sur lequel le réseau est entraîné. Dans sa définition classique, un réseau de Kohonen est constitué d'une couche d'entrée de M neurones et d'une couche de sortie de $N \times N$ neurones qui a une certaine topologie définie par une fonction de voisinage Λ . L'apprentissage consiste en deux étapes, une première qui consiste à déterminer le neurone le plus activé et une seconde qui est la phase d'apprentissage à proprement parler. Chaque neurone j de la couche de sortie est connecté aux neurones de la couche d'entrée par les poids $\vec{P}_j = \{w_{ji}\}_{i \in [1:M]}$. Ces poids définissent un *prototype* pour lequel la réponse du neurone est maximale. Lorsqu'on présente un exemple \vec{E} de l'espace d'entrée, chaque neurone évalue son activité par $o_j = f(\vec{P}_j \times \vec{E}) =$

$f(\sum_{i=1}^M w_{ji}E_i)$. Le neurone j^* d'activité maximale est appelé neurone gagnant. Dans sa définition standard, il est déterminé par un algorithme *winner-take-all* et correspond au neurone dont le prototype est le plus proche de l'exemple présenté. Le neurone gagnant j^* vérifie :

$$\|P_{j^*} - E\| = \min_j \|P_j - E\|$$

La seconde phase consiste à modifier le prototype de la cellule gagnante ainsi que d'un certain voisinage selon la règle :

$$\forall j, \Delta \vec{P}_j = \eta \Lambda(j, j^*) (\vec{E} - \vec{P}_j)$$

où η est le taux d'apprentissage et $\Lambda(j, j^*)$ définit une fonction de voisinage qui peut, par exemple, être la distance Euclidienne entre les neurones j et j^* en supposant qu'une position spatiale est attribuée à chaque neurone en accord avec la topologie du réseau. Le principe de l'apprentissage consiste à rapprocher le prototype du neurone gagnant de l'exemple présenté, ainsi que les prototypes des voisins du gagnant, avec une amplitude dépendant de leur "distance" au neurone gagnant. Une fois l'apprentissage terminé, on constate que les neurones voisins topologiquement dans la carte ont des prototypes proches dans l'espace d'entrée. On appelle ce phénomène *auto-organisation* car le système s'organise sur l'espace d'entrée sans jamais spécifier ce que doit calculer chaque cellule.

Dans l'approche des cartes auto-organisatrices de Kohonen, la structure est un *a priori* choisi avant l'exécution de l'algorithme d'apprentissage. Comme nous l'avons vu dans le cas de l'apprentissage supervisé, il y a différents paramètres qui définissent la structure d'un réseau :

- le nombre de couches et le nombre d'unités par couche ;
- la connectivité entre les unités ;
- la fonction de transfert des unités.

L'algorithme de Kohonen présuppose un nombre d'unité et une topologie, ce qui peut fortement contraindre l'apprentissage, au même titre que la définition d'un nombre de couches et du nombre d'unités par couches d'un perceptron multi-couches est critique. D'autres algorithmes comme le *Growing Neural Gas* permettent de s'affranchir de cette contrainte [Fritzke, 1995]. Le principe de cet algorithme est de partir d'un réseau de petite dimension et d'ajouter des cellules si nécessaire. La figure 2.9 illustre l'état du réseau au fur et à mesure de l'apprentissage ainsi que la distribution des entrées sur lequel ce réseau est entraîné.

Ainsi, comme pour l'apprentissage supervisé, il existe des algorithmes non-supervisés qui permettent d'apprendre les différents paramètres qui définissent un réseau mais, au contraire de l'apprentissage supervisé, ces algorithmes ne nécessitent pas qu'un terme d'erreur soit explicitement fourni au réseau. L'apprentissage supervisé a le désavantage, d'un point de vue de sa plausibilité biologique, de nécessiter un terme d'erreur pour chaque cellule du réseau. Une solution intermédiaire entre l'apprentissage supervisé et non-supervisé est l'apprentissage par renforcement, que nous détaillons dans le prochain paragraphe.

Apprentissage par renforcement

L'apprentissage par renforcement peut être vu comme un intermédiaire entre l'apprentissage supervisé, qui nécessite qu'un terme d'erreur soit calculé pour chaque cellule de manière explicite,

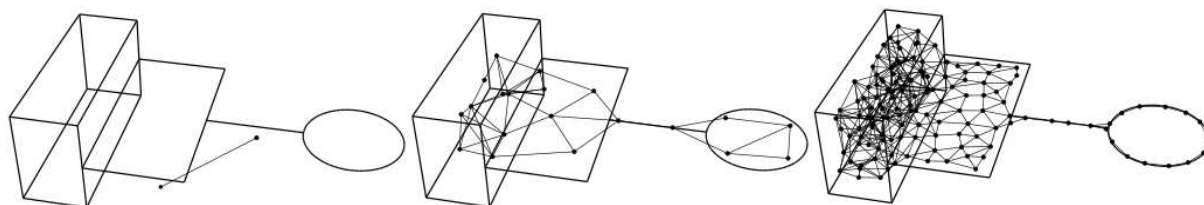


FIG. 2.9 – L’algorithme *Growing Neural Gas* part d’un réseau de petite taille qui croît au fur et à mesure de l’apprentissage, si l’ajout de nouvelles cellules est nécessaire, à comparer aux cartes de Kohonen dont la structure est fixée a priori. Adapté de [Fritzke, 1995]

et l’apprentissage non-supervisé qui extrait les régularités d’un espace d’entrée sans fournir de terme d’erreur. L’apprentissage par renforcement repose sur la donnée d’une “erreur” plus abstraite que pour l’apprentissage supervisé : un signal de récompense. Ce signal peut être une récompense négative ou une récompense positive, mais en aucun cas n’est défini pour chaque cellule.

[Barto et al., 1983] introduit l’apprentissage par renforcement comme l’apprentissage des actions qu’un agent doit réaliser dans un environnement pour maximiser une certaine récompense. Il est important de noter une différence fondamentale avec l’apprentissage supervisé. Alors que l’apprentissage supervisé enseigne au système quelle action réaliser, l’apprentissage par renforcement implique que le système apprenne de lui-même, par son expérience de ses interactions avec l’environnement, quelle action maximise sa récompense ; aucun signal n’indique de manière explicite la meilleure action à engager. Sutton et Barto étudient l’apprentissage par renforcement qu’ils considèrent comme un problème à la fois plus complexe que les apprentissages supervisés et non-supervisés mais également plus réaliste d’un point de vue biologique quant à la délivrance du signal de récompense. Dans leurs travaux ultérieurs, ils se rapprochent de la communauté du contrôle optimal qui leur permet de formaliser l’apprentissage par renforcement dans le cadre des Processus de Décision Markoviens (MDP), en considérant l’apprentissage par renforcement comme un problème plus contraint que le contrôle optimal, dans lequel la fonction de transition et la fonction de récompense ne sont pas connues par avance [Sutton et Barto, 1998]. L’apprentissage par renforcement, ainsi formalisée dans le cadre des MDP, est une théorie computationnelle générique non spécifique aux réseaux de neurones artificiels. Récemment, [Scherrer, 2003] a montré qu’il est possible de repasser du formalisme des MDP à une formulation connexionniste de l’apprentissage par renforcement.

La seconde raison pour mentionner l’apprentissage par renforcement est que différentes régions du cerveau ont des cellules dont les activités sont liées à une récompense attendue ou obtenue. C’est notamment le cas des ganglions de la base par leurs interactions avec le système dopaminergique et dont nous reparlerons au chapitre 3.3.4 [Hikosaka, 2007]. C’est également le cas d’aires oculomotrices, qui, lorsqu’elles sont enregistrées pendant la réalisation de tâches impliquant des mouvements oculaires récompensés (dans le cas d’un singe, la récompense peut être une certaine quantité de jus d’orange), déchargent en relation avec la récompense espérée ou obtenue [Amador et al., 2000, Kobayashi et al., 2002].

2.2.5 Bilan du calcul local, numérique, distribué et adaptatif

Dans les paragraphes précédents, nous avons illustré les quatre principes d'un paradigme de calcul qui nous paraissent intéressants, à savoir :

- le calcul local et l'émergence d'un phénomène global à partir de l'interaction locale d'unités élémentaires ;
- le calcul numérique ;
- les calculs et les représentations distribués et massivement parallèles ;
- le calcul adaptatif par la modification de la structure du réseau par apprentissage.

La localité des calculs est un élément essentiel du système nerveux puisqu'elle n'implique pas l'existence d'un superviseur ou d'un système qui contrôle/supervise le système. Dans ce cadre, les fonctions cognitives telles que la mémoire ou l'attention visuelle ne seraient que le résultat émergent de l'interaction des cellules élémentaires du système nerveux que sont les neurones. L'interaction des neurones semble se faire principalement par l'envoi de potentiels d'actions. C'est ce message que s'envoient les neurones que nous modélisons par une quantité numérique. Certains modèles, comme les modèles de neurones à spikes ou le modèle de Hudgkin-Huxley, essaient d'être aussi proches que possible du neurone biologique. Nous prenons le parti de simplifier ce message en considérant des neurones à taux de décharge moyen. En faisant ce choix, on pourrait nous reprocher de faire l'hypothèse que les temps d'émission des spikes et la synchronisation entre les neurones ne sont pas importants. Il serait plus juste de considérer que dans l'optique d'étudier le comportement de grandes populations de cellules, nous faisons l'hypothèse simplificatrice de se limiter à une grandeur qui pourrait s'assimiler au taux de décharge moyen. Nous ne prétendons pas que les temps de décharge ne sont pas importants pour le fonctionnement du système mais nous considérons qu'un modèle de neurone à taux de décharge moyen est un niveau d'abstraction suffisant pour observer des phénomènes intéressants avec des populations de telles cellules. De manière équivalente, la communauté qui utilise des neurones à spikes ne prétend pas que les canaux ioniques ne jouent pas un rôle dans le fonctionnement des neurones mais qu'en en faisant l'abstraction, on conserve des propriétés intéressantes des neurones. Ces modèles de neurone à différents niveaux d'abstraction sont des approches complémentaires. Le niveau d'abstraction choisi est un compromis entre ce que peut expliquer le modèle et les ressources nécessaires pour le simuler.

Nous avons également illustré l'intérêt des représentations distribuées par rapport aux représentations localisées en soulignant leur robustesse à la perte de cellules et leur pouvoir de représentation. Enfin, on dispose d'algorithmes efficaces pour que le système modifie sa structure, soit en modifiant les poids synaptiques, soit en modifiant le nombre ou la dynamique des neurones impliqués dans le calcul. C'est une propriété essentielle du système que de s'adapter à l'environnement dans lequel il se trouve.

Dans la prochaine section, nous présentons le paradigme de calcul des champs neuronaux qui s'inscrit dans le cadre des mécanismes de calcul local, numérique, distribué et adaptatif. Nous présentons l'intérêt des champs neuronaux grâce à quelques exemples illustratifs avant de proposer un mécanisme de déploiement de l'attention visuelle. La présentation de ce mécanisme est faite en plusieurs étapes en présentant ses différents composants fonctionnels.

2.3 Les champs neuronaux

Dans les années 1930, Nicolas Rashevsky, dont l'un des élèves fut Walter Pitts, entreprend de développer une biophysique mathématique en vue d'introduire des modèles mathématiques dans l'étude du système nerveux. Le principe est alors de décrire un phénomène macroscopique en définissant un système d'équations différentielles décrivant les interactions quantitatives entre les éléments constitutifs [Dupuy, 1999]. Cette démarche sera poursuivie, notamment par les travaux de [Beurle, 1956, Griffith, 1963, Griffith, 1965, Wilson et Cowan, 1972, Wilson et Cowan, 1973], en considérant une limite spatiale continue du tissu cortical. En s'intéressant à la proportion de cellules devenant actives par unité de temps dans un volume donné de neurones dans un réseau connecté aléatoirement, Beurle a été capable d'étudier le déclenchement et la propagation d'activités dans ce réseau. Il n'a étudié que des réseaux de neurones excitateurs et l'extension aux populations excitatrices et inhibitrices ainsi que l'introduction de la période réfractaire est due à [Wilson et Cowan, 1972, Wilson et Cowan, 1973]. Un peu plus tard, différents travaux [Amari, 1977, Taylor, 1999] ont identifié les conditions pour que des *patterns* d'activités apparaissent dans des champs neuronaux à une ou deux dimensions. Ils montrent en particulier l'existence de quatre comportements :

- des profils périodiques en temps ou en espace ;
- des régions d'activités localisées dans l'espace et stables au cours du temps ;
- des profils qui évoluent dynamiquement avec le temps comme des ondes qui se propagent dans le champ neuronal ;
- des solutions “triviales” : un réseau sans excitation ou au contraire complètement excité.

Dans cette section, nous commençons par illustrer quelques comportements des champs neuronaux à une dimension, avec une ou deux couches de cellules excitatrices ou inhibitrices, en s'inspirant des travaux de [Amari, 1977]. Nous avons pris le parti de ne présenter les champs neuronaux qu'au travers de quelques exemples illustratifs des comportements qu'on peut obtenir en formant des réseaux à partir de telles unités et en explicitant pourquoi les comportements obtenus peuvent être intéressants pour la modélisation du système nerveux.

2.3.1 Quelques exemples illustratifs

Dans [Amari, 1977], l'auteur s'intéresse à la dynamique d'un champ neuronal à une dimension spatiale définie par l'équation 2.7, dans laquelle la fonction de poids est homogène, isotrope et définie par une différence de gaussiennes, plus couramment appelée “chapeau mexicain”. L'évolution de l'activité $u(x, t)$ d'un neurone à la position x , au temps t dépend de son activité, de l'intégration de l'activité des neurones latéraux, de l'influence des neurones afférents et de la stimulation extérieure $I(x, t)$ et du taux de décharge spontanée h . La fonction de transfert des cellules est l'échelon de Heaviside.

$$\begin{aligned}
\frac{\partial u}{\partial t}(x, t) &= -u(x, t) + \int w(x - x')f[u(x', t)]dx' + I(x, t) + h \\
f(x) &= \begin{cases} 0 & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \\
w(x) &= A_+e^{-\frac{x^2}{\sigma_+^2}} - A_-e^{-\frac{x^2}{\sigma_-^2}}
\end{aligned} \tag{2.7}$$

La figure 2.10 représente l'allure des poids en différence de gaussiennes. La gaussienne de paramètres (A_+, σ_+) est dite excitatrice et la gaussienne de paramètres (A_-, σ_-) est dite inhibitrice. La gaussienne excitatrice a une amplitude plus grande et une variance plus petite que la gaussienne inhibitrice de telle sorte que la différence de ces deux gaussiennes conduit à une influence localement excitatrice qui devient inhibitrice sur de plus grandes distances puis tend vers zéro.

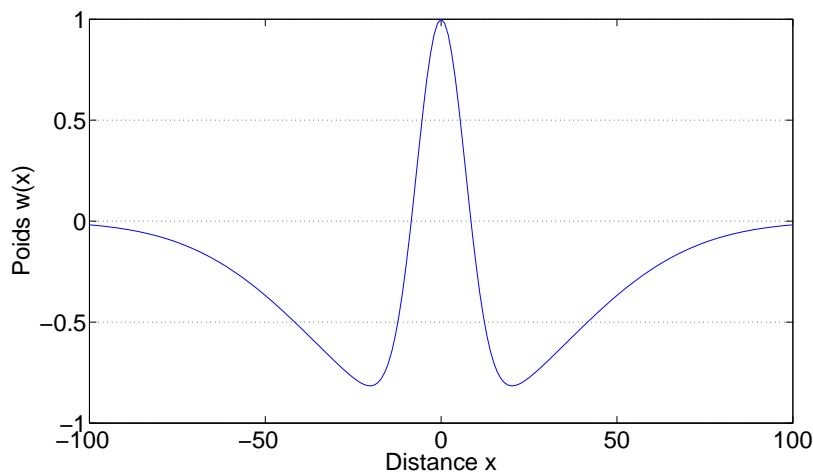


FIG. 2.10 – Le profil de poids en “chapeau mexicain” est la différence d’une gaussienne excitatrice étroite et d’une gaussienne inhibitrice large. Avec l’équation 2.7, la fonction de poids représentée est définie avec $A_+ = 2.0, \sigma_+ = 10.0, A_- = 1.0, \sigma_- = 50.0$

Dans les prochains paragraphes, nous considérons quelques exemples illustratifs des champs neuronaux. Tout d’abord, nous illustrons le cas d’une population à une dimension dont les connexions latérales sont définies par une différence de gaussiennes, et à laquelle on applique une stimulation extérieure constituée de deux excitations gaussiennes. Pour une même connectivité, le comportement de la population diffère en fonction de la distance entre les deux sites stimulés. Dans ce même paragraphe, nous présentons également le cas d’un réseau dont les connexions latérales contiennent une composante asymétrique. Puis nous présentons le cas de deux populations qui interagissent. Sous certaines conditions, ces populations peuvent exhiber, sans stimulation, une activité oscillante. Lorsqu’une stimulation est appliquée, on peut observer la propagation d’ondes d’activité. Une étude exhaustive de ces deux derniers phénomènes (oscillation et vague d’activité) est proposée dans [Amari, 1977].

Un réseau à une population de cellules : Formation de patterns statiques ou dynamiques

Dans le cas d'une seule population de cellules, en fonction de l'interaction latérale entre les cellules et de la stimulation appliquée au réseau, on peut observer différents comportements dont la coexistence, la sélection ou la fusion des stimulations. Considérons un réseau 1D de cellules dont les poids latéraux sont définis par une différence de gaussiennes avec les paramètres $A_+ = 0.5, \sigma_+ = 0.25, A_- = 0.46, \sigma_- = 0.4$. On stimule le réseau à deux positions spatiales. On observe alors, en fonction de la distance entre les sites stimulés, différents comportements, illustrés sur la figure 2.11 :

- lorsque les deux sites stimulés sont suffisamment distants, ils ne s'influencent pas de telle sorte que deux excitations demeurent dans le réseau (figure 2.11a),
- lorsque les deux sites stimulés sont plus proches, les deux régions excitées entrent en compétition de telle sorte que seule une des deux excitations demeure. Puisque le réseau est symétrique, le vainqueur peut être l'une ou l'autre des deux excitations (figures 2.11b,c),
- Lorsque les deux sites stimulés sont encore plus proches, il y a coopération : les deux excitations fusionnent (figure 2.11d).

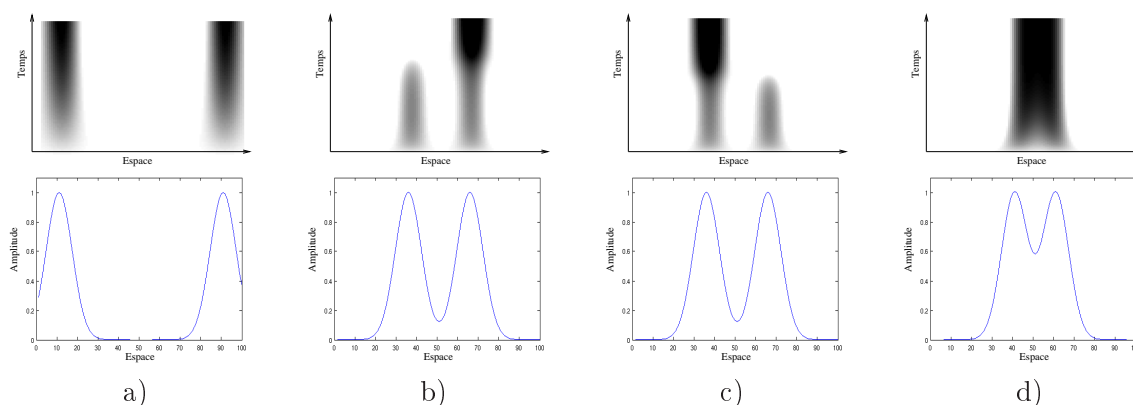


FIG. 2.11 – Différents comportements d'une population de 100 cellules, connectées latéralement par une différence de gaussiennes, et soumises à deux excitations dont on fait varier la distance. a) Lorsque les deux régions stimulées sont suffisamment éloignées, deux excitations coexistent. b,c) Lorsque les deux régions stimulées sont plus proches, seule l'une des deux excitations demeure. d) Si on rapproche encore les deux sites stimulés, les excitations fusionnent. Les paramètres de la différence de gaussiennes sont $A_+ = 0.5; \sigma_+ = 0.25; A_- = 0.46; \sigma_- = 0.4$.

Sous certaines conditions, on observe ainsi la formation de *patterns* d'activités, stables au cours du temps. Nous avons illustré le cas où les connexions latérales sont symétriques. On peut également considérer des connexions latérales asymétriques. Un exemple simple, proposé par [Zhang, 1996], consiste à définir les connexions latérales comme une différence de gaussiennes à laquelle on ajoute une composante qui est la dérivée de cette différence de gaussiennes. Puisque la différence de gaussiennes est paire, sa dérivée est impaire. En pondérant l'influence de cette composante impaire, on introduit une instabilité dans le réseau qui permet de déplacer continuellement des patterns d'activités, sa vitesse étant dépendante du poids attribué à la composante impaire [Zhang, 1996]. Un exemple d'une telle fonction de poids est présenté sur la figure 2.12a. La figure 2.12b illustre le comportement de la population pour une stimulation avec une variance assez large. Une excitation prend naissance au niveau du maximum de la stimulation puis, à cause

de l'instabilité causée par l'asymétrie des poids, l'excitation se propage vers la gauche du réseau. Puisque l'inhibition latérale décroît à partir d'une certaine distance, il arrive un moment où la région excitée s'est suffisamment déplacée pour qu'une deuxième région excitée puisse émerger et qui "chasse" la première excitation avec plus de force.

Ce mécanisme de déplacement dynamique d'un profil d'activité modulé par une information de vitesse est notamment utilisé dans le mécanisme proposé par [Dominey et Arbib, 1992]. Dans ce mécanisme, les auteurs mémorisent, dans un réseau de neurones, la position spatiale de cibles qui lui sont présentées. Les auteurs introduisent des projections latérales asymétriques afin de mettre à jour continûment cette mémoire pendant le déplacement des yeux. Nous reviendrons sur cette hypothèse de *continuous remapping* au chapitre 3, l'objet ici étant uniquement d'illustrer certains comportements de ces populations de cellules.

Les deux exemples illustratifs précédents permettent d'illustrer la formation de pattern statiques ou dynamiques dans une population de cellules, dont la fonction de voisinage est définie par une différence de gaussiennes (excitatrice dans le voisinage immédiat puis inhibitrice sur de plus longues distances), à laquelle on ajoute ou non une composante asymétrique. Le prochain paragraphe illustre l'apparition d'activité oscillante, sans stimulation extérieure (c'est à dire générée et entretenue par la dynamique de la population), ainsi que la diffusion d'activité par vague (*travelling waves*) dans un réseau construit à partir de deux populations de cellules.

Un réseau à deux populations de cellules

On considère deux populations de cellules qu'on appelle population excitatrice EXC et population inhibitrice INH. Les unités de la population EXC ont des projections latérales excitatrices ainsi que des projections inhibitrices afférentes de la population INH (figure 2.13). Les unités de la population INH possèdent des afférences ponctuelles de la population EXC. Nous utilisons par la suite les notations introduites dans [Amari, 1977]. La seule différence est que les équations 2.8 sont une version discrétisée en espace des équations introduites par l'auteur. On note u_1 les activités de la population EXC et u_2 les activités de la population INH.

$$\begin{aligned} \forall i, \tau \frac{du_1}{dt}(i, t) &= -u_1(i, t) + \sum_j w_1(i - j) f[u_1(j, t)] - \sum_j w_2(i - j) f[u_2(j, t)] + h_1 + s_1 \\ \forall i, \tau \frac{du_2}{dt}(i, t) &= -u_2(i, t) + w_3 f[u_1(i, t)] + h_2 + s_2 \\ \forall i \in \{1, 2, 3\}, w_i(x) &= \frac{A_i}{\sqrt{2\pi\sigma_i}} e^{-\frac{x^2}{\sigma_i\sqrt{2}}} \end{aligned} \quad (2.8)$$

Les deux phénomènes d'oscillation et de vague d'activité sont obtenus avec la même architecture, en changeant simplement la valeur du taux de décharge spontanée h_1 de la population EXC. Les paramètres de simulation sont résumés dans la table suivante :

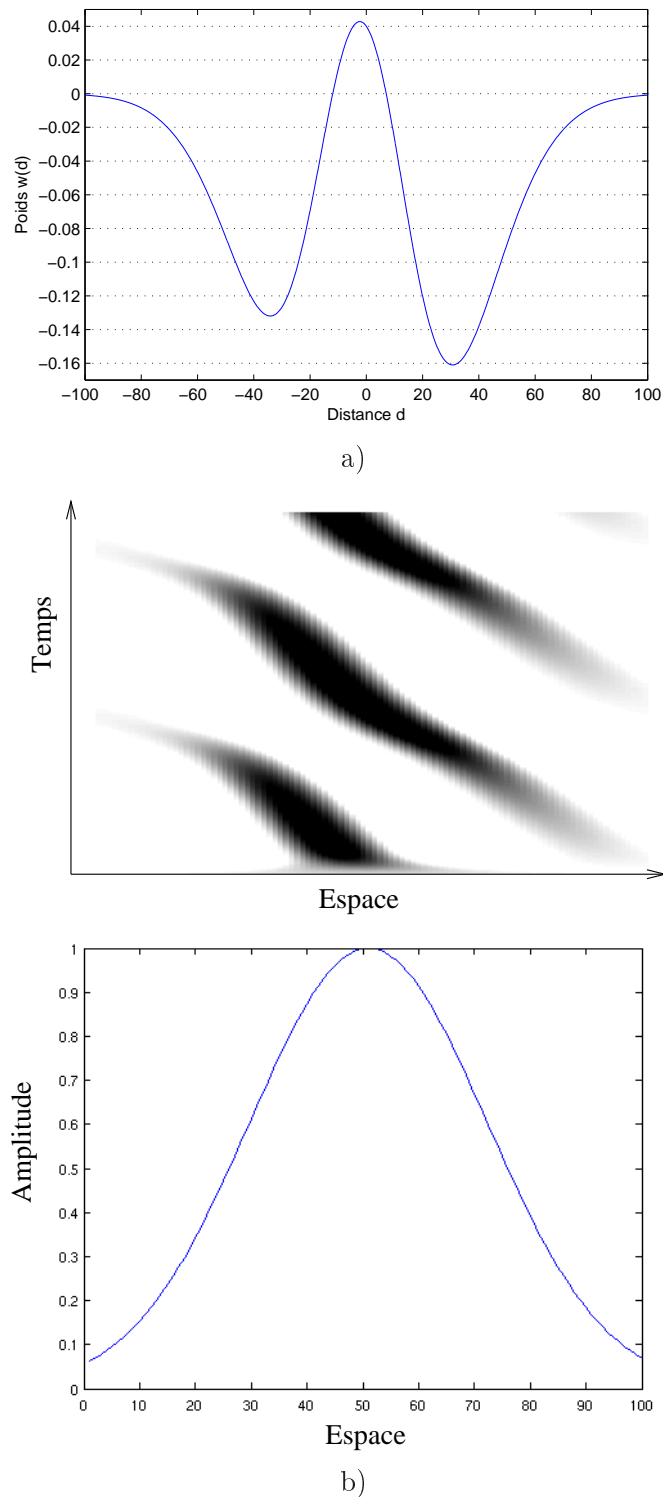


FIG. 2.12 – a) Une fonction de poids asymétrique peut être obtenue en ajoutant à la différence de gaussiennes, sa dérivée pondérée par un poids qui module l’amplitude de la composante asymétrique b) Activités de la population de 100 cellules au cours du temps, étant données une connectivité latérale représentée sur la figure a) et une stimulation représentée en bas de la figure b). L’excitation qui émerge au niveau du maximum de la stimulation, est instable à cause de l’asymétrie des projections latérales. Lorsque la région excitée s’est suffisamment déplacée, une deuxième région excitée émerge et chasse la première avec plus de force, jusqu’à ce qu’à son tour, elle se fasse chasser par une nouvelle région excitée. Les paramètres des projections latérales sont $A_+ = 0.5, \sigma_+ = 0.25, A_- = 0.46, \sigma_- = 0.4$.

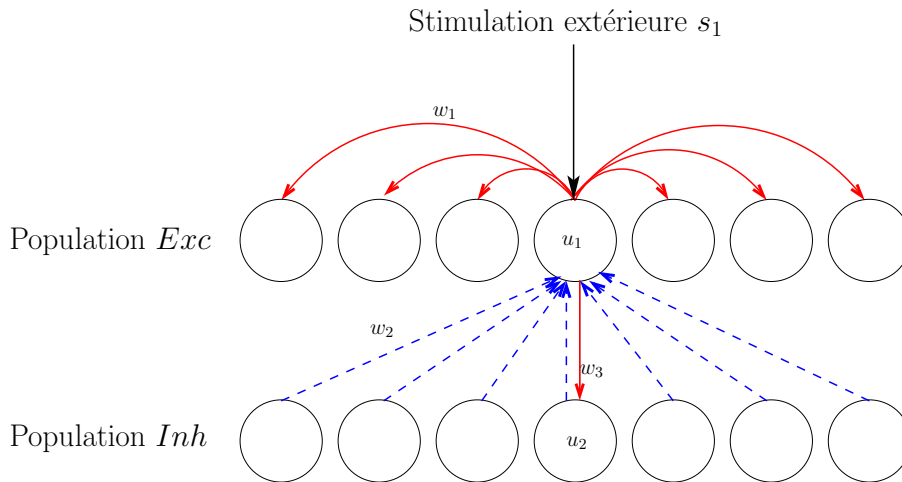


FIG. 2.13 – Illustration du réseau à deux populations de cellules. Les cellules de la population excitatrice *Exc* ont des projections latérales excitatrices et des afférences inhibitrices de la population *Inh*. Les cellules de la population *Inh* ont des afférences ponctuelles excitatrices de la population *Exc*. Les paramètres des projections ainsi que de la dynamique des cellules sont donnés dans la table ci-dessous. Les poids excitateurs sont représentés en trait plein tandis que les projections inhibitrices sont représentées en tirets.

Nom	Type	Valeur
w_1	Gaussienne	Amplitude 2.0, Variance 1.0
w_2	Gaussienne	Amplitude 4.0, Variance 1.5
w_3	Ponctuel	Amplitude 2.0
h_1	Taux de décharge spontanée	0.1 (Oscillation) ou -0.1 (Vague)
h_2	Taux de décharge spontanée	-1.0
s_1	Stimulation extérieure	0.0 (Oscillation), 1.0 ponctuel (Vague)
s_2	Stimulation extérieure	0.0

Les figures 2.14 illustrent les activités des populations excitatrices et inhibitrices dans un graphique espace-temps. L'axe des abscisses représente les différentes cellules, le temps évoluant le long de l'axe des ordonnées. La figure 2.14a représente l'activité de la population *Exc* et la figure 2.14b représente l'activité de la population *Inh*. Sans stimulation extérieure, les populations exhibent des activités oscillantes qui sont uniquement le résultat de la dynamique du réseau.

En modifiant légèrement le taux de décharge spontanée, on observe un deuxième comportement intéressant : la propagation d'une onde d'activité. Les figures 2.15 illustrent les activités dans les populations selon les mêmes conventions que précédemment. Cette fois-ci, on excite à intervalle régulier la cellule centrale de la population excitatrice. On observe alors qu'une onde d'activités se propage dans les deux populations.

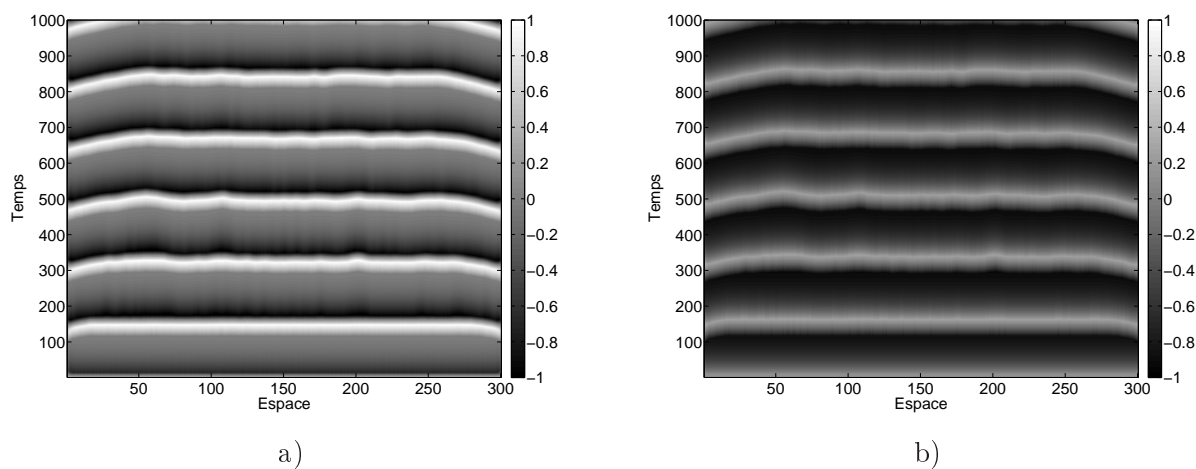


FIG. 2.14 – Activités dans les deux populations faisant apparaître des oscillations. Ces oscillations sont uniquement la conséquence de la dynamique interne du réseau puisqu’aucune stimulation extérieure n’est appliquée. a) Activité de la population EXC au cours du temps. b) Activité de la population INH au cours du temps.

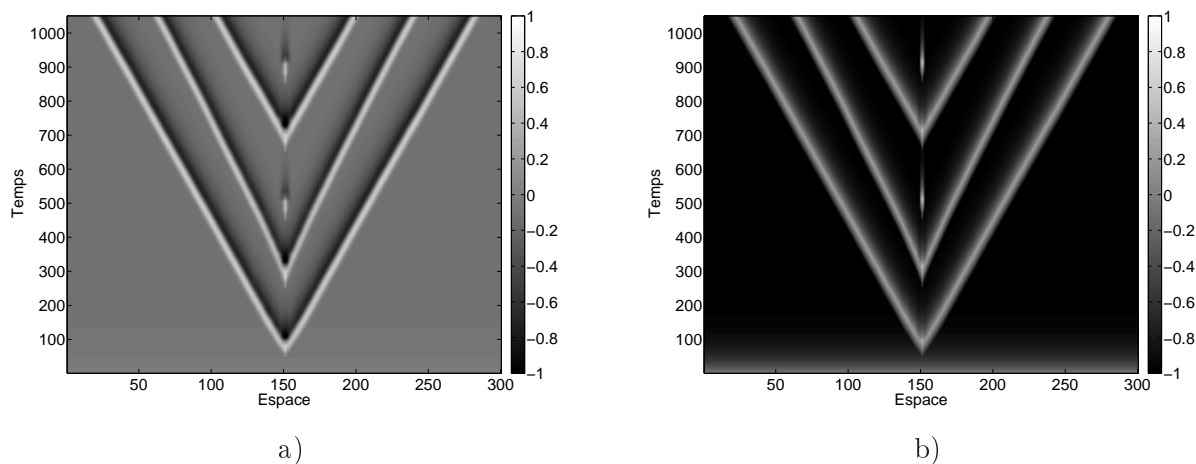


FIG. 2.15 – Activités dans les deux populations au cours du temps faisant apparaître la propagation d’ondes d’activités, depuis un site stimulé à intervalle régulier. La stimulation est ponctuelle, et uniquement dans la population excitatrice. a) Activité de la population EXC au cours du temps. b) Activité de la population INH au cours du temps.

Discussion

Pour conclure ce paragraphe d'introduction des champs neuronaux, revenons sur leurs intérêts à la fois pour l'informatique et pour la modélisation du système nerveux. D'un point de vue informatique, les champs neuronaux sont un exemple de paradigme de calcul qui s'inscrit dans le cadre des calculs locaux, numériques, distribués et adaptatifs¹⁰ que nous avons détaillé dans le paragraphe 2.2. Par ailleurs, les équations différentielles proposées sont relativement simples. Malgré leur simplicité, nous avons vu qu'il peut y avoir différents comportements émergents de l'interaction de populations de cellules, que nous avons illustré au travers de quelques exemples. Certains de ces comportements ont été observés dans le vivant. En particulier [Prechtl et al., 1997] montre qu'un stimulus visuel produit des vagues d'activité dans le cortex visuel de la tortue. On trouvera également d'autres travaux faisant le lien entre différents troubles comme l'épilepsie ou la migraine et la propagation d'activités dans le cortex dans [Pinto et Ermentrout, 2001]. Même si nous ne l'avons pas présenté sous cet angle, les champs neuronaux constituent un outil permettant d'étudier des mécanismes du système nerveux [Schwabe et al., 2006]. Nous avons préféré présenter les champs neuronaux au travers de quelques exemples mais on trouvera des travaux faisant le lien avec le système nerveux ainsi que différents outils mathématiques pour les étudier dans [Beurle, 1956, Wilson et Cowan, 1972, Amari, 1977, Bressloff et Coombes, 1997, Ermentrout, 1998, Vogels et al., 2005, Coombes, 2005].

2.4 Un mécanisme d'attention sélective dans le cadre des champs neuronaux

Nous avons introduit et illustré dans les paragraphes précédents les champs neuronaux. Les prochains paragraphes sont consacrés à la présentation d'un mécanisme de "déploiement de l'attention visuelle". Il apparaît nécessaire de nuancer ce que nous appelons déploiement de l'attention visuelle. Nous avons présenté dans le chapitre précédent l'état de l'art concernant la compréhension de l'attention visuelle chez le primate. Nous avons notamment vu qu'il y a principalement deux interprétations pour le déploiement de l'attention qui ne sont pas encore départagées. D'une part, certains proposent que l'attention spatiale est déployée en parallèle. D'autre part, certains proposent que l'attention spatiale est déployée de manière séquentielle. Le mécanisme que nous proposons dans le reste de ce chapitre repose sur cette seconde hypothèse selon laquelle l'attention visuelle spatiale est déployée séquentiellement sur les stimuli visuels. Nous l'illustrons à la fin du chapitre dans une tâche de recherche visuelle qui consiste à trouver une cible parmi un ensemble de distracteurs.

L'équation différentielle discrétisée (en espace et en temps) régissant la dynamique de l'activité $u(x, t)$ d'un neurone à la position x au temps t utilisée dans le reste du chapitre est la

¹⁰L'adaptabilité de la structure du réseau n'est pas, à proprement parler, une propriété du formalisme des champs neuronaux. Néanmoins, puisque ce formalisme ne décrit que les lois d'évolution des activités des unités, rien n'empêche d'y adjoindre des algorithmes pour modifier la structure du réseau

suivante :

$$\begin{aligned}
 u(x, t + \delta t) &= f(u(x, t) + \Delta u(x, t)) \\
 \tau \Delta u(x, t) &= -u(x, t) + \frac{1}{\alpha} \left(\sum_y w(x, y) u(y, t) + I(x, t) \right) + h \\
 f(x) &= \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x > 0 \end{cases} \\
 w(x) &= A_+ e^{-\frac{x^2}{\sigma_+^2}} - A_- e^{-\frac{x^2}{\sigma_-^2}}
 \end{aligned} \tag{2.9}$$

où h est le taux de décharge spontanée, $\alpha, \tau \in \mathbb{R}^{+*}$ paramètrent la dynamique, f est la fonction de transfert, ici considérée comme une rectification, et $w(x)$ la fonction de poids latéraux en différences de gaussiennes (pour des poids gaussiens, il suffit de prendre A_+ ou A_- nul).

Le mécanisme est présenté par étapes en décomposant ce qui nous paraît être des mécanismes fonctionnels élémentaires :

- un mécanisme de sélection, présenté dans le paragraphe 2.4.1
- auquel est adjoint un mécanisme pour biaiser le déploiement de l'attention spatiale vers des régions sur lesquelles elle n'a pas été récemment déployée, présenté dans le paragraphe 2.4.2
- auquel est adjoint un mécanisme pour biaiser le déploiement de l'attention en faveur de certains attributs visuels, ce que nous avons appelé "l'attention guidée par les attributs visuels" dans le chapitre 1, et présenté dans le paragraphe 2.4.3

Lorsque, dans les prochains paragraphes, nous parlons d'attention visuelle pour décrire les mécanismes proposés, elle doit être comprise comme une analogie fonctionnelle à l'attention visuelle chez le primate.

2.4.1 Un mécanisme de sélection

Présentation du mécanisme

Dans ce paragraphe nous proposons d'illustrer un mécanisme de sélection qui repose sur le paradigme des champs neuronaux. On commence par présenter le mécanisme sans faire référence à la biologie puis nous terminons par une discussion sur l'analogie entre ce mécanisme et les mécanismes attentionnels que nous avons présentés au chapitre précédent.

L'architecture est constituée de deux cartes à deux dimensions INPUT et FOCUS (figure 2.16a). La carte INPUT fournit l'entrée au système. Les unités de la carte FOCUS ont des afférences de la carte INPUT ainsi que des projections latérales en différence de gaussiennes pour lesquelles l'amplitude de la gaussienne excitatrice étroite est égale à l'amplitude de la gaussienne inhibitrice large. L'équation 2.10 est un exemple d'une telle fonction, illustrée sur la figure 2.16. Les connexions latérales sont ainsi uniquement inhibitrices et leur amplitude dépend de la distance entre les unités pré- et post-synaptiques.

$$\begin{aligned}
 w_1(x) &= Ae^{-\frac{x^2}{\sigma^2}} \\
 w_2(x) &= 0.65e^{\frac{x^2}{5.0^2}} - 0.65e^{\frac{x^2}{(\sqrt{2}N)^2}}
 \end{aligned}
 \tag{2.10}$$

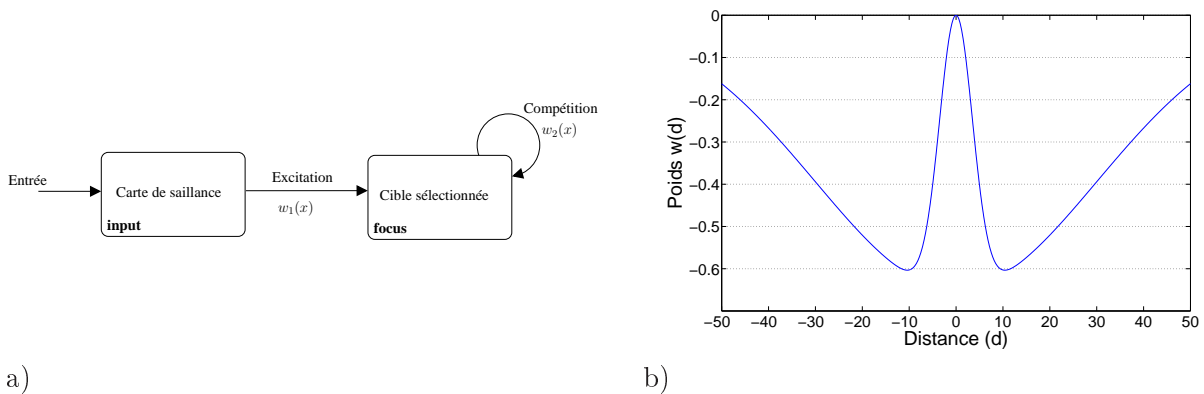


FIG. 2.16 – a) Architecture du système. La carte INPUT excite la carte FOCUS dont les unités ont des projections latérales en différence de gaussiennes. b) Les poids des projections, définis par les équations 2.10, sont uniquement inhibiteurs.

Ces projections latérales inhibitrices induisent de la compétition entre les cellules comme on va le voir dans le prochain paragraphe.

Simulation et résultats

Pour illustrer ce comportement de sélection, on considère une entrée constituée de deux stimuli gaussiens de même variance et de même amplitude (figure 2.17a). On lance la simulation du réseau et, à intervalle régulier, on réinitialise l'activité des unités de la carte FOCUS. La figure 2.17 illustre les activités de la carte INPUT et de la carte FOCUS dans une représentation espace/temps. On observe sur la figure 2.17b qu'à chaque instant, une seule cible est présente dans la carte FOCUS. Cette sélection d'une des cibles émerge des interactions latérales inhibitrices.

La simulation précédente permet d'illustrer que des projections latérales inhibitrices suffisent à obtenir un mécanisme de sélection. Nous aimerions aller plus loin en liant la cible sélectionnée et des paramètres (dans le cas d'une cible gaussienne, ces paramètres peuvent être l'amplitude, la variance mais pourquoi pas également des mesures moins directes qui permettraient de caractériser la "saillance" d'une cible) qui la caractérisent. En particulier, on considère en entrée deux cibles gaussiennes de même variance mais d'amplitude différente. Pour rendre le mécanisme de sélection exploitable, il serait intéressant que la cible dont l'amplitude est la plus importante gagne statistiquement plus la compétition que le distracteur d'amplitude moins importante. Des expériences préliminaires montrent que le mécanisme précédent ayant uniquement des connexions latérales inhibitrices, ne permet pas d'avoir cette propriété :

- soit l'une des deux cibles est sélectionnée avec une probabilité de 1.0
- soit elles sont sélectionnées avec une probabilité de 0.5

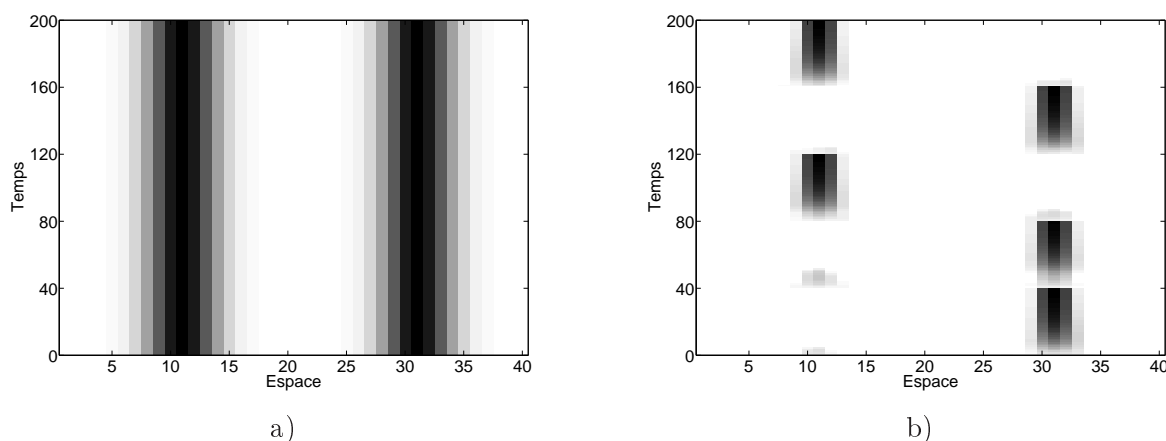


FIG. 2.17 – a) L'entrée du réseau est constituée de deux stimulations gaussiennes de même amplitude et de même variance. b) Les connexions latérales inhibitrices dans la carte FOCUS induisent de la compétition entre les régions excitées de telle sorte que seule une des deux régions gagne la compétition. L'activité dans la carte FOCUS est réinitialisée tous les 40 pas de temps pour relancer la compétition.

On considère alors un mécanisme de sélection dans lequel les projections latérales ont des composantes excitatrice et inhibitrice définies par l'équation suivante :

$$w_2(x) = 1.0e^{\frac{x^2}{5.0^2}} - 0.65e^{\frac{x^2}{(\sqrt{2}N)^2}}$$

L'expérience consiste à présenter les deux cibles et à mesurer quelle cible a été sélectionnée. On reproduit cette expérience un grand nombre de fois (ici 500 fois), en fixant l'amplitude de la première cible (notée s_1) à 1.0 et en faisant varier l'amplitude de la deuxième cible (notée s_2) entre 0.5 et 1.5 (figure 2.18a) ; la distance entre les deux cibles est fixée. La figure 2.18b représente la fonction de poids en trait plein ainsi que les deux cibles s_1 et s_2 relativement à la cellule la plus excitée par s_1 . On observe en particulier que la cible s_2 est dans son voisinage inhibiteur. La figure 2.18c représente la probabilité de sélectionner la cible s_1 en fonction de l'amplitude de la cible s_2 . On observe que la probabilité de sélectionner s_1 diminue si l'amplitude de s_2 augmente. Si les deux cibles ont la même amplitude, la probabilité de les sélectionner est de 50%, ce qui se comprend intuitivement puisque, le problème étant symétrique, il n'y a aucune raison pour qu'une des cibles gagne plus la compétition que l'autre. Ce qui est intéressant, c'est qu'un biais en amplitude d'une cible favorise sa sélection. Nous utiliserons ce principe dans le paragraphe 2.4.3 où nous présenterons une manière d'introduire un biais dans la carte de saillance pour favoriser certains attributs visuels. Puisque la sélection est biaisée en faveur des cibles dont l'amplitude est la plus importante, il suffit d'apporter un biais en amplitude vers certaines cibles pour qu'elles soient statistiquement plus souvent sélectionnées.

Discussion

Le mécanisme présenté dans ce paragraphe, construit dans le cadre du paradigme des champs neuronaux, permet de sélectionner une région spatiale lorsque plusieurs sont excitées. Bien

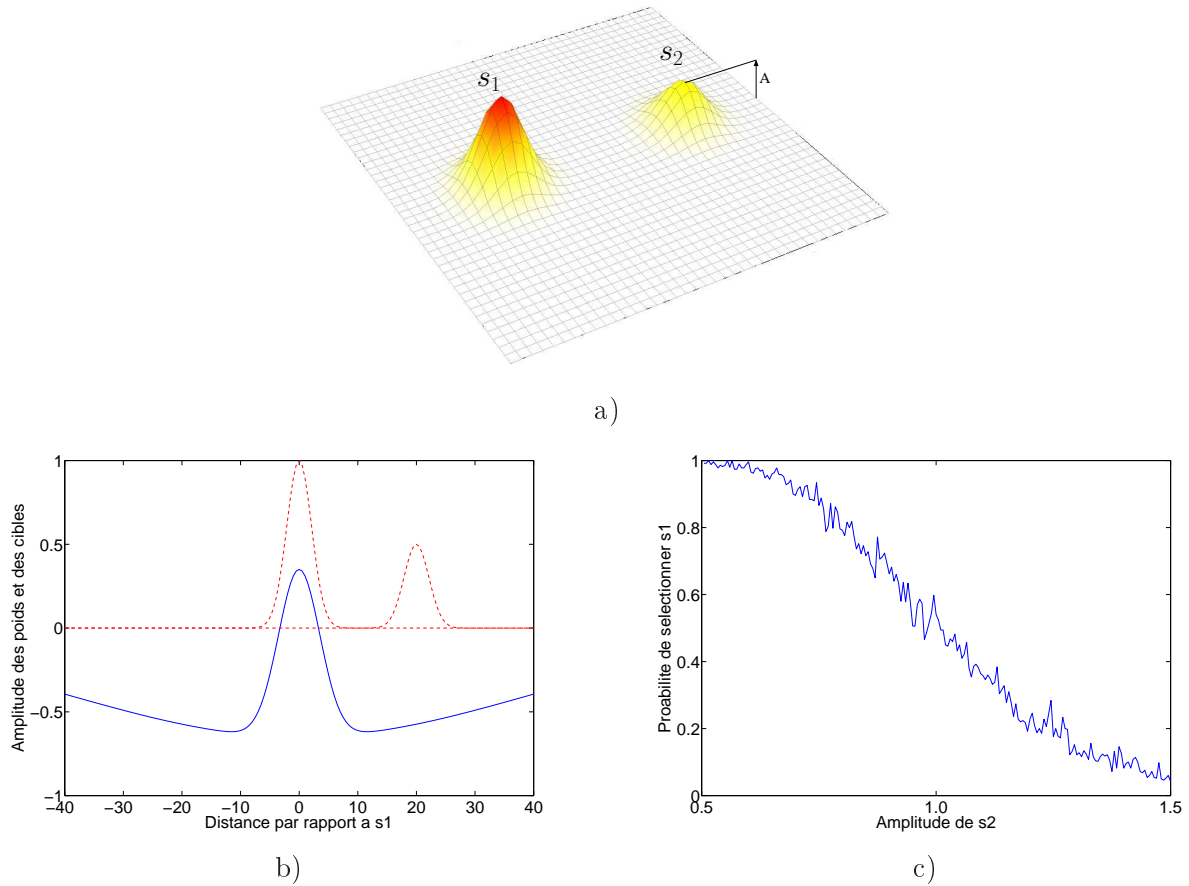


FIG. 2.18 – a) On présente deux cibles gaussiennes en entrée du mécanisme. Ces deux cibles ont la même variance. L'amplitude de s_1 est fixée tandis qu'on fait varier l'amplitude de s_2 . On estime, en répétant 500 fois l'expérience et en mesurant la cible qui a gagné la compétition, la probabilité de sélectionner la cible s_1 en fonction de l'amplitude de s_2 b) La fonction de poids qui définit les projections latérales est représentée en trait plein. Les deux cibles s_1 et s_2 sont représentées sur le même graphique, en pointillés, pour souligner que la cible s_2 appartient au voisinage inhibiteur des cellules excitées par la cible s_1 . c) L'amplitude de s_1 étant fixée à 1.0, la probabilité de la sélectionner est une fonction décroissante de l'amplitude de s_2 . Lorsque les deux cibles ont la même amplitude, la probabilité de sélectionner l'une ou l'autre des cibles est identique.

évidemment, les conditions pour que le mécanisme fonctionne dépendent de plusieurs paramètres comme la forme des stimuli et les paramètres des poids latéraux. Nous avons illustré son fonctionnement par une approche empirique puisque nous ne disposons pas, à ma connaissance, d'outils permettant de formaliser son fonctionnement. Les résultats nous montrent que sous certaines conditions, la probabilité de sélectionner une cible dépend de son amplitude. D'autres propriétés de ce mécanisme de sélection sont étudiées dans [Vitay, 2006] dont la robustesse au bruit et la robustesse aux distracteurs lorsqu'une cible est sélectionnée.

Ce mécanisme est une analogie au déploiement de l'attention spatiale que nous avons présentée au chapitre précédent. Il repose sur l'hypothèse qu'à un certain point, seule une région spatiale est sélectionnée. L'hypothèse que l'attention spatiale est déployée séquentiellement est encore controversée. Néanmoins, si on voit ce mécanisme comme un mécanisme de sélection d'une position spatiale vers laquelle engager une action (comme un mouvement oculaire), la sélection d'une unique région prend du sens. Par ailleurs, d'après la théorie prémotrice de l'attention, ce sont les mêmes aires corticales qui sont impliquées dans le déploiement de l'attention spatiale et la préparation d'un mouvement.

Avant de pouvoir vraiment parler d'un mécanisme analogue au déploiement de l'attention spatiale, il nous reste encore au moins une chose à inclure : la possibilité de désengager l'attention spatiale pour qu'elle puisse être déployée sur une autre région. C'est cette extension que nous abordons dans la prochaine section.

2.4.2 La nécessité d'une forme de mémoire spatiale

Présentation du mécanisme

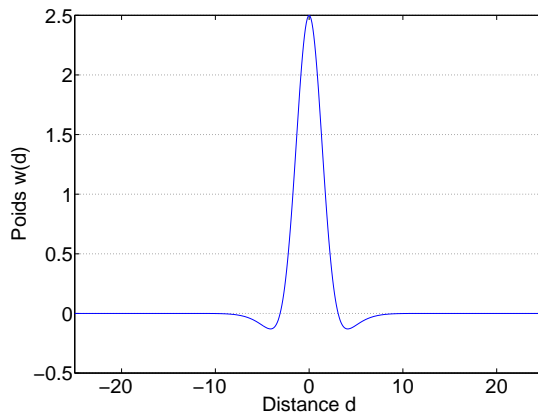
Comme nous l'avons vu au paragraphe 1.1, il existe un mécanisme qui évite que l'attention visuelle ne soit redéployée sur des positions spatiales sur lesquelles elle s'est déployée récemment. Il existe différentes hypothèses quant aux mécanismes neuronaux responsables de ce phénomène dit "d'inhibition de retour". Ce pourrait être une mémoire de travail spatiale dans laquelle les positions sur lesquelles l'attention s'est déployée sont stockées. Mais ce pourrait également être une période réfractaire qui empêche les cellules, dont le champ récepteur couvre la position sur laquelle l'attention s'est récemment déployée, de décharger. Pour la définition d'un modèle, le phénomène d'inhibition de retour, qu'il soit causé par l'un ou l'autre des mécanismes mentionnés, peut être modélisé par une représentation spatiale.

Dans le mécanisme précédent, nous avons utilisé une différence de gaussiennes définie de telle manière que l'inhibition soit suffisamment large pour qu'une région qui gagne la compétition inhibe les autres avec une amplitude suffisamment forte pour que plusieurs régions excitées ne puissent pas coexister. Dans le cas d'une mémoire de travail, on peut considérer des connexions en différences de gaussiennes locales. Dans le cas précédent, l'inhibition couvrait toute la carte. Ici, on considère des connexions uniquement locales et qui par ailleurs ont une composante excitatrice locale. L'excitation locale permet aux unités voisines de s'auto-entretenir et l'inhibition évite que l'activité ne se propage dans toute la carte. Par exemple, considérons les connexions latérales

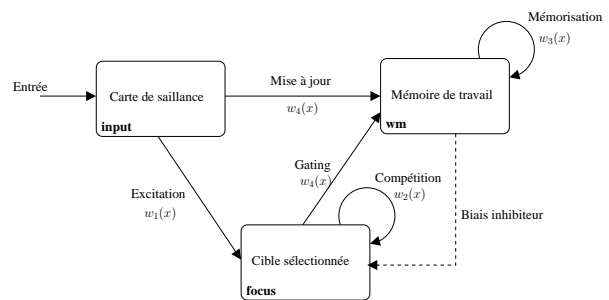
définies par l'équation 2.11 et illustrées sur la figure 2.19a.

$$w_3(x, y) = 3.0e^{-\frac{\|x-y\|^2}{2.0^2}} - 0.5e^{-\frac{\|x-y\|^2}{4.0^2}} \quad (2.11)$$

La figure 2.19b présente l'architecture considérée. L'objectif de ce mécanisme est de sélectionner séquentiellement les cibles présentées en entrée. Pour ce faire, on adjoint au mécanisme de sélection présenté dans le paragraphe précédent la mémoire de travail WM qui mémorise les cibles sélectionnées. L'entrée d'une cible en mémoire de travail est conditionnée par la coactivation des cartes INPUT et FOCUS. Cela signifie qu'une cible entre en mémoire de travail si et seulement si elle est à la fois présente dans la carte INPUT et dans la carte FOCUS. L'excitation de la carte INPUT permet également de conserver une cible en mémoire de travail lorsque le focus attentionnel change de cible, c'est à dire lorsqu'un foyer d'activation de la carte WM n'est plus excité par la carte FOCUS. La mémoire de travail permet de biaiser la compétition dans la carte FOCUS pour favoriser les cibles sur lesquelles l'attention se s'est pas encore déployée. L'influence de la mémoire de travail sur la sélection dans la carte FOCUS est représentée par la projection de principe en pointillés entre WM et FOCUS. Si ces projections inhibitrices connectent directement les deux cartes, le focus attentionnel est changé dès que la cible, sur laquelle l'attention est déployée, entre en mémoire de travail. On verra dans le paragraphe 2.4.3 un mécanisme grâce auquel l'attention est désengagée seulement sous certaines conditions.



a)



b)

FIG. 2.19 – a) Illustration des poids latéraux de la carte WM, définis par l'équation 2.11 b) On adjoint, au mécanisme de sélection présenté dans le paragraphe précédent une mémoire de travail spatiale qui n'est rien d'autre qu'une carte dont les unités ont des projections latérales locales avec une composante excitatrice non nulle.

Simulation et résultats

Pour illustrer le mécanisme, on commence par présenter une simulation n'impliquant que la mémoire de travail puis une simulation de l'architecture complète avec les cartes INPUT, FOCUS et WM.

Mémoire de travail seule. On considère dans ce paragraphe l'architecture constituée uniquement des cartes INPUT et WM (figure 2.19b). Considérons une entrée constituée de trois stimuli dont on fait varier l'amplitude. L'expérience consiste à simuler la sélection par l'attention de chacune des cibles de manière séquentielle, en augmentant l'amplitude des cibles en entrée. La figure 2.20a illustre les entrées qu'on applique à la mémoire de travail. Pour le moment, on ne s'intéresse qu'à ce qui se passe entre 0 et 600 pas de temps. Les trois cibles ont initialement une amplitude assez faible, puis l'amplitude de la première est augmentée au pas de temps 150 avant d'être diminuée. L'amplitude de la seconde cible est ensuite augmentée puis diminuée et on fait de même avec la troisième cible. La figure 2.20b illustre l'activité dans la mémoire de travail. Initialement, l'excitation d'entrée est trop faible pour qu'une excitation émerge dans la mémoire de travail. Au pas de temps 150, l'activité de la première cible devient suffisamment importante pour qu'elle émerge en mémoire de travail. Il est intéressant de constater que la diminution d'amplitude de la cible qui s'ensuit n'affecte pas la présence de la cible en mémoire de travail. Plus précisément, l'amplitude de la cible reste suffisamment importante pour que, combinée avec l'excitation latérale de la mémoire de travail, la cible demeure en mémoire de travail malgré sa diminution d'amplitude ; le même raisonnement est valable pour les deux autres cibles.

Les projections latérales excitatrices permettent de maintenir une information en mémoire de travail lorsque l'amplitude de l'entrée diminue, en compensant la diminution de l'excitation afférente par des excitations latérales. Il est également intéressant de remarquer que la mémoire de travail reste sensible à l'entrée. Pour l'illustrer, on déplace les stimuli à partir du pas de temps 600. On observe alors que les cibles en mémoire de travail suivent l'évolution de l'entrée et restent cohérentes avec les stimuli d'entrée. A partir du pas de temps 1000, on diminue progressivement l'amplitude des stimuli. A partir d'un certain seuil, l'excitation afférente à la mémoire de travail n'est plus suffisante et l'information mémorisée est perdue.

Les paramètres utilisés pour la simulation sont (les notations utilisées sont celles de la figure 2.19b et de l'équation 2.9) :

$$\begin{aligned} N &= 50 \\ w_3(x) &= 3.0e^{-\frac{x^2}{4.0}} - 0.5e^{-\frac{x^2}{16.0}} \\ w_4(x) &= 0.35e^{-\frac{x^2}{4.0}} \\ \text{WM : } \tau &= .75, h = -0.2, \alpha = 12.0 \end{aligned}$$

Mécanisme de déploiement de l'attention spatiale. On considère maintenant l'architecture complète avec les trois cartes INPUT, FOCUS et WM (figure 2.19b). Il y a trois aspects intéressants de la simulation précédente :

- une excitation initiale sous une valeur critique ne suffit pas à faire émerger une cible en mémoire de travail,
- lorsque l'excitation atteint la valeur critique, la cible émerge en mémoire de travail,
- dès qu'une cible est en mémoire de travail, une faible excitation suffit à la conserver en mémoire de travail grâce à des excitations latérales suffisamment fortes pour compenser la diminution de l'excitation afférente et suffisamment faibles pour rester sensible à l'entrée.

Dans l'architecture complète, l'excitation initiale, trop faible pour faire émerger une cible en mémoire de travail, est fournie par l'entrée. Dans l'architecture complète, l'augmentation d'activité correspond à la sélection par la carte FOCUS d'une des cibles. La figure 2.21 illustre le fonction-

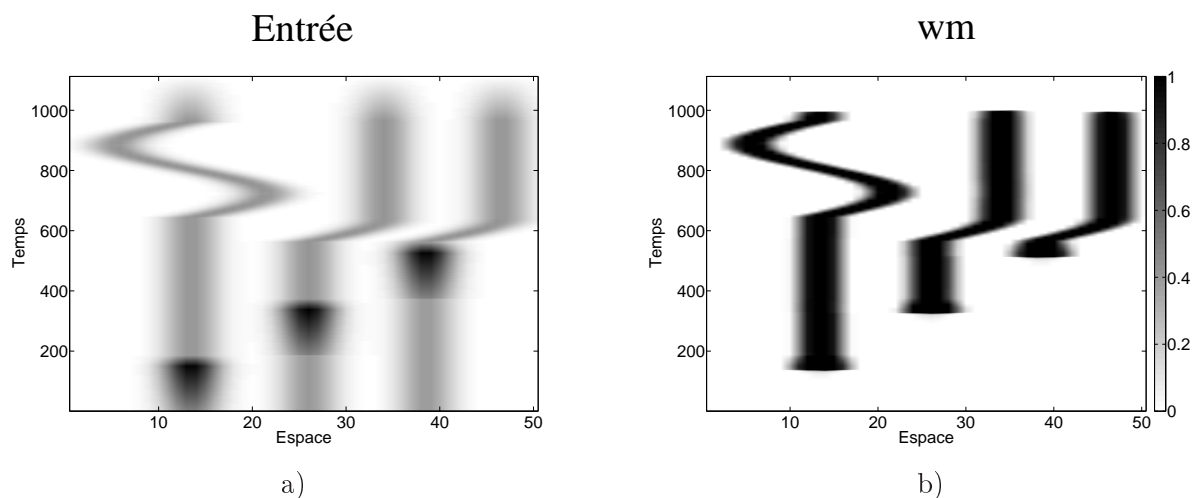


FIG. 2.20 – a) L'entrée est constituée de trois stimuli gaussiens. Entre les pas de temps 0 et 600, on augmente brièvement l'amplitude des cibles les unes après les autres pour simuler une forme de sélection spatiale. A partir du pas de temps 600, les stimuli sont déplacés pour illustrer les propriétés dynamiques de la mémoire. b) Activité dans la carte WM au cours du temps. A partir d'une amplitude critique, une cible émerge dans la mémoire et y reste même si l'amplitude de la cible diminue jusqu'à sa valeur initiale. Les connexions latérales excitatrices sont suffisamment fortes pour compenser la diminution de l'excitation afférente. Elles sont également suffisamment faibles pour rester sensible à l'entrée : à partir du pas de temps 600, les cibles mémorisées suivent la dynamique de l'entrée.

nement du système.

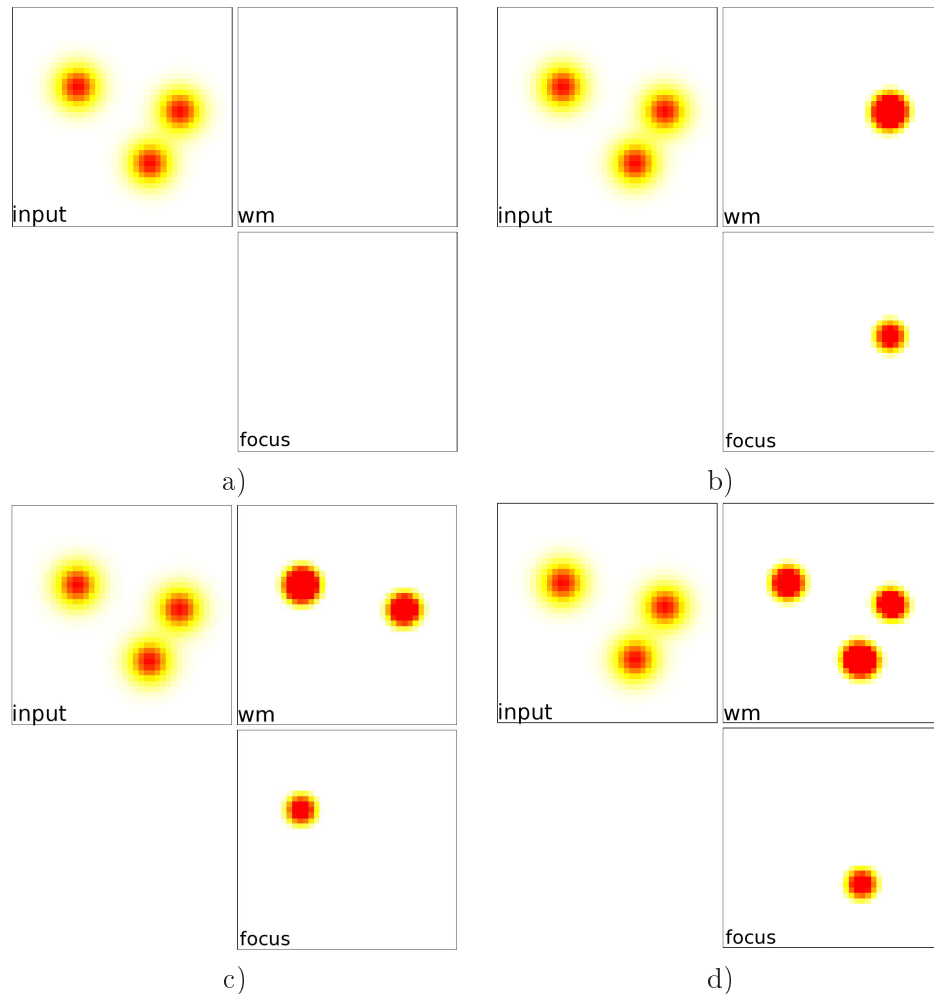


FIG. 2.21 – a) On présente trois cibles en entrée. b) Les trois régions excitées dans la carte FOCUS entrent en compétition et seule l'une d'elle la gagne. La coactivation des cartes INPUT et FOCUS permet l'émergence de la cible sélectionnée dans la mémoire de travail WM. c) Les positions mémorisées en mémoire de travail sont inhibées dans la carte FOCUS, de telle sorte qu'une nouvelle compétition s'engage mais, cette fois-ci, en faveur des régions non encore sélectionnées. La deuxième cible sélectionnée entre en mémoire de travail. d) Les deux cibles précédentes étant correctement mémorisées, la dernière cible gagne la compétition dans la carte FOCUS.

Trois cibles sont présentées en entrée (figure 2.21a). Les trois régions de la carte FOCUS, excitées par l'entrée, entrent en compétition jusqu'à ce qu'une des cibles la gagne (figure 2.21b). La coactivation des cartes INPUT et FOCUS produit une excitation suffisante pour que la cible émerge dans la mémoire de travail WM. Lorsque l'attention doit être désengagée, la mémoire de travail est utilisée pour inhiber les positions mémorisées dans la carte FOCUS par un mécanisme représenté en pointillés sur la figure 2.19b qu'on ne détaillera pas dans ce paragraphe. Nous verrons dans le paragraphe 2.4.3 un exemple d'un tel mécanisme. En conséquence, une nouvelle compétition s'engage entre les trois régions, avec néanmoins un biais en défaveur de la position mémorisée.

Une position spatiale sur laquelle l'attention ne s'est pas encore déployée est ainsi sélectionnée (figure 2.21c). On observe deux phénomènes intéressants. Le premier est que la cible sélectionnée entre en mémoire de travail par la coactivation des cartes INPUT et FOCUS. Le second est que malgré le désengagement de l'attention, qui implique que la région précédemment sélectionnée n'est plus excitée dans WM par FOCUS, la cible reste en mémoire de travail. Cela est dû aux projections latérales excitatrices qui, combinées avec l'excitation de la carte INPUT suffisent à garder la cible en mémoire. Enfin, l'attention est à nouveau désengagée pour se déployer sur la dernière cible (figure 2.21d).

Les paramètres utilisés pour la simulation sont (les notations utilisées sont celles de la figure 2.19 et de l'équation 2.9) :

$$\begin{aligned}
 N &= 40 \\
 w_1(x) &= 1.0e^{-\frac{x^2}{4.0}} \\
 w_2(x) &= 0.65e^{-\frac{x^2}{25.0}} - 0.65e^{-\frac{x^2}{2N^2}} \\
 w_3(x) &= 3.0e^{-\frac{x^2}{4.0}} - 0.5e^{-\frac{x^2}{16.0}} \\
 w_4(x) &= 0.25e^{-\frac{x^2}{4.0}} \\
 \text{FOCUS : } \tau &= 0.75, h = 0.0, \alpha = 7.0 \\
 \text{WM : } \tau &= 0.75, h = -0.25, \alpha = 12.0
 \end{aligned}$$

Discussion

Dans cette section, nous avons montré comment une mémoire spatiale peut émerger d'interactions latérales locales. Si la composante excitatrice de ces interactions latérales est trop forte, la mémoire de travail spatiale ainsi construite peut devenir insensible aux modifications de l'entrée. Nous avons au contraire utilisé une composante excitatrice modérée de telle sorte qu'une fois que la position spatiale d'une cible est mémorisée, si la cible bouge, la mémoire est mise à jour dynamiquement. Cette mise à jour dynamique est uniquement la conséquence des interactions latérales des cellules.

Le mécanisme finalement obtenu, construit à partir du mécanisme de sélection de la section 2.4.1 et du mécanisme de mémoire de travail spatiale, peut être considéré comme une analogie fonctionnelle au déploiement séquentiel de l'attention spatiale. A chaque instant, seule une région spatiale est sélectionnée. Lorsqu'elle est sélectionnée, elle est mémorisée. La mémoire spatiale peut ensuite être utilisée pour biaiser la compétition en faveur d'une région qui n'a pas encore été sélectionnée. Par analogie avec le déploiement de l'attention spatiale, la mémoire de travail spatiale que nous utilisons pourrait modéliser le phénomène d'inhibition de retour. Néanmoins, l'inhibition de retour a une portée temporelle limitée. Ici, il est important de noter que la mémoire spatiale que nous avons proposée maintient l'information tant que le stimulus est présent dans le champ. Si nous voulons que l'information ne soit mémorisée que pendant une certaine durée, il est nécessaire soit de modifier les paramètres de la mémoire, soit de modifier l'architecture. Quoi qu'il en soit, ne serait-ce que pour guider le déploiement de l'attention vers des régions sur lesquelles elle ne s'est pas encore déployée, le mécanisme est fonctionnel.

Nous avons vu dans le chapitre précédent que parmi les différentes facettes de l'attention visuelle, elle peut être guidée par les attributs visuels. Jusqu'à maintenant, nous n'avons traité que des informations spatiales. Dans le prochain paragraphe, nous proposons un mécanisme grâce auquel on peut inclure les attributs visuels comme la couleur ou l'orientation.

2.4.3 Intégrer des attributs non spatiaux : un mécanisme de recherche visuelle

Présentation du mécanisme

Les deux mécanismes précédents permettent de parcourir séquentiellement plusieurs cibles présentées en entrée. Pour le moment, nous nous sommes intéressés uniquement à l'information spatiale. Dans ce paragraphe, nous proposons d'étendre le mécanisme précédent en incluant des attributs visuels en vue d'obtenir un mécanisme d'exploration visuelle guidée par des attributs visuels. Cette extension est en partie inspirée des travaux de [Hamker, 2004, Hamker, 2005c] que nous avons présentés succinctement au paragraphe 1.4.3.

Comme nous l'avons vu au paragraphe 2.4.1 traitant du mécanisme de sélection, la probabilité de sélectionner une cible dépend de son amplitude. Ainsi, si on souhaite qu'une cible soit privilégiée par rapport aux autres, il suffit de lui apporter un biais en amplitude. C'est l'un des principes que nous allons utiliser dans le mécanisme présenté dans ce paragraphe. Le mécanisme comprenant le mécanisme de sélection ainsi que la mémoire de travail peut être considéré comme un mécanisme d'exploration spatiale. Nous souhaitons l'étendre pour lui ajouter deux propriétés :

- biaiser l'exploration spatiale en faveur des cibles ayant certains attributs visuels (couleur, orientation, ...)
- extraire les attributs visuels de la cible sélectionnée spatialement

L'architecture proposée est illustrée sur la figure 2.22. Elle se compose d'une partie qui traite uniquement l'information spatiale (voie "Where") et qui est constituée du mécanisme présenté dans le paragraphe précédent (mécanisme de sélection et mémoire de travail spatiale) ainsi que d'une partie qui traite uniquement les attributs visuels (voie "What").

Un point essentiel dans ce paragraphe est de comprendre comment la carte de saillance PRE-MOTOR est modulée pour favoriser certaines positions spatiales qu'occupent des cibles potentiellement intéressantes. L'information visuelle est tout d'abord filtrée selon plusieurs dimensions (on considérera dans l'exemple les deux couleurs bleu/vert et les deux orientations 45° et 135°). Le résultat de ces filtres définit l'activité des cartes I_{bleu} , I_{vert} , I_{45° , I_{135° . Par la suite on note M_* l'ensemble des cartes M_{bleu} , M_{vert} , M_{45° , M_{135° où M peut être I , $V4$, IT ou PF . Les cartes $V4_*$ représentent un niveau intermédiaire dans le réseau ; elles intègrent :

- l'information visuelle montante, provenant des cartes I_*
- un biais spatial provenant de la carte PRE-MOTOR et qui favorise la position spatiale sélectionnée dans la voie "Where" en agissant comme un gain sur l'information montante provenant de I_* ,
- un biais en faveur de certains attributs provenant de la voie "What" et qui favorise les

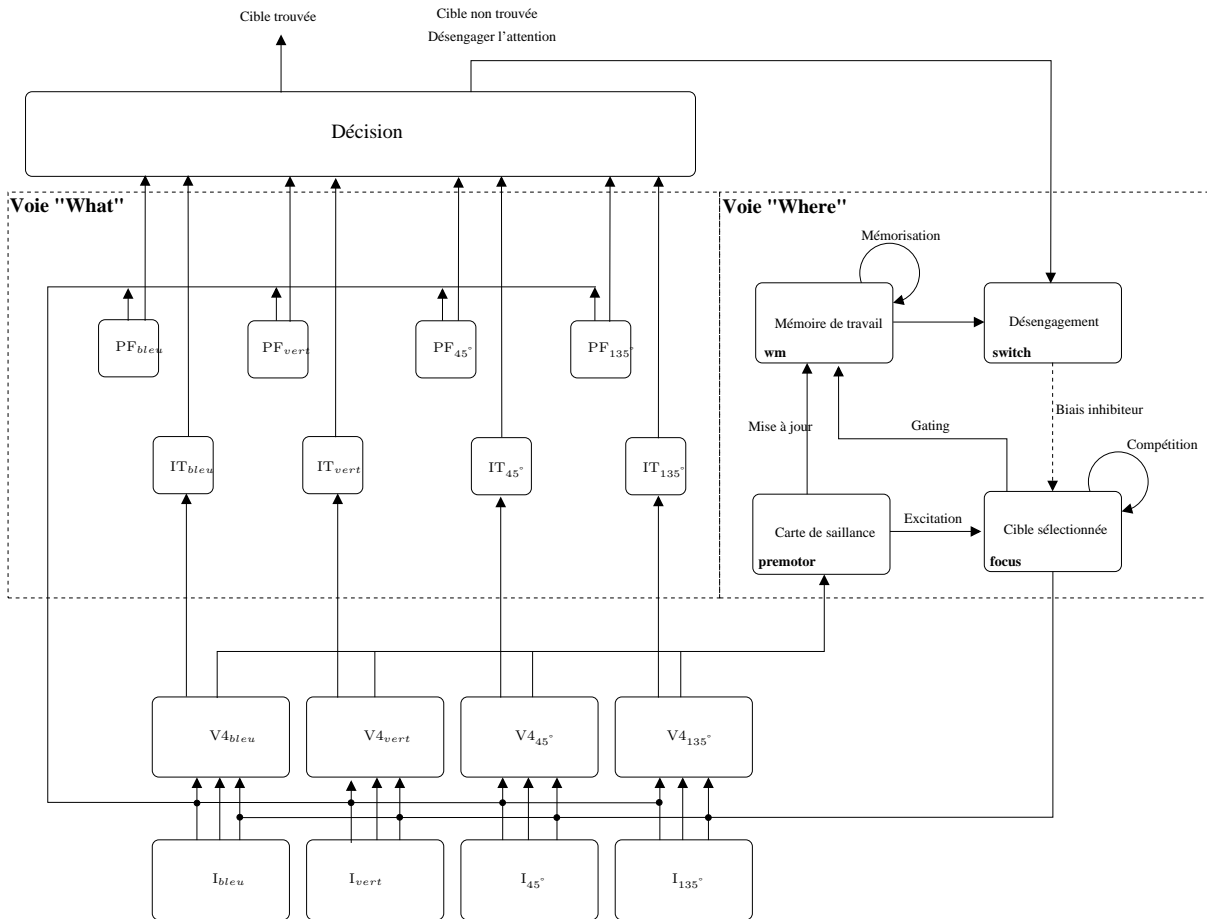


FIG. 2.22 – Mécanisme de déploiement *covert* de l'attention visuelle. A la voie "Where" détaillée dans les paragraphes précédents, on adjoint des mécanismes qui permettent de moduler la carte de saillance PREMOTOR en fonction des attributs visuels recherchés. L'entrée visuelle est filtrée selon plusieurs dimensions. Le résultat de ces filtres définit l'activité des unités des cartes I_* . Ces cartes excitent les cartes $V4_*$ qui servent de niveau intermédiaire dans le système. Ces cartes ont des afférences de l'entrée visuelle, des unités représentant les attributs à rechercher PF_* ainsi que de la carte de sélection spatiale FOCUS. Toutes les équations d'évolution des unités ainsi que les paramètres du modèle sont disponibles dans les annexes A.1 et A.2

positions spatiales qui contiennent une cible partageant l'un des attributs, en agissant comme un gain sur l'information montante provenant des cartes I_* .

Puisque nous avons déjà détaillé la voie "Where" dans les paragraphes précédents, nous allons passer plus de temps sur la voie "What". Cette voie est constituée de deux principaux ensembles de cartes. D'une part, les cartes IT_* ont des afférences des cartes $V4_*$ qui leur sont associées. Par exemple, la carte IT_{bleu} n'a des afférences que de la carte $V4_{bleu}$. Par ailleurs, chaque carte IT n'est constituée que d'une seule unité dont le champ récepteur couvre tout le champ visuel et qui calcule le maximum des activités des unités dans son champ récepteur. Les activités de ces unités fournissent ainsi une indication sur la "quantité" d'un attribut dans tout le champ visuel, extraite des activités des cartes $V4_*$. Toute l'astuce consiste à modifier les activités des unités dans $V4_*$ pour que les informations extraites par les unités IT_* représentent les attributs de la cible sélectionnée spatialement dans la voie "Where". Les cartes PF_* représentent la mémoire des attributs visuels qui sont pertinents pour la tâche. Ces cartes ne contiennent également qu'une unité. Ce sont ces unités qui fournissent un biais aux représentations $V4_*$ pour favoriser la représentation des positions spatiales qu'occupent des stimuli qui partagent au moins un attribut avec la cible. Il est possible d'associer une sémantique à l'activité des deux ensembles de cartes IT_* et PF_* :

- les cartes IT_* permettent de décoder les attributs visuels de la cible sélectionnée spatialement dans la voie "Where"
- les cartes PF_* déterminent les attributs visuels recherchés

Puisque les cellules PF_* favorisent la représentation dans $V4_*$ des positions spatiales occupées par des stimuli qui partagent au moins un attribut visuel avec la cible, la carte `PREMOTOR`, qui intègre les activités des cartes $V4_*$ peut être considérée comme une carte de saillance. En effet, elle constitue une représentation spatiale, indépendante des attributs visuels, qui représente la pertinence comportementale de chaque position du champ visuel. C'est sur la base de cette représentation que la carte `FOCUS` sélectionne la prochaine région sur laquelle déployer l'attention spatiale. La région sélectionnée excite en retour les cartes $V4_*$ en agissant comme un gain sur l'information montante provenant de I_* . Ainsi, les représentations dans $V4_*$, en plus d'être modulées par les attributs pertinents pour la tâche, sont également modulées par la position spatiale sélectionnée. Cette seconde modulation a la conséquence indirecte de biaiser les représentations dans les cartes IT_* , de telle sorte que les attributs visuels d'une cible sélectionnée spatialement dans la voie "Where", peuvent être décodés dans la voie "What" des activités des cartes IT_* .

L'adéquation entre les attributs visuels recherchés et les attributs visuels de la cible sélectionnée spatialement est estimée par des cellules du module `DECISION`. Le résultat de l'intégration de ces cellules permet de décider quel comportement adopter :

- soit la région sélectionnée contient la cible recherchée, auquel cas la recherche est terminée
- soit la région sélectionnée ne contient pas la cible recherchée, auquel cas l'attention spatiale doit être désengagée

Le désengagement de l'attention spatiale est obtenu en inhibant dans la carte `FOCUS` les régions sur lesquelles l'attention spatiale s'est déjà déployée. Le mécanisme qui permet de désengager l'attention est la carte `SWITCH` de la voie "Where" qui a des afférences des unités de décision ainsi que de la mémoire de travail et dont les unités inhibent la carte `FOCUS`. Le mécanisme le plus simple consiste à utiliser une carte, telle que l'excitation de la mémoire de travail seule ne suffise pas à exciter ses cellules. Une excitation complémentaire des unités de décision suffit à exciter ces cellules qui inhibent ainsi la carte `FOCUS`. On trouvera dans [Vitay et Rougier, 2005] un mécanisme de désengagement de l'attention plus sophistiqué et inspiré du fonctionnement des

ganglions de la base.

Simulation et résultats

On considère une tâche de recherche visuelle très simple qui consiste uniquement à localiser une cible définie par ses attributs visuels. Les stimuli sont des barres colorées bleues ou vertes et orientées à 45° ou 135° . On considère que la cible à localiser est la barre bleue orientée à 45° qui est unique dans les exemples proposés, et entourée de distracteurs.

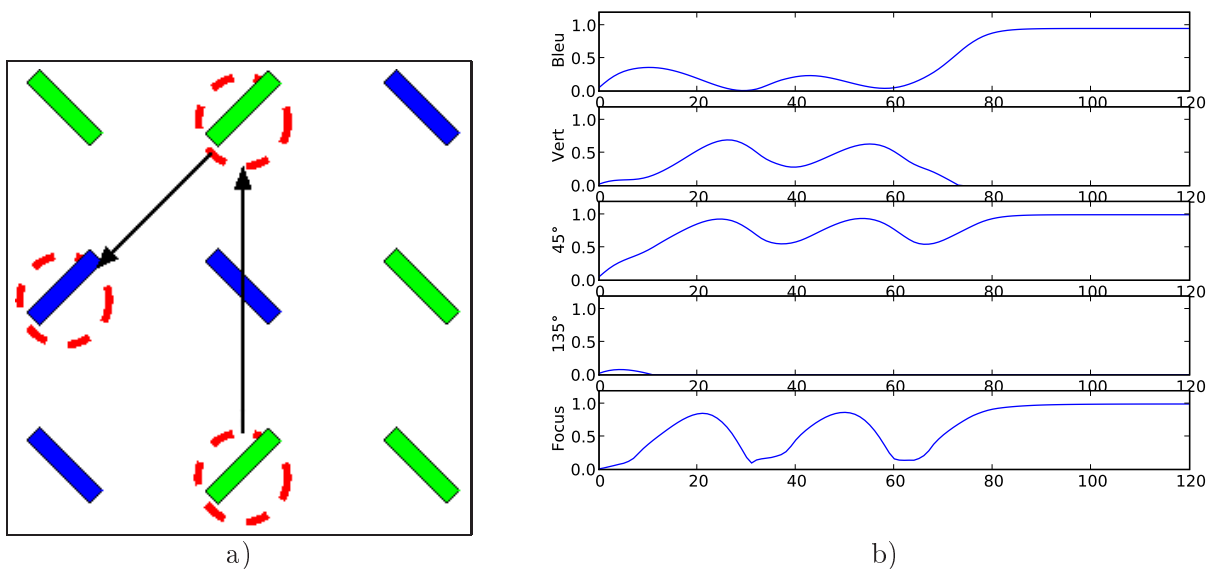


FIG. 2.23 – a) Exemple de parcours de l'attention visuelle spatiale. La tâche consiste à localiser la cible bleue orientée à 45° . L'attention visuelle est d'abord déployée sur deux distracteurs verts avant de sélectionner la cible. b) Pendant que le système réalise la tâche, on enregistre les activités des cellules IT_{bleu} , IT_{vert} , IT_{45° , IT_{135° ainsi que l'activité moyenne normalisée de la carte FOCUS. Ces cartes constituent de bons indicateurs de l'état du système à un instant donné.

La figure 2.23 illustre le déploiement de l'attention ainsi que l'enregistrement de certaines cellules du mécanisme pendant la réalisation de la tâche. Sur la figure 2.23a, le parcours par l'attention visuelle spatiale, dont la position est décodée dans la carte FOCUS, est représenté par les cercles en pointillés. Pour expliciter ce que "perçoit" le système, l'activité des cellules IT_* ainsi que l'activité moyenne normalisée dans la carte FOCUS sont représentées sur la figure 2.23b. L'activité dans la carte FOCUS permet de savoir quand une cible est sélectionnée spatialement. A chaque sélection, il est possible de décoder les attributs visuels de la cible sélectionnée spatialement à partir des activités des cellules IT_* . Ces cinq mesures constituent ainsi de bons indicateurs de l'état du système. On observe ainsi que les cibles sélectionnées sont :

- une cible verte orientée à 45°
- suivie d'une cible verte orientée à 45°

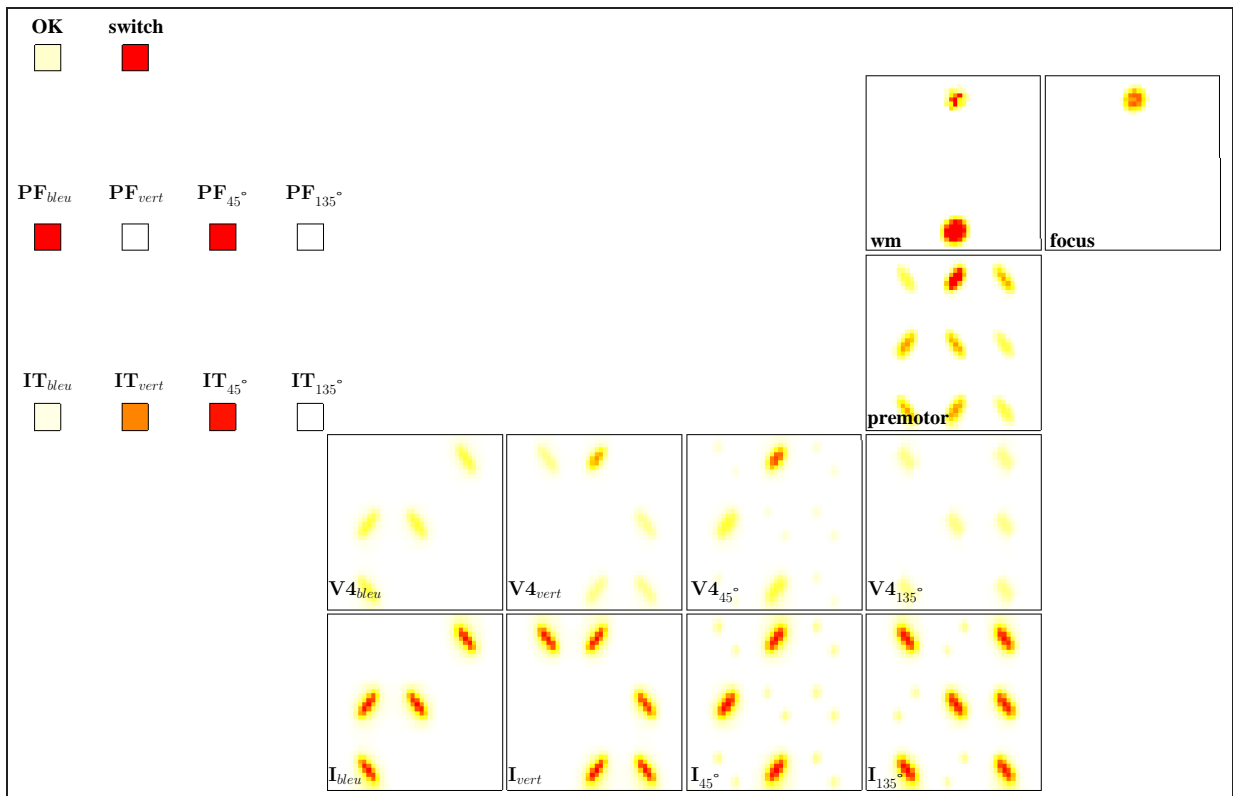
– suivie d'une cible bleue orientée à 45°

La figure 2.24 représente l'état du système après la sélection du deuxième stimulus ainsi qu'une fois la tâche terminée et la cible localisée. Sur la figure 2.24a on peut noter plusieurs choses intéressantes. Puisque la cible à rechercher est bleue et orientée à 45° , l'amplitude des activités des cartes $V4_{bleu}$ et $V4_{45^\circ}$ est plus importante que l'amplitude des activités des cartes $V4_{vert}$ et $V4_{135^\circ}$. Sur les illustrations, ces modulations sont relativement faibles et difficilement perceptibles. Elles sont beaucoup plus facilement perceptibles dans la carte de saillance PRE-MOTOR. On constate en effet que toutes les positions spatiales occupées par un stimulus qui partage au moins un attribut avec la cible ont une amplitude plus importante. On observe également que parmi toutes les cibles représentées dans $V4_*$, celle qui est sélectionnée spatialement dans la carte FOCUS a un biais en amplitude supplémentaire. C'est ce biais en amplitude qui permet de biaiser indirectement les activités des cartes IT_* desquelles sont extraites les attributs de la cible qui occupe la position spatiale sur laquelle l'attention est déployée. Enfin, puisque le stimulus sélectionné n'a pas tous les attributs recherchés, le neurone SWITCH décharge et module l'inhibition de la mémoire de travail spatiale WM (qui contient les deux premières régions spatiales sélectionnées) sur la carte FOCUS. La figure 2.24b représente l'état du système une fois la tâche terminée. La région sélectionnée dans la carte FOCUS contient la cible recherchée. Ainsi, le neurone que nous avons noté OK sur les illustrations décharge, indiquant que la cible est localisée.

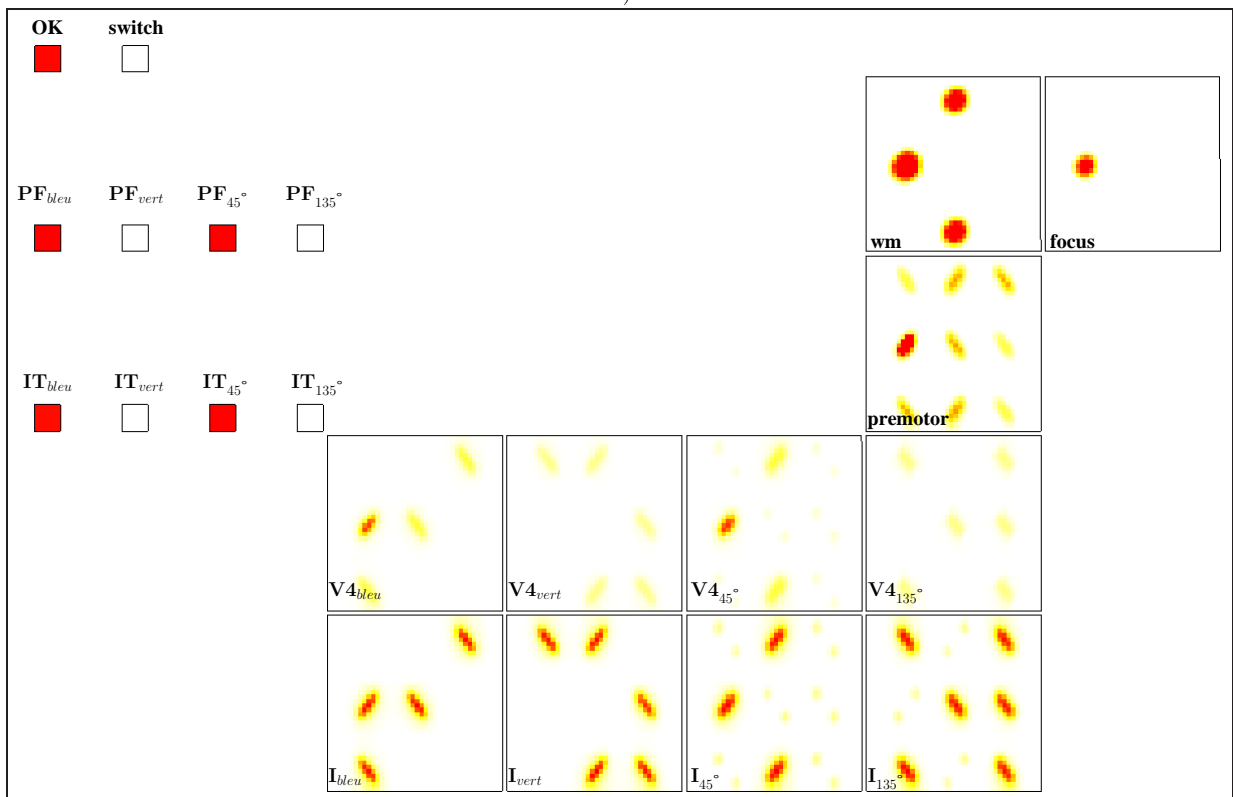
D'autres exemples de parcours par l'attention sont présentés sur les figures 2.25 pour différentes configurations de l'entrée.

Discussion

Dans cette section, nous avons montré comment il était possible d'inclure les attributs visuels pour guider le déploiement de l'attention. Non seulement les attributs cibles guident le déploiement de l'attention spatiale mais il est également possible d'extraire les attributs d'une cible sélectionnée spatialement. Il y a une forme de symétrie entre la sélection spatiale et la sélection par les attributs. On pourrait par exemple imaginer que la sélection spatiale se fasse dans une autre modalité, par exemple auditive. Si le signal sonore est localisé, sa position pourrait venir en *feedback* biaiser les représentations $V4_*$, ayant pour conséquence de biaiser les représentations IT_* . Ainsi, avec la même architecture, il est possible d'inclure la sélection spatiale dans d'autres modalités. L'architecture que nous proposons considère un ensemble limité d'attributs visuels. Les travaux classiques sur le calcul de la carte de saillance utilisent néanmoins des filtres plus complexes avec notamment le calcul d'oppositions de couleurs (vert/rouge ou bleu/jaune) ou le calcul de plusieurs orientations. On distingue en général la **dimension** et la **valeur**. Pour les couleurs, les dimensions seraient par exemple vert/rouge et bleu/jaune, la valeur étant le résultat du filtrage. Dans l'architecture que nous venons de présenter, nous confondons dimension et valeur : une cellule de la carte I_{bleu} décharge si il y a beaucoup de bleu dans son champ récepteur ; une cellule de la carte I_{45° décharge uniquement si dans son champ récepteur, le stimulus est orienté à 45° . Une approche plus réaliste serait de considérer des populations de cellules pour chaque dimension, de telle sorte qu'une cellule dans la population ait une certaine courbe de sélectivité (*tuning curve*) pour les valeurs de la dimension considérée. En pratique c'est ce que propose [Hamker, 2005c] mais pour des raisons de simplicité, nous ne l'avons pas considéré. Inclure ces propriétés aurait deux avantages. D'une part, les stimuli dans des scènes



a)



b)

FIG. 2.24 – a) Activités des différentes unités après la sélection de la deuxième cible. La cible sélectionnée est en train d'émerger dans la mémoire de travail. Au même moment, le biais spatial rétropropagé vers les cartes intermédiaires $V4_*$ biaise ces représentations, de telle sorte que la voie "What" "sait" déjà que le stimulus sélectionné est un distracteur. b) Une fois que la cible a été localisée, le système reste dans un état stationnaire. En particulier, la mémoire de travail conserve l'ensemble des positions spatiales qui ont été sélectionnées.

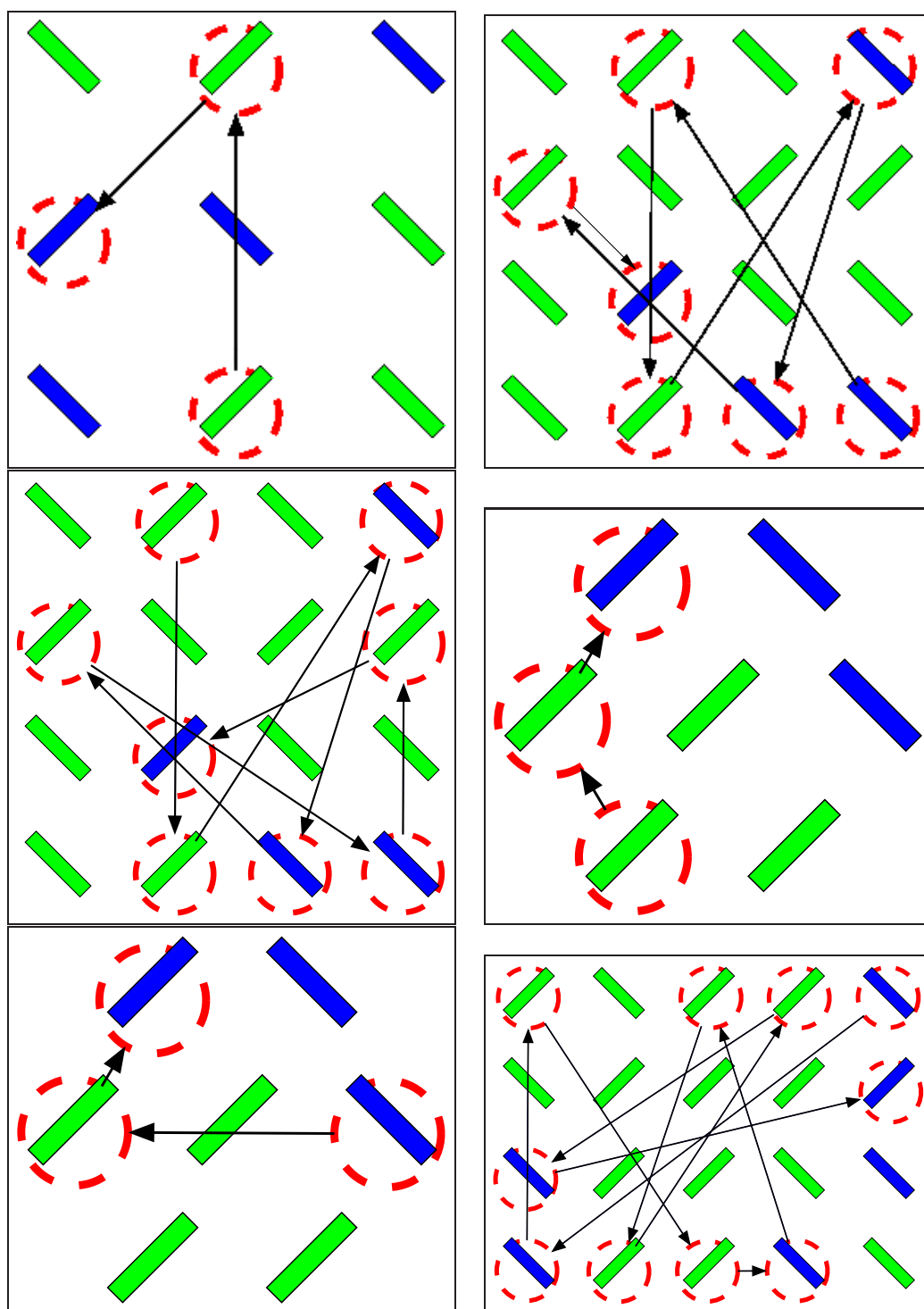


FIG. 2.25 – Exemples de parcours par l'attention spatiale pour différentes configurations de stimuli. La cible recherchée est la barre bleue orientée à 45° . L'attention visuelle est déployée séquentiellement, et en priorité sur les cibles qui partagent au moins un attribut avec la cible (en particulier, les cibles vertes orientées à 135° ne sont jamais sélectionnées). Les cibles sélectionnées spatialement sont représentées par un cercle en pointillé. Le sens du parcours est orienté par les flèches.

naturelles ne pourraient pas être discriminés uniquement sur la base des attributs élémentaires que nous considérons ici. Enrichir la représentation des attributs est un pas vers l'application de ces mécanismes en vision par ordinateur. Par ailleurs, [Navalpakkam et Itti, 2006] montrent que des variations de valeur dans une même dimension peuvent guider les mouvements oculaires. Sous l'hypothèse que l'attention se déploie sur une région de l'espace avant qu'un mouvement oculaire ne soit dirigée vers elle, on peut imaginer que l'attention peut être guidée en faveur d'un intervalle de valeur d'une même dimension plutôt qu'uniquement en faveur de la dimension.

Les extensions citées précédemment sont des extensions possibles du mécanisme. Néanmoins, nous pensons qu'en principe, cette architecture est une architecture minimale relativement simple, définie dans le cadre des champs neuronaux, dont le comportement peut être considéré comme une analogie fonctionnelle au déploiement de l'attention visuelle spatiale ou guidée par les attributs visuels.

Conclusion

Dans ce chapitre, nous avons proposé un modèle d'attention visuelle *covert* défini dans le cadre des champs neuronaux. La question à laquelle nous cherchions à répondre est : est-il possible de construire un mécanisme tel que l'attention visuelle puisse émerger de l'interaction locale de cellules élémentaires. Pour ce faire, nous avons commencé par présenter les fondements du connexionnisme que sont le calcul local, numérique, distribué et adaptatif et nous les avons détaillés. Puis, nous avons présenté le paradigme des champs neuronaux qui est un exemple de paradigme de calcul qui inclut ces contraintes. Nous avons illustré les intérêts de ce paradigme de calcul au travers de quelques exemples. Notre choix s'est porté sur ce paradigme en particulier puisqu'il nous paraît être un bon compromis entre les ressources nécessaires pour le simuler et son réalisme biologique.

Dans la deuxième partie du chapitre, nous avons présenté un modèle de déploiement de l'attention visuelle en le décomposant en ce qui nous paraissent être des composants fonctionnels élémentaires et qui sont :

- un mécanisme de sélection,
- auquel est adjoint un mécanisme de mémoire spatiale,
- auquel est adjoint un mécanisme pour inclure les attributs visuels.

L'architecture obtenue a enfin été appliquée à une tâche de recherche visuelle. On peut considérer cette architecture comme une architecture minimale qui contient uniquement les composants nécessaires au déploiement de l'attention sous l'hypothèse qu'elle est déployée séquentiellement. Nous avons à plusieurs reprises souligné que l'analogie avec l'attention spatiale est une analogie fonctionnelle. Ici, nous évoquons également que l'architecture est minimale. Cela signifie que la comparaison entre les cartes de neurones et les aires corticales n'est pas immédiate. En particulier, les fonctions élémentaires que nous présentons sont très localisées dans le système, même si elles émergent de l'interaction en parallèle des différentes cellules. Dans le système nerveux, ces "fonctions" sont certainement plus distribuées.

Dans le reste du manuscrit, on étudie comment il est possible d'inclure les mouvements

oculaires dans cette architecture en se reposant sur l'hypothèse que la même architecture peut être utilisée pour le déploiement *overt* et *covert* de l'attention, c'est à dire avec ou sans mouvement oculaire. Dans le prochain chapitre, nous résumons l'état actuel de nos connaissances sur le contrôle des mouvements oculaires et nous proposons dans le dernier chapitre des mécanismes qui permettent de les prendre en compte.

Chapitre 3

Le contrôle cortical et sous cortical des saccades oculaires : représentation et manipulation de l'information spatiale

The more we find out about the way in which our eyes seek out and capture objects of interest in the visual world, the more remarkable does its seemingly effortless precision and sophistication appear. At the same time, it functions at a level of complexity between the banality of the spinal reflex and the inscrutability of the voluntary act

Roger Carpenter
Movements of the eyes, 2nd ed., 1988

Introduction

L'organe sensoriel de la vision est l'œil, dont la rétine est tapissée de photorécepteurs qui convertissent un signal lumineux en signal électrique. Comme nous le verrons plus en détail dans ce chapitre, ces photorécepteurs sont classés en deux catégories : les bâtonnets et les cônes. Les bâtonnets sont essentiels pour la vision nocturne, et ne sont sensibles qu'à un domaine restreint de longueurs d'onde, tandis que les cônes permettent d'analyser en détail une information visuelle. Ces cônes sont en effet divisés en trois catégories, en fonction du domaine de longueurs d'onde auquel ils sont sensibles, et plus denses dans une région localisée de la rétine qu'on appelle la fovéa. Les dimensions de la fovéa correspondent environ aux dimensions de la projection, sur la rétine, d'un citron qu'on observe lorsqu'il est tenu dans la main, bras tendu. Cette grande densité de photorécepteurs en une région localisée de l'espace permet d'extraire les détails fins d'un stimulus visuel. La densité des cônes décroît rapidement avec l'excentricité par rapport à la fovéa. L'un des avantages de cette densité des photorécepteurs est de pouvoir concentrer son

analyse sur une région restreinte de l'espace visuel mais l'un des inconvénients est que les informations visuelles extraites dans la vision périphérique sont beaucoup moins précises.

Chez les primates, l'exploration d'une scène visuelle consiste alors à "déplacer" la fovéa sur les régions de l'espace visuel. Il existe différents mouvements oculaires pour réaliser cette fonction, certaines sont monoculaires, d'autres binoculaires, certains déplacent lentement les yeux pour suivre une cible, d'autres changent rapidement et brutalement l'orientation du regard. Les saccades oculaires appartiennent à cette dernière catégorie de mouvement oculaire : elles sont rapides, potentiellement de grande amplitude et fréquents. Les primates effectuent en effet de l'ordre de trois saccades par seconde. L'une des conséquences de ces saccades est que la région de l'espace visuel, de laquelle la rétine reçoit un signal lumineux, est brutalement modifiée. Pour expliciter les conséquences de ces modifications brutales, considérons un exemple imagé. Imaginez que vous soyez entrain de tourner un film et que vous bougiez la caméra à une fréquence de trois fois par seconde. Arrêtez l'enregistrement et regardez le film que vous venez tourner, il est probable que vous trouviez cela désagréable, mais peut être arrivez vous à en comprendre le sens puisque c'est vous qui avez tourné le film. Demandez à un ami de regarder ce film, et peut être le trouvera t'il encore plus désagréable. Et pour cause, l'interprétation du film est d'autant plus compliquée qu'il ne maîtrise pas les mouvements de la caméra, de telle sorte qu'il est très difficile d'en construire une interprétation cohérente. Pour en revenir aux mouvements saccadiques, même si l'analogie entre l'œil et une caméra est douteuse, nous en retiendrons néanmoins cette idée que l'information visuelle traitée par la rétine est brutalement et fréquemment modifiée.

Dans ce chapitre, nous allons dresser un panorama des structures corticales et sous corticales impliquées dans le contrôle oculomoteur dans l'objectif d'en extraire des principes fonctionnels qui permettent de guider le développement des modèles que nous présenterons dans le chapitre 4. Nous mettrons en particulier l'accent sur la manière dont l'information spatiale est représentée dans ces structures, que cette information soit sensorielle ou motrice.

Le contrôle moteur des yeux est assez simple puisque les yeux des primates sont mis en mouvement à l'aide d'uniquement six paires de muscles et des contraintes physiques réduisent le nombre de degrés de liberté d'un œil à deux (à comparer, par exemple, à la complexité des mouvements du bras qui impliquent une vingtaine de degrés de liberté). Malgré la simplicité apparente du contrôle des yeux, un vaste ensemble d'aires corticales et sous corticales sont impliquées. Le rôle spécifique de chacune de ces aires et leurs interactions peuvent être étudiés en combinant des études psychophysiques ou des enregistrements cellulaires à des perturbations temporaires (stimulation électrique ou magnétique, inactivation réversible grâce à l'injection d'agents pharmacologiques) ou définitives (lésion, ablation) des circuits oculomoteurs.

Nous proposons dans une première partie une revue des différents mouvements oculaires observés chez le primate (section 3.1) qui nous permet de situer les saccades oculaires par rapport aux autres mouvements oculaires, d'en présenter les propriétés dynamiques ainsi que quelques protocoles expérimentaux qui permettent d'étudier leur contrôle. Nous commençons ensuite la description du circuit oculomoteur en partant des muscles extraoculaires qui mettent en mouvement les yeux. Nous verrons ensuite que les motoneurones, qui contrôlent les muscles extraoculaires, sont sous le contrôle d'une structure qui code temporellement une saccade oculaire, à savoir

la formation réticulée. Dans la section 3.2, nous présenterons également le colliculus supérieur, une des afférences de la formation réticulée, qui, à la différence de la formation réticulée, code spatialement la cible d'une saccade. Nous aborderons ensuite, dans la section 3.3, les différentes structures corticales dont on connaît, à l'heure actuelle, l'implication dans le contrôle des saccades oculaires. Puisque plusieurs de ces aires se projettent sur le colliculus supérieur ou la formation réticulée, il nous paraît moins clair de continuer la présentation des aires oculomotrices en remontant la hiérarchie de ces aires en partant des structures motrices. C'est la raison pour laquelle, dans la section 3.3, nous partirons de l'organe sensoriel de la vision et étudierons les différentes structures efférentes au cortex visuel. Nous terminerons le chapitre en présentant, en section 3.4, deux mécanismes pour résoudre les problèmes de transformation et de mémorisation d'une information spatiale.

3.1 Les mouvements oculaires

On distingue généralement trois catégories de mouvements oculaires :

- les mouvements de duction ;
- les mouvements de version ;
- les mouvements de vergence.

Les mouvements de duction sont des mouvement monoculaires. Les mouvements de version et de vergence sont des mouvements binoculaires, respectivement conjugués (de même amplitude et de même direction) et disconjugués (d'amplitude différente et de direction opposée). Une seconde classification des mouvements oculaires est une classification fonctionnelle. On distingue les mouvements de fixation, de stabilisation du regard et de déplacement du regard. Bien que cela puisse paraître contradictoire, on parle de mouvement de fixation car, comme nous le verrons par la suite, certaines structures du cerveau sont actives lorsque les yeux doivent rester fixes. Dans la catégorie fonctionnelle des mouvements de stabilisation du regard, on distingue le réflexe vestibulo-oculaire qui permet de compenser la rotation de la tête et le réflexe opto-cinétique qui permet de compenser le déplacement de la scène visuelle. Dans la catégorie fonctionnelle des mouvements d'orientation du regard, on distingue la vergence qui est un mouvement coordonné des deux yeux dans une direction opposée servant à faire converger la fovéa des deux rétines sur un même objet, la poursuite lente qui permet de maintenir la fovéa sur un objet en mouvement, sous la condition que cet objet ne bouge pas trop vite et les saccades qui sont des mouvements rapides, de grande amplitude, qui changent brutalement la position des yeux et qui servent à amener le plus rapidement possible la fovéa des yeux sur un même objet. Ces différents mouvements ne sont pas toujours indépendants et nous verrons par exemple que le nystagmus vestibulo-oculaire est constitué d'un déplacement lent et d'un déplacement rapide des yeux, et que la poursuite lente peut parfois s'accompagner d'une saccade.

3.1.1 Fixation

On parle de mouvements de fixation pour au moins deux raisons : la première est que, comme nous le verrons dans le paragraphe 3.2.1, certaines cellules sont actives lorsque les yeux restent fixes, et la seconde est que ces mouvements dits de fixation semblent avoir un rôle important dans la perception visuelle bien que le lien de causalité ne soit pas encore établi. Nous reviendrons sur ce point à la fin de ce paragraphe.

On distingue trois mouvements de fixation, ou mouvements miniatures : le tremblement, la dérive et la microsaccade [Carpenter, 1988, Yarbus, 1967]. La figure 3.1a permet d'observer expérimentalement certains de ces mouvements de fixation [Verheijen, 1961]. L'expérience consiste à fixer le point noir central pendant une minute, puis à diriger son regard sur le point blanc. Si l'expérience fonctionne, la persistance rétinienne permet de voir des lignes en mouvement continu et éventuellement un déplacement rapide de l'image ; ces deux déplacements correspondent respectivement à la dérive et aux microsaccades. La troisième composante des mouvements miniatures est le tremblement. Ces trois composantes sont représentées sur la figure 3.1b. Le tremblement ne s'observe pas facilement sur la figure puisqu'il est une oscillation très rapide des yeux. La dérive est la variation plus lente de leurs positions tandis qu'une microsaccade, indicée

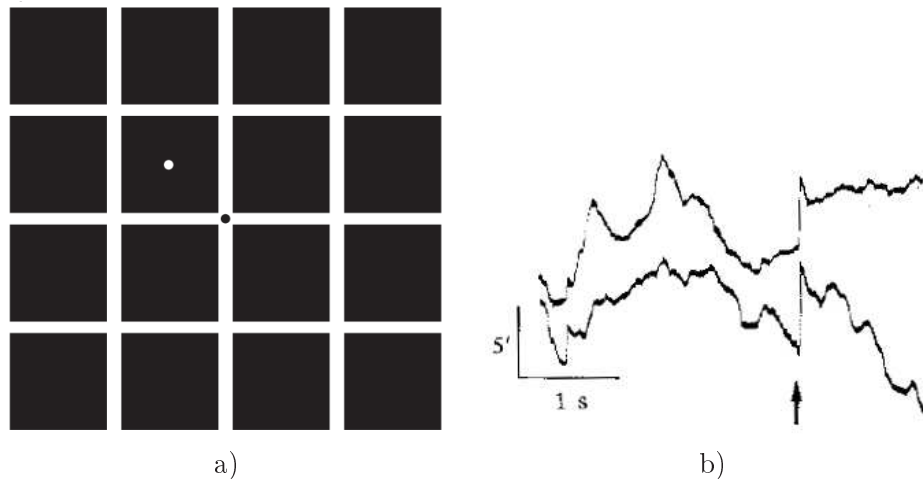


FIG. 3.1 – a) En fixant pendant une minute le point noir puis en regardant le point blanc, la persistance rétinienne permet d’observer les mouvements de dérive et les microsaccades. Adapté de [Verheijen, 1961] b) Illustration des mouvements miniatures des deux yeux. Les tremblements sont des mouvements d’oscillations à haute fréquence que l’on ne distingue pas très bien sur l’illustration. Le mouvement de dérive est l’évolution plus lente de la position des yeux. Une microsaccade, indiquée par une flèche, est un changement brutal de leurs positions. Adapté de [Yarbus, 1967]

par la flèche, correspond à un changement brutal de leurs positions.

Le tremblement est un mouvement très rapide des yeux à une fréquence de l’ordre de 90 Hz et d’amplitude avoisinant le diamètre d’un cône, une des deux catégories de photorécepteurs de la rétine. La dérive, simultanée avec le tremblement, est un mouvement beaucoup plus lent, qui peut être conjugué ou non, dont l’amplitude peut couvrir une dizaine de photorécepteurs. Les périodes de dérive et de tremblement sont parfois séparées par des microsaccades. Ce sont des mouvements rapides, d’une durée d’environ 25ms, d’amplitude allant d’une dizaine à une centaine de photorécepteurs. Au même titre que les saccades de plus grande amplitude, les paramètres des microsaccades (durée, latence, amplitude, vitesse) sont liés par des relations que l’on appelle séquence principale, et dont on reparlera dans le paragraphe 3.1.5 traitant des saccades. On trouvera une revue de plusieurs études sur ces différents mouvements miniatures ainsi que des données quantitatives dans [Martinez-Conde et al., 2004].

Le débat sur le rôle fonctionnel de ces mouvements miniatures n’est pas clos [Martinez-Conde et al., 2004]. Certains auteurs suggèrent qu’ils n’ont aucune utilité [Kowler et Steinman, 1980], d’autres proposent qu’ils ont un rôle fonctionnel bien précis que nous allons expliciter [Ditchburn et al., 1959, Martinez-Conde et al., 2006]. La fovéa est la région où l’acuité visuelle est la plus importante. Différentes études montrent que la vision d’un objet reste nette si sa vitesse n’excède pas 5 degrés par seconde et s’il reste dans une région de 0.5 degrés autour de la fovéa [Burr et Ross, 1982]. Par contre, si on utilise un dispositif expérimental pour fixer l’information rétinienne malgré les mouvements des yeux, elle “s’évanouit” très rapidement [Clarke, 1957, Clarke, 1960]; c’est un phénomène connu sous le nom de *neural adaptation* et qui correspond au fait

qu'une stimulation constante suscite une réponse forte à sa présentation de la part des cellules sensorielles, qui s'atténue si la stimulation perdure. Ainsi, comme le souligne [Martinez-Conde et al., 2004], le système visuel est confronté au paradoxe de devoir maintenir l'information visuelle pour pouvoir la traiter avec le niveau de détail le plus fin tout en s'assurant qu'elle ne reste pas fixe trop longtemps sans quoi elle s'estompe rapidement. D'après les auteurs, le rôle fonctionnel des mouvements miniatures seraient d'éviter que l'image rétinienne ne reste fixe afin de conserver une réponse de la part des cellules visuelles.

3.1.2 Vergence

Avec le développement de la vision frontale, il est devenu possible d'orienter la fovéa des deux yeux sur un même objet, même si certains animaux ayant la vision latéralisée, comme le lapin ou certains oiseaux [Wallman et Pettigrew, 1985, Zuidam et Collewijn, 1979], ont également cette faculté. Les mouvements de vergence sont des mouvements disconjugués des yeux (au contraire par exemple des saccades oculaires) permettant d'amener un même objet sur les deux fovéa. La façon la plus simple de les observer est de fixer un objet qu'on déplace en profondeur.

Il y a principalement deux propriétés visuelles qui conduisent à ces mouvements disconjugués : la disparité entre les images des deux rétines et le flou visuel [Leigh et Zee, 2006]. Les mouvements de vergence liés à la disparité visuelle sont mis en évidence en plaçant un prisme devant l'un des deux yeux, ce qui a pour conséquence de déplacer l'image rétinienne de cet œil, induisant de la disparité entre les deux images rétiniennes. Cette disparité est compensée par un mouvement de l'œil appareillé qui peut être rapide et brutal (comme par exemple lorsque notre regard passe d'un objet proche à un objet lointain), ou lent et continu (comme par exemple lorsque nous fixons un objet qu'on déplace lentement en profondeur). La deuxième propriété utilisée pour contrôler les mouvements de vergence est le flou visuel. En effet, si on dispose un cache devant un œil et que l'on déplace un objet selon l'axe optique de l'œil non caché comme illustré sur la figure 3.2a, on constate deux choses : l'œil non caché bouge très peu (puisque l'objet se déplace dans son axe optique, l'œil n'a pas besoin de bouger pour le maintenir sur la fovéa) et l'œil caché converge si la cible se rapproche (figure 3.2b) [Müller, 1843]. Or, l'œil caché n'apporte aucune information visuelle supplémentaire pour supprimer le flou causé par le rapprochement de la cible donc on peut interpréter son déplacement comme un mouvement réflexe conjoint à l'adaptation de la lentille de l'œil non caché.

Lorsque le déplacement de la cible fixée est prévisible, les mouvements de vergence sont en général réalisés par anticipation [Kumar et al., 2002]. Lorsque le déplacement de la cible n'est pas prévisible, la latence des mouvements de vergence causés par la disparité ou le flou est de l'ordre de 150 ms à 200 ms et leur vitesse maximale est dépendante de l'amplitude du mouvement [Leigh et Zee, 2006]. Dans des conditions naturelles, les mouvements de vergence accompagnent des mouvements de stabilisation (par exemple le réflexe vestibulo-oculaire) ou de déplacement du regard (par exemple une saccade). Une question intéressante concerne alors l'interaction entre les mouvements binoculaires disconjugués de vergence et les mouvements binoculaires conjugués de stabilisation et de déplacement du regard. Par exemple, lorsqu'un mouvement de vergence et une saccade sont combinés, on observe que la saccade est ralentie et que la vergence est accélérée ce qui pourrait souligner une certaine dépendance des circuits impliqués dans le contrôle de la

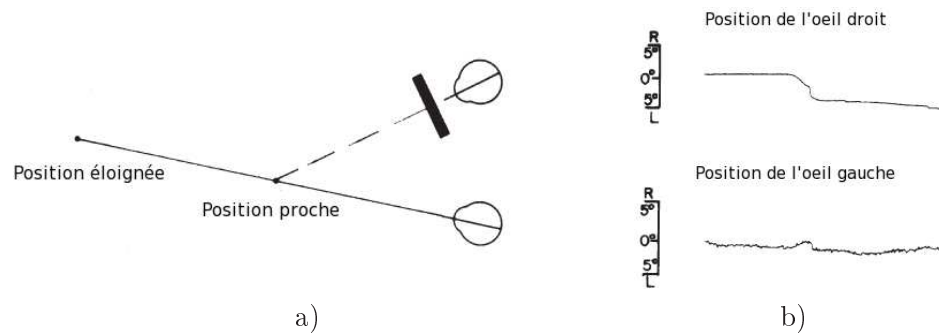


FIG. 3.2 – a) Le protocole expérimental de [Müller, 1843] permettant d’observer la vergence causée par un flou visuel consiste à cacher un œil pendant qu’une cible se déplace le long de l’axe optique de l’œil non caché et à enregistrer leurs positions b) Lorsque la cible se rapproche en passant d’une position éloignée à une position proche, les deux yeux commencent par converger puis l’œil non caché reprend sa position initiale alors que l’œil caché poursuit son mouvement. Adapté de [Leigh et Zee, 2006]

vergence et des saccades [Collewyn et al., 1995, Kumar, 2005, Busetini et Mays, 2005a, Busetini et Mays, 2005b].

3.1.3 Réflexe vestibulo-oculaire

[Carpenter, 1988] distingue deux classes de réflexes vestibulo-oculaires, statiques et dynamiques, selon qu’ils sont produits par une accélération linéaire ou en rotation, stimulant respectivement les otolithes et les canaux semi-circulaires, des organes sensoriels situés dans l’oreille interne. Chez le primate, dont les yeux sont disposés dans le plan frontal, le réflexe dynamique prédomine tandis que pour les animaux dont les yeux sont latéralisés, comme le lapin ou la tortue, on observe à la fois le réflexe statique et le réflexe dynamique.

D’après Carpenter, le réflexe vestibulo-oculaire dynamique sert à aligner la vitesse de rotation des yeux avec celle de la tête, afin de stabiliser l’image rétinienne pendant une rotation de la tête. Ainsi, lorsque la tête tourne dans une direction, à une vitesse donnée, les yeux tournent dans la direction opposée à la même vitesse. Il y a deux façons d’observer ce réflexe : dans des conditions expérimentales ou dans des conditions naturelles. La première, dans des conditions de laboratoire, consiste à mettre en rotation un sujet en le disposant sur un siège tournant. Dans ces conditions, il est possible de tourner la tête (en supposant que la tête est fixe par rapport au corps) d’un angle quelconque. L’œil compense alors dans un premier temps la rotation de la tête. Mais à cause de contraintes physiques, il finit par arriver en butée. Dans ce cas, un rapide mouvement oculaire dans la même direction que la rotation de la tête permet de “réinitialiser” la position de l’œil par rapport à la tête. Le mouvement de l’œil est ainsi une succession de phases lentes, dans la direction opposée de la rotation de la tête, suivies de phases rapides dans la même direction que la rotation de la tête, comme illustré sur la figure 3.3a. C’est à cause de ce mouvement oscillant que le réflexe vestibulo-oculaire est également appelé nystagmus¹¹.

¹¹Nystagmus :n.m. Mouvement d’oscillation involontaire et saccadé du globe oculaire, caractérisé par une suc-

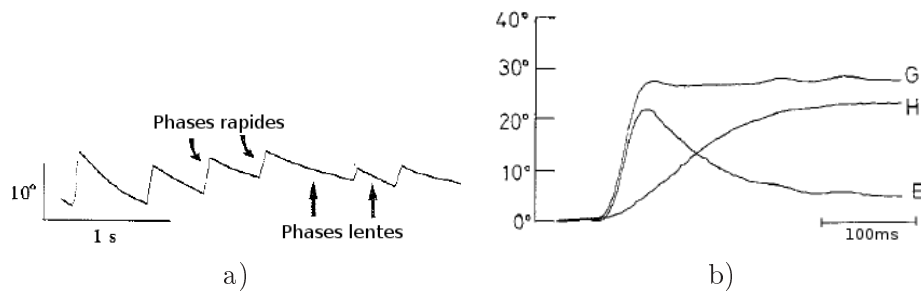


FIG. 3.3 – a) Lorsqu'un primate est mis en rotation sur une chaise tournante, l'accélération en rotation conduit à une succession de phases lentes dans le sens contraire de la rotation suivies de phases rapides dans le sens de la rotation. Cette succession de phases lentes et rapides s'appelle le nystagmus vestibulo-oculaire. b) Les mouvements de la tête qui servent principalement, dans des conditions naturelles, à orienter le regard sont plus lents que les mouvements des yeux. Ainsi, un déplacement rapide du regard est réalisé par un mouvement lent de la tête accompagné d'un mouvement des yeux dans la direction de la cible suivi d'un mouvement des yeux dans la direction opposée. C'est ce mouvement des yeux qui compense le déplacement de la tête que l'on appelle réflexe vestibulo-oculaire. E : position angulaire de l'œil par rapport à la tête, H : position angulaire de la tête par rapport au corps, G : position angulaire du regard, calculée sommant E et H. Adapté de [Carpenter, 1988]

Le réflexe vestibulo-oculaire observé dans le cas de mouvements naturels de la tête est différent de celui observé dans des conditions de laboratoire. En effet, la plupart des mouvements naturels de la tête servent à orienter le regard¹². La vitesse de déplacement de la tête est plus petite que la vitesse de déplacement des yeux étant donnée son inertie. Or, si on considère que le déplacement du regard doit être le plus rapide possible, une solution consiste à déplacer rapidement les yeux sur la cible puis, pendant que la tête est en mouvement, à compenser ce mouvement grâce au réflexe vestibulo-oculaire afin de maintenir la direction du regard fixe. Ce déplacement lent de la tête accompagné d'un mouvement rapide des yeux dans la même direction suivi d'une compensation dans la direction opposée est illustré sur la figure 3.3b où **E** représente la position des yeux, **H** la position de la tête et **G** la position du regard. On observe que la tête se déplace progressivement vers la cible, que les yeux se déplacent rapidement vers la cible puis se déplacent dans le sens opposé afin que le regard soit orienté le plus rapidement possible puis maintenu fixe.

3.1.4 Réflexe opto-cinétique et poursuite lente

Si on dispose un sujet sur une chaise mise en rotation à vitesse constante, l'accélération étant nulle, le réflexe vestibulo-oculaire ne peut plus compenser la rotation de la tête puisqu'il repose sur des informations vestibulaires, fournies par les otolithes et les canaux semi-circulaires qui ne sont sensibles qu'à l'accélération. Dans ce cas, ce sont les informations visuelles, et principalement le flot optique, qui prennent le relais des informations vestibulaires en alimentant le réflexe

cession rythmée de mouvements conjugués changeant alternativement de sens.

¹²Le déplacement du regard est défini par la somme du déplacement de la tête et des yeux.

opto-cinétique. Ce réflexe est observé lorsqu'une grande partie de l'information visuelle est en mouvement, que ce déplacement soit causé par un mouvement de l'animal (disposé sur une chaise mise en rotation) ou par un mouvement propre de la scène (par exemple lorsqu'on observe le paysage par la fenêtre d'un train) au contraire des mouvements de poursuite lente provoqués par le déplacement d'un objet ou d'une petite région de l'espace visuel.

Lorsque les yeux suivent une cible en mouvement, un premier déplacement rapide des yeux amenant la cible sur la fovéa est suivi d'un déplacement plus lent et continu dont la vitesse est très proche de celle de la cible. C'est cette seconde composante du mouvement qui est appelée poursuite lente (*smooth pursuit*). Comme le souligne [Leigh et Zee, 2006], les mouvements de poursuite lente permettent d'atténuer (et d'annuler dans l'idéal) le déplacement d'une cible en mouvement par rapport à la fovéa, en ignorant le déplacement de l'arrière plan. Le déplacement de la cible peut être induit soit par un mouvement propre de la cible [Dodge, 1919], le sujet restant fixe, soit par le mouvement du sujet lui-même [Miles, 1998]. La poursuite lente se distingue du réflexe opto-cinétique parce qu'elle sert à suivre une cible en mouvement, même lorsque l'arrière plan est fixe, tandis que dans ces mêmes conditions, le réflexe opto-cinétique tend à garder les yeux fixes. Elle se distingue également du réflexe vestibulo-oculaire car elle est dépendante des informations visuelles et non pas d'informations vestibulaires. Dit autrement, la poursuite lente repose sur le mouvement local d'une cible tandis que le réflexe opto-cinétique dépend du mouvement global de la scène et le réflexe vestibulo-oculaire dépend d'informations non-visuelles. Plus encore, puisque le réflexe opto-cinétique compense le déplacement d'une scène, il doit être inhibé lors d'un mouvement de poursuite lente, qui induit un déplacement de l'arrière plan opposé au déplacement des yeux [Lindner et Ilg, 2006].

3.1.5 Saccade oculaire

Rôle fonctionnel

Comme nous venons de le voir dans le cas de la poursuite lente, un premier mouvement oculaire rapide amène la fovéa sur la cible. Ce mouvement rapide et brutal est une saccade oculaire. Le rôle fonctionnel des saccades oculaires est d'amener la fovéa sur une cible le plus rapidement possible [Dodge, 1919], en particulier parce que ce mouvement induit un déplacement de l'information visuelle sur la rétine qui rend son traitement difficile. Avec le développement de la fovéa, il est devenu crucial de pouvoir diriger cette région où l'acuité visuelle est maximale sur les objets d'intérêt. On en exécute environ trois par seconde, séparées par des périodes de fixation de l'ordre de 150 à 200 ms. Les saccades peuvent être dirigées vers des objets vus, entendus, mémorisés ou encore dirigées vers des cibles virtuelles (comme dans le cas des tâches d'antisaccade où la cible saccadique est symétrique, par rapport à un point de fixation, du stimulus présenté). Elles peuvent être volontaires (par exemple en faisant partie d'une stratégie de parcours d'une scène visuelle) ou réflexes (par exemple en réponse à l'apparition soudaine d'un stimulus).

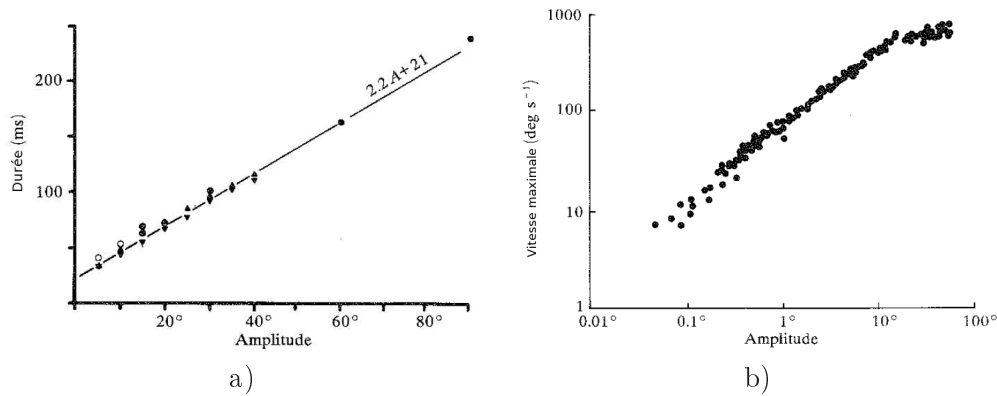


FIG. 3.4 – La séquence principale (*main sequence*), pour les saccades oculaires, est formée de deux relations liant l’amplitude, la durée et la vitesse maximale des saccades oculaires, suffisamment caractéristiques pour classer un mouvement oculaire dans l’une des catégories : saccade normale, saccade anormale, mouvement oculaire non saccadique. a) Relation entre la durée du mouvement et son amplitude. b) Relation entre la vitesse maximale du mouvement et son amplitude. Adapté de [Carpenter, 1988]

Propriétés dynamiques

Les saccades oculaires sont des mouvements stéréotypés. Elles peuvent être décrites par quatre paramètres : l’amplitude, la vitesse maximale, la durée et la latence. Il existe des relations entre l’amplitude, la vitesse et la durée du mouvement, connues sous le nom de “séquence principale” (*main sequence*) [Bahill et al., 1975], un terme emprunté à l’astronomie qui permet dans ce domaine de classer les étoiles en fonction de leur luminosité et de leur température. Pour les mouvements oculaires, ces relations sont suffisamment cohérentes pour constituer un indice permettant de savoir si le mouvement considéré est une saccade normale, anormale ou un mouvement non-saccadique [Leigh et Zee, 2006]. Ces deux relations sont illustrées sur la figure 3.4. En particulier, la durée d’une saccade est proportionnelle à son amplitude (figure 3.4a), et la vitesse maximale est proportionnelle à l’amplitude pour des amplitudes inférieures à 10° puis sature pour des amplitudes supérieures (figure 3.4b).

Le dernier paramètre dont nous n’avons pas encore parlé est la latence, ou temps de réaction, défini par le temps séparant l’ordre de déclencher un mouvement oculaire et le début de la saccade vers cette cible. Ce temps de réaction est très variable et dépend de différents facteurs comme par exemple la modalité de la cible (visuelle, auditive) [Zambarbieri, 2002], ses propriétés physiques (intensité, contraste) [Doma et Hallett, 1988] ou la tâche considérée (saccade réflexe, saccade mémorisée, antisaccade¹³) [Montagnini et Chelazzi, 2005].

¹³Paradigme d’antisaccade : Dans ce paradigme, le sujet doit exécuter une saccade vers une cible imaginaire, située à l’opposé d’une cible visuelle par rapport au point de fixation [Hallett, 1978]

Différents paradigmes pour étudier les saccades oculaires

Différents paradigmes sont utilisés en laboratoire pour étudier le contrôle des saccades et souligner la contribution relative de différentes structures corticales et sous-corticales. La contribution de ces aires est mise en avant en étudiant les conséquences de perturbations, temporaires (e.g. stimulation ou inactivation réversible par injection d'un agent pharmacologique) ou définitives (e.g. lésions), sur la réalisation de ces tâches, ou en enregistrant l'activité de ces aires pendant que le sujet réalise la tâche.

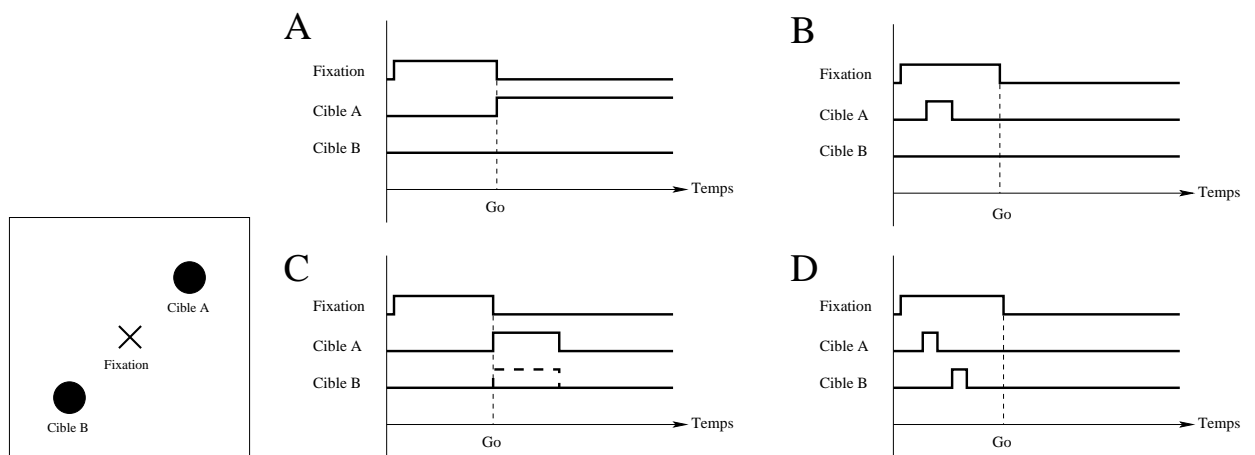


FIG. 3.5 – Gauche : configuration des cibles saccadiques éventuelles et du stimulus de fixation. Droite : Représentation de l'allumage ou de l'extinction des cibles et du stimulus de fixation. L'indication "go" signale quand le sujet doit relâcher la fixation pour exécuter la tâche. a) Dans le cas d'une tâche de saccade simple, une cible (ici la cible A) est allumée au moment de l'extinction de la fixation. Le sujet est récompensé s'il exécute une saccade vers cette cible ; b) Dans le cas d'une tâche de saccade mémorisée, la cible est flashée pendant la fixation. A l'extinction du stimulus de fixation, le sujet doit exécuter une saccade vers la position mémorisée de la cible présentée c) Dans le cas d'une tâche d'anti-saccade, le sujet doit exécuter une saccade vers une cible imaginaire (cible B) opposée à la cible présentée (cible A) ; d) Dans le cas d'une séquence de saccades (double step), plusieurs cibles sont brièvement allumées dans un certain ordre, pendant la fixation (cible A puis cible B). Le sujet doit exécuter les saccades vers les cibles, dans le même ordre qu'à leur présentation.

La figure 3.5 présente certains de ces paradigmes standards dans lesquels on considère un point de fixation et une ou deux cibles (cible A et cible B). Les tâches varient en fonction des séquences d'apparition et d'extinction du point de fixation et des cibles. Dans une tâche de saccade simple (figure 3.5A), le sujet doit exécuter une saccade vers la cible allumée lorsque la fixation est éteinte. Dans une tâche de saccade vers une cible mémorisée (figure 3.5B), le sujet doit garder les yeux fixés sur le point de fixation tant qu'il est allumé. Pendant le temps de fixation, une cible est brièvement allumée en périphérie et le sujet doit faire une saccade vers cette cible mémorisée une fois la fixation éteinte. La tâche d'antisaccade (figure 3.5C) est un peu plus compliquée puisqu'elle nécessite que le sujet exécute une saccade vers une cible virtuelle (cible B) symétrique de la cible visuelle présentée (cible A), lorsque la fixation est éteinte [Hallett, 1978]. Enfin, les tâches de séquence de saccades (figure 3.5D) requièrent de la part du sujet d'exécuter

une séquence de saccades vers les cibles qui ont été présentées pendant la fixation, dans le même ordre que leur présentation (cible A puis cible B) [Hallett et Lightstone, 1976]. Cette dernière tâche nécessite d'une part que le sujet mémorise les cibles et d'autre part qu'il soit capable de prendre en compte son premier mouvement oculaire pour déterminer la position de la deuxième cible, qui sera différente de sa position rétinienne à sa présentation. En effet, si le sujet considère l'erreur motrice pour le second mouvement au moment de la présentation de cette cible, sans la mettre à jour après la première saccade, la séquence ne sera pas reproduite convenablement. Par exemple, dans la configuration illustrée sur la figure 3.5, ne pas être capable de mettre à jour la position de la cible B après la saccade vers la cible A pourrait conduire à une seconde saccade vers le point de fixation.

3.2 Contrôle sous-cortical des saccades oculaires

Les yeux sont mis en mouvement par six paires de muscles antagonistes (trois paires par œil). La figure 3.6 représente les trois paires de muscles extraoculaires de l'œil droit. Les mouvements horizontaux des yeux sont produits principalement par les muscles médial et latéral rectus (**lr** et **mr** sur la figure). Les mouvements verticaux et de torsion sont réalisés par la coopération des muscles rectus supérieur et inférieur et oblique supérieur et inférieur (**sr**, **ir**, **so**, **io** sur la figure). Les mouvements oculaires, qui ne sont ni purement horizontaux, ni purement verticaux, impliquent la contraction de l'ensemble de ces muscles. Théoriquement, il existe une infinité de configurations pour les muscles extraoculaires de telle sorte que le regard soit orienté vers une cible. On imagine facilement que si le regard est orienté vers une cible, il le reste malgré toutes les rotations autour de son axe optique. Or, on constate que la position de l'œil répond à différentes lois et que l'œil n'occupe qu'un sous ensemble de toutes les positions envisageables. En particulier, selon la loi de Donders(1847), si la tête est fixe, la torsion ne dépend que de la position de l'œil dans l'orbite : si l'orientation du regard est fixée, il n'y a qu'une position de l'œil, quelle que soit la trajectoire qu'il a emprunté pour atteindre cette position. La loi de Listing(1854) quant à elle, spécifie quelles rotations sont autorisées étant donnée une position de départ de l'œil : à partir d'une position primaire de l'œil, et étant donné le plan perpendiculaire à la direction du regard que l'on appelle plan de Listing, seules les positions que l'œil peut atteindre par une rotation autour d'un axe appartenant à ce plan sont autorisées.

Les muscles extraoculaires sont innervés par un ensemble de neurones moteurs qui se trouvent au niveau des noyaux prémoteurs du tronc cérébral (région PMN de la figure 3.7). Ces motoneurons sont sous l'influence de la formation réticulée qui produit les signaux permettant de contracter ou de relâcher les muscles (paragraphe 3.2.1). La formation réticulée est elle-même sous l'influence du colliculus supérieur (paragraphe 3.2.2). La formation réticulée code temporellement la cible d'un mouvement oculaire tandis que le colliculus supérieur la code spatialement. Cette section se focalise sur les structures sous-corticales impliquées dans le contrôle des mouvements oculaires, les structures corticales seront abordées dans le paragraphe 3.3. On notera néanmoins que les ganglions de la base, qui sont un ensemble de noyaux sous-corticaux, ne seront abordés qu'à la fin de la section sur les structures corticales, essentiellement parce que nous mettrons l'accent sur leurs interactions avec ces structures corticales.

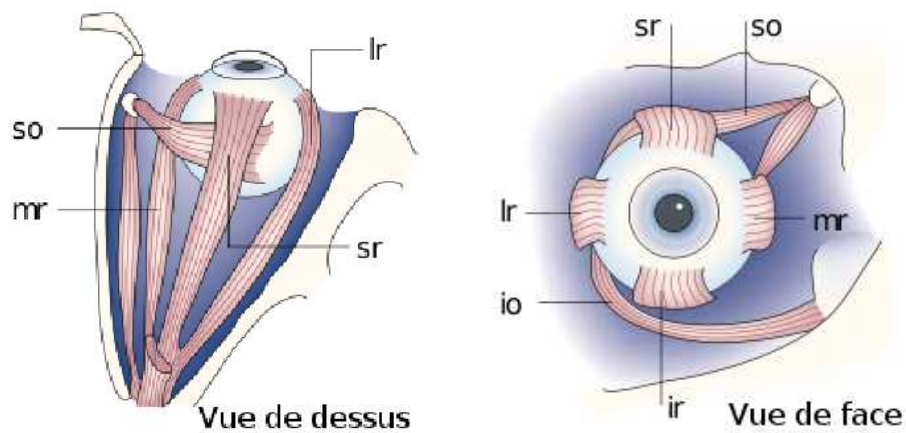


FIG. 3.6 – a) Vue de dessus et de face des trois paires de muscles extraoculaires antagonistes de l'œil droit. Les mouvements horizontaux sont contrôlés par les muscles rectus latéral (lr) et médial (mr). Les mouvements verticaux et de torsion sont réalisés par une combinaison de la contraction des muscles rectus supérieur (sr) et inférieur (ir) et oblique supérieur (so) et inférieur (io). Adapté de [Sparks, 2002]

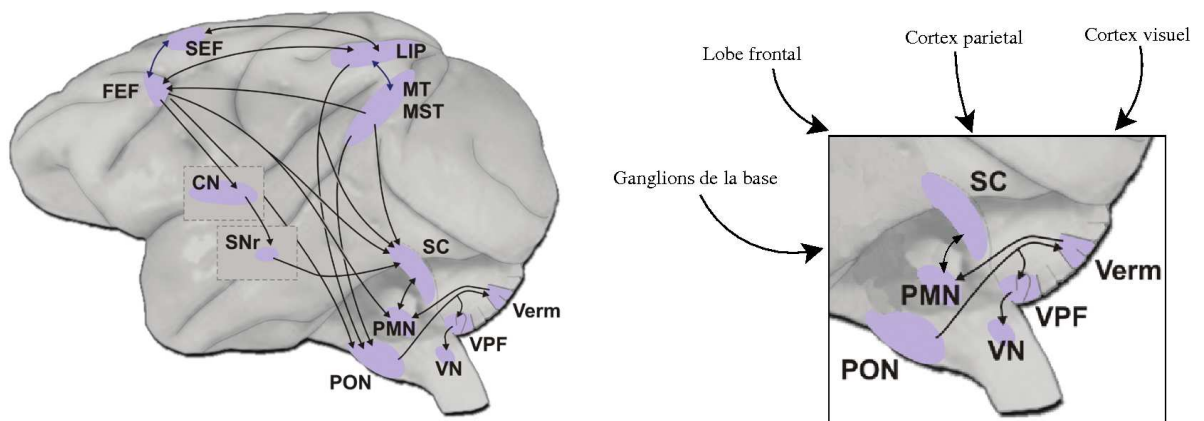


FIG. 3.7 – Gauche : Principales structures corticales, sous corticales et cérébelleuses impliquées dans le contrôle des mouvements saccadiques et de poursuite des yeux. Droite : zoom sur les structures sous-corticales et cérébelleuses. CN=caudate nucleus; FEF=frontal eye field; LIP=lateral intraparietal area; MT=middle temporal area; MST=medial superior temporal area; PMN=brainstem premotor nuclei; PON=precerebellar pontine nuclei; SC=superior colliculus; SEF=supplementary eye field; SNr=substantia nigra pars reticulata; Verm=oculomotor vermis; VN=vestibular nuclei; VPF=ventral paraflocculus. Adapté de [Krauzlis, 2005]

3.2.1 La formation réticulée

Les muscles extraoculaires sont contrôlés par un ensemble de neurones moteurs dont le profil d'activité contient deux principales composantes : une bouffée d'activité phasique ou une pause et une décharge tonique (figure 3.8a). Dans ce qui suit, nous traitons du contrôle des mouvements horizontaux des deux yeux et nous reviendrons à la fin du paragraphe sur le contrôle des mouvements verticaux et obliques. Les muscles horizontaux sont contrôlés par les noyaux abducens. Le muscle agoniste reçoit une décharge d'activité phasique qui permet de mettre en mouvement l'œil en luttant contre les forces de frottement. Cette bouffée d'activité contracte le muscle. Pendant ce temps, les motoneurones contrôlant le muscle antagoniste sont silencieux, permettant de le relâcher. Ce principe de contraction du muscle agoniste et de relâchement du muscle antagoniste avait déjà été proposé par [Sherrington, 1894], connu sous le nom de loi d'innervation réciproque (*law of reciprocal innervation*). A la fin du mouvement, les motoneurones des deux muscles déchargent à une fréquence constante, proportionnelle à la position de l'œil dans l'orbite [Sparks, 2002].

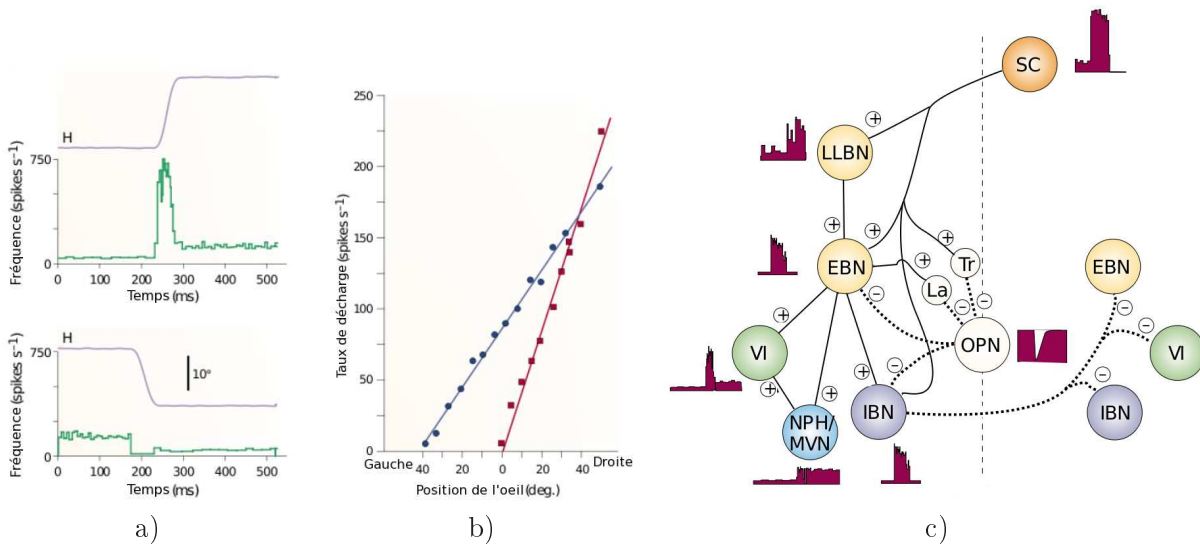


FIG. 3.8 – a, haut) : Profil de décharge du motoneurone contrôlant le muscle agoniste. Une décharge phasique, qui contracte le muscle et met en mouvement l'œil, est suivie d'une décharge tonique proportionnelle à l'excentricité de l'œil. a, bas) : Profil de décharge du motoneurone contrôlant muscle antagoniste. Le motoneurone est d'abord inhibé, ce qui a pour conséquence de relâcher le muscle, puis il décharge à une fréquence constante dépendant de l'excentricité de l'œil. b) La fréquence de décharge de la composante tonique des motoneurones est proportionnelle à l'excentricité de l'œil. c) Illustration des différentes cellules de la formation réticulée *paramedian pontine* qui transforment une commande motrice spatiale envoyée par le colliculus supérieur (SC) en profil temporel phasique/tonique transmis au motoneurone abducens (VI) qui contrôle les muscles horizontaux. SC : superior colliculus ; LLBN : Long lead burst neurons ; EBN : Excitatory burst neurons ; IBN : Inhibitory burst neurons ; VI : noyau abducens ; OPN : Omnipause neurons ; NPH : nucleus prepositus hypoglossi ; MVN : medial vestibular nucleus. Adapté de [Sparks, 2002]

Les noyaux abducens sont sous l'influence d'un ensemble de cellules de la formation réticulée

paramedian pontine qui se distingue de la formation réticulée mésencéphalique qui se projette sur les noyaux innervant les muscles extraoculaires responsables des mouvements verticaux des yeux. Dans la formation réticulée *paramedian pontine*, les neurones responsables de la décharge phasique-tonique, que l'on regroupe sous le nom de générateur saccadique du tronc cérébral (*brainstem saccadic burst generator*) sont répartis en trois classes [Sparks, 2002, Horn, 2005] :

- les neurones omnipause (OPN)
- les *medium lead burst neurons* (MLBN) répartis en neurones excitateurs (*Excitatory Burst Neurons*, EBN) et neurones inhibiteurs (*Inhibitory Burst Neurons*, IBN)
- les *long-lead burst neurons* (LLBN)

La figure 3.8c illustre l'ensemble des cellules de la formation réticulée, leur profil de décharge pendant une saccade ainsi que les projections qu'ils partagent. Les OPN déchargent à une fréquence constante lorsque l'œil doit rester fixe et sont silencieux pendant une saccade, quelle que soit sa direction. Les EBN et LLBN déchargent de manière phasique pour des saccades dans l'hémichamp ipsilatéral¹⁴. L'amplitude, la durée et la vitesse des saccades (ipsilatérales) sont reliées respectivement au nombre de spikes, à la durée et à la fréquence de la bouffée d'activité que les EBN envoient aux motoneurones. Les IBN inhibent les motoneurones contralatéraux afin qu'ils soient silencieux pendant des saccades ipsilatérales. Certains auteurs proposent également que les IBN sont responsables d'une bouffée d'activité envoyée au muscle antagoniste à la fin de la saccade, ce qui permet de le contracter et ainsi de freiner, voire arrêter, le déplacement de l'œil. A titre d'exemple, pour une saccade dirigée vers la gauche, les LLBN, EBN et IBN qui contrôlent le muscle gauche déchargent, mais les OPN sont silencieux. La décharge des LLBN et EBN permet de contracter le muscle gauche tandis que la décharge des IBN inhibe le motoneurone du muscle droit.

Concernant les mouvements obliques des yeux, qui ont une composante verticale et horizontale, les signaux qui déplacent les yeux horizontalement et verticalement ne sont pas indépendants. Si c'était le cas, on devrait observer des trajectoires courbées pour des saccades obliques, or elles sont relativement droites [King et al., 1986]. Les structures qui génèrent les deux composantes sont largement indépendantes à ceci près qu'elles interagissent par l'intermédiaire des neurones omnipause [Curthoys et al., 1984]. On observe alors que la composante la plus courte (en amplitude) dure plus longtemps afin qu'elle dure aussi longtemps que la composante la plus grande (en amplitude) [King et al., 1986, Smit et al., 1990].

Le codage utilisé par les neurones de la formation réticulée pour contrôler le déplacement des yeux est un codage temporel (caractérisé par une bouffée d'activité phasique suivie d'une décharge tonique). Les structures afférentes (par exemple le colliculus supérieur ou les champs oculomoteurs) qui fournissent la cible d'une saccade oculaire la codent spatialement, donnant ainsi lieu au problème connu sous le nom de *Spatial to Temporal Transform Problem (STTP)*, que l'on ne détaillera pas mais dont on trouvera une discussion dans [Sparks, 2004].

Parmi les afférences sous-corticales de la formation réticulée, on trouve principalement le

¹⁴Hémichamp ipsilatéral : Moitié du champ visuel ou moteur du côté du muscle contrôlé par le neurone. Ex : l'hémichamp ipsilatéral d'un EBN ou LLBN contrôlant le muscle gauche d'un œil, contient l'ensemble des saccades dirigées vers la gauche. L'hémichamp opposé est appelé hémichamp contralatéral.

colliculus supérieur qui fournit la cible d'un mouvement oculaire et le cervelet qui semble avoir un rôle dans l'adaptation à long terme.

3.2.2 Le colliculus supérieur

Le colliculus, structure au sommet du mésencéphale, est divisé en deux sous-structures : le colliculus inférieur et le colliculus supérieur (SC, figure 3.7). Le colliculus inférieur est impliqué dans le traitement des informations auditives [Winer et Schreiner, 2005]. Le colliculus supérieur est quant à lui très impliqué dans le contrôle coordonné du mouvement des yeux et de la tête comme le montrent plusieurs études chez des primates dont le colliculus supérieur est stimulé ou inactivé temporairement [Moschovakis et al., 1996, May, 2005]. Il a une structure laminaire que l'on divise généralement en trois couches : superficielles, intermédiaires et profondes. Les couches superficielles sont sensorielles et possèdent des afférences de la rétine et du cortex visuel [May, 2005]. Les neurones de ces couches, appelés cellules visuelles, déchargent lorsqu'une cible est présentée dans une région localisée de l'espace visuel. Cette dernière définit leur champ récepteur, que ce stimulus soit ou non la cible d'un mouvement d'orientation du regard [Mays et Sparks, 1980]. On distingue également une catégorie particulière de cellules visuelles dites quasi-visuelles. Comme les cellules visuelles, ces cellules déchargent lorsqu'un stimulus est présenté dans leur champ récepteur. Mais, contrairement aux cellules visuelles, elles déchargent de manière prédictive lorsqu'une saccade va amener une cible dans leur champ récepteur [Walker et al., 1995]

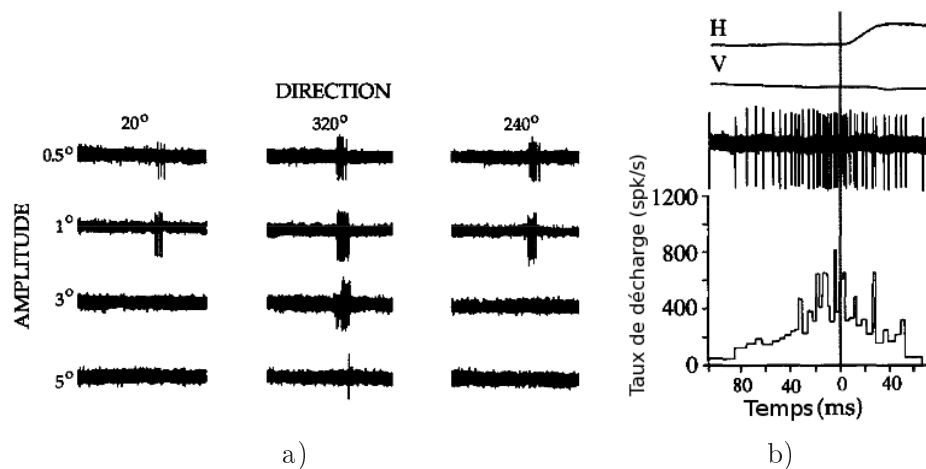


FIG. 3.9 – a) Certaines cellules dont l'activité est corrélée à l'exécution d'une saccade déchargent une bouffée d'activité peu de temps avant un mouvement d'une amplitude et d'une direction données. b) D'autres cellules motrices ont une activité qui croît progressivement jusqu'à l'exécution d'une saccade. Adapté de [Sparks et al., 1976]

Les neurones des couches intermédiaires et profondes peuvent être sensoriels, moteurs ou sensorimoteurs et ont des afférences visuelles, auditives et somatosensorielles¹⁵ [King et Palmer, 1985, May, 2005]. On distingue deux types de cellules motrices : certaines déchargent une bouffée d'activité peu de temps (18 à 20 ms) avant l'exécution d'une saccade (figure 3.9a), d'autres

¹⁵Somatosensoriel : relatif à la perception de stimuli sensoriels sur la peau ou les organes internes.

déchargent progressivement entre l'apparition d'une cible et l'exécution de la saccade qui dirige le regard vers cette cible (figure 3.9b). Ces deux types de cellules sont respectivement appelées neurones *saccade-related* ou *burst* et neurones *prelude* ou *build-up* selon les auteurs [Sparks et al., 1976, Sparks, 1978, Mays et Sparks, 1980, Munoz et Wurtz, 1995a].

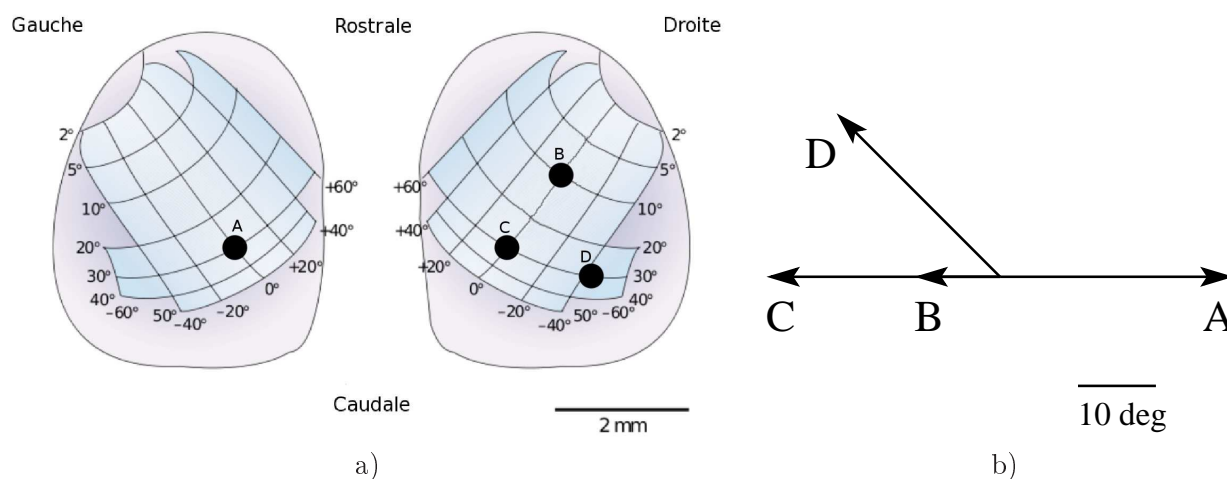


FIG. 3.10 – a) Les cellules motrices des couches profondes du colliculus supérieur sont organisées topographiquement : chaque site colliculaire code une saccade d'une amplitude et d'une direction données. Adapté de [Sparks, 2002] b) Représentation des mouvements saccadiques produits par la microstimulation des sites représentées sur a).

Les études par microstimulation de [Robinson, 1972] révèlent que les cellules motrices déchargent préférentiellement pour une saccade d'une direction et d'une amplitude données, qui définit leur champ de mouvement (*movement field*). Ces études conduisent à définir une carte motrice (figure 3.10) organisée selon un axe rostral-caudal pour des saccades d'amplitude de plus en plus grande. La région rostrale, qui représente les saccades de petite amplitude, est également très active lorsque les yeux doivent être maintenus fixes. Par ailleurs, le mouvement codé par un site du colliculus supérieur profond est indépendant de la position des yeux : chaque neurone code préférentiellement un déplacement désiré plutôt qu'une position de l'œil dans l'orbite. Enfin, [Sparks, 2004] souligne qu'un site colliculaire code une direction et une amplitude maximale, qui n'est atteinte que si la stimulation de ce site est soutenue, une stimulation trop courte évoquant une saccade hypométrique (trop courte).

Certains auteurs suggèrent que l'activité dans le colliculus se propage du site représentant la cible d'une saccade vers la région rostrale, que l'on appelle région de fixation. Selon cette proposition, lorsqu'une saccade est exécutée, une bulle d'activité se forme dans la carte colliculaire au niveau du site codant cette saccade puis se déplace, pendant le mouvement, jusqu'à la région de fixation. Une fois que l'activité a atteint la région de fixation, les yeux sont stoppés puisqu'en particulier, la région de fixation se projette sur les neurones omnipauses de la formation réticulée. Cette hypothèse, qui porte le nom de *moving hill* (bulle glissante), et proposée par [Munoz et al., 1991, Munoz et Wurtz, 1995b], est séduisante mais a été sévèrement critiquée dans des travaux ultérieurs [Sparks, 1993, Soetedjo et al., 2002].

L'une des raisons pour lesquelles cette idée était séduisante est qu'elle proposait une solution pour arrêter la saccade lorsque le regard a atteint sa cible, qui est l'un des aspects du problème plus général de la transformation d'un code spatial vers un code temporel (*Spatial to Temporal Transform Problem*) nécessaire pour contrôler les muscles extraoculaires. Même si le colliculus ne code pas l'erreur motrice dynamique (écart instantané entre la direction du regard et la direction de la cible), nous avons vu précédemment que le site colliculaire excité est caractéristique de la saccade qui va être exécutée. Les structures en aval, en particulier la formation réticulée, doivent être capable de décoder la cible d'un mouvement oculaire afin de produire le codage temporel nécessaire pour contrôler les muscles extraoculaires. Deux hypothèses ont été proposées pour le décodage de l'information colliculaire : par sommation ou par moyennage. Selon la première, la cible d'un mouvement oculaire est décodée par sommation des directions privilégiés des cellules actives du colliculus [vanGisbergen et al., 1987]. Selon la seconde, la cible d'un mouvement oculaire est décodée par une somme pondérée des directions privilégiées [Sparks et al., 1976]. Ces deux alternatives, discutées récemment par [Badler et Keller, 2002], ne sont pas encore départagées.

Nous avons mentionné au début de ce paragraphe que les couches superficielles du colliculus supérieur reçoivent des afférences de la rétine et du cortex visuel. Elles reçoivent également des afférences d'autres structures corticales comme le cortex pariétal et le lobe frontal (en particulier les champs oculomoteurs frontaux, les champs oculomoteurs supplémentaires, le cortex préfrontal dorsolatéral) par des projections directes ou indirectes via les ganglions de la base. Chacune de ces structures, qui contribuent à la sélection d'une cible motrice et à l'exécution d'une saccade oculaire, est discutée dans le paragraphe 3.3.

3.3 Les structures corticales impliquées dans le contrôle des saccades oculaires

Le contrôle des saccades oculaires implique un vaste réseau d'aires corticales, en plus des aires sous-corticales que nous venons de détailler (figure 3.11). D'après [Lynch et Tian, 2005], on dispose d'au moins quatre critères pour déterminer qu'une aire participe au contrôle oculomoteur :

- sa stimulation électrique produit ou modifie un mouvement oculaire
- son activité, obtenue par enregistrement intracellulaire ou par imagerie, est corrélée à l'un des aspects du contrôle oculomoteur (e.g. sélection d'une cible, déclenchement du mouvement oculaire)
- sa lésion ou son inactivation temporaire conduit à des déficits du contrôle des mouvements oculaires
- elle possède des projections directes sur les principales structures oculomotrices du tronc cérébral.

En se limitant aux saccades oculaires suscitées par des stimuli visuels, on distingue notamment l'organe sensoriel, c'est à dire l'œil, dont la rétine est tapissée de photorécepteurs qui convertissent la lumière qu'ils reçoivent en influx nerveux voyageant alors le long du nerf optique pour atteindre le cortex visuel. L'information visuelle est ensuite traitée dans les aires visuelles striées et extrastriées que nous abordons dans le paragraphe 3.3.1. Nous nous limiterons par la suite à

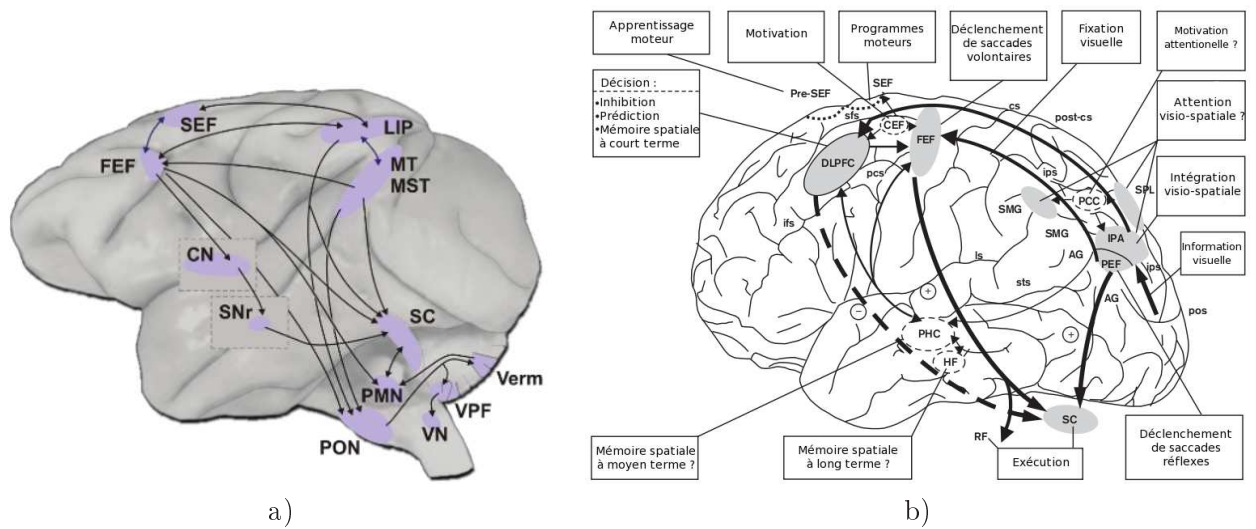


FIG. 3.11 – a) Illustration des structures corticales chez le singe impliquées dans le contrôle des saccades oculaires et des mouvements de poursuite. Adapté de [Krauzlis, 2005]. b) Illustration des structures corticales chez l'homme ainsi que leur rôle respectif dans le contrôle oculomoteur. Adapté de [Pierrot-Deseilligny et al., 2004]. La signification de l'ensemble des acronymes utilisés est disponible dans le glossaire.

L'ensemble des structures corticales qui extraient et transforment l'information visuelle spatiale pour fournir aux structures corticales les informations nécessaires à la production d'une saccade oculaire. En particulier, on discutera dans le paragraphe 3.3.2 de l'implication de deux aires du lobe pariétal (LIP et 7a) qui se projettent directement sur le colliculus supérieur et indirectement via des aires du lobe frontal. Dans le lobe frontal, on distinguera notamment les champs oculomoteurs frontaux (FEF, *Frontal Eye Fields*) et supplémentaires (SEF, *Supplementary Eye Fields*). On explicitera leur rôle dans la sélection "cognitive" (à distinguer des saccades réflexes qui sont supposées faire intervenir des projections plus directes, par exemple du pariétal, sur le colliculus supérieur) d'une cible saccadique (par exemple sur la base d'informations mémorisées), et le cortex préfrontal dorsolatéral dont on connaît l'implication dans la mémorisation d'informations spatiales (paragraphe 3.3.3). Enfin, le dernier paragraphe est consacré aux ganglions de la base dont les interactions avec l'ensemble des aires corticales impliquées dans le contrôle oculomoteur ainsi qu'avec les structures limbiques pourraient lui permettre de contribuer à des processus cognitifs tels que la mémoire de travail ou la sélection de l'action.

On trouvera différentes revues du contrôle oculomoteur en général ou de certains mouvements oculaires dans [Schiller et Tehovnik, 2005, Büttner et Büttner-Ennever, 2005, Lynch et Tian, 2005, Krauzlis, 2005, Goldberg, 2000, Pierrot-Deseilligny et al., 2004, Schall, 1995]

3.3.1 De la rétine au cortex visuel strié et extrastrié

Les primates capturent l'information visuelle grâce à leurs rétines qui occupent le fond des deux yeux, et qui sont tapissées d'un ensemble de photorécepteurs appartenant à deux catégories :

les cônes et les bâtonnets. Les cônes sont eux mêmes divisés en trois catégories en fonction de la longueur d'onde privilégiée de la lumière pour laquelle ils sont le plus sensible. On distingue ainsi les cônes S, M et L dont le nom vient du domaine de longueurs d'ondes auxquelles ils sont sensibles (S pour *Short*, M pour *Medium* et L pour *Long* qui correspondent respectivement à des couleurs proches du bleu, du vert et du jaune, figure 3.12a). Les bâtonnets ne sont quand à eux sensibles qu'à une plage de longueurs d'onde. La répartition de ces photorécepteurs n'est pas uniforme. En particulier, la densité des cônes est très importante dans une région localisée de la rétine, que l'on appelle fovéa, et décroît rapidement avec l'excentricité. L'évolution de la densité des bâtonnets est inverse de celle des cônes : ils sont moins denses à la fovéa qu'en périphérie (figure 3.12b). Ainsi, lorsque vous observez les étoiles dans la nuit noire, il est préférable d'excentrer son regard afin que les étoiles à observer se projettent sur la périphérie de la rétine où les bâtonnets sont plus denses.

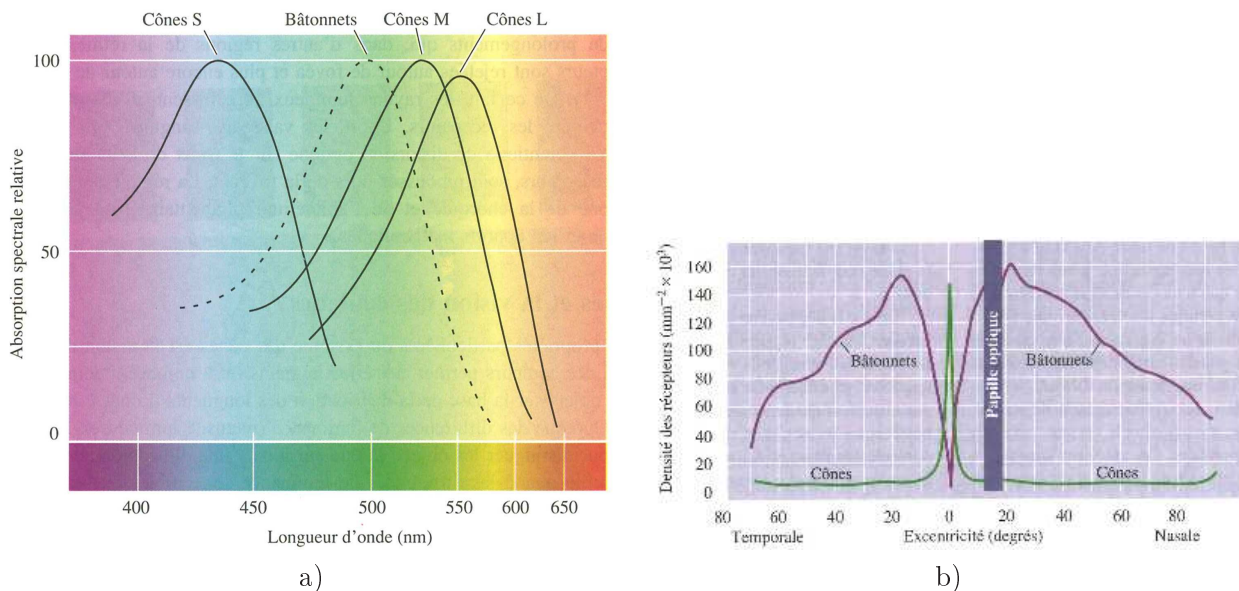


FIG. 3.12 – a) Les photorécepteurs sont divisés en deux catégories, les cônes et les bâtonnets. Les bâtonnets sont sensibles à une longueur d'onde voisine de 500 nm. Les cônes se répartissent en trois catégories, sensibles à des petites, moyennes et grandes longueurs d'ondes. b) La densité des photorécepteurs n'est pas uniforme sur la rétine. La densité des cônes est maximale à la fovéa et décroît rapidement avec l'excentricité. Au contraire, les bâtonnets sont plus denses en périphérie qu'à la fovéa. Adapté de [Purves et al., 2005]

Les photorécepteurs transforment la lumière qu'ils reçoivent en signal électrique à travers une succession de cellules horizontales, bipolaires, amacrines et ganglionnaires (figure 3.13a). Les axones des cellules ganglionnaires se regroupent pour former le nerf optique qui quitte la rétine en une région localisée où l'absence de photorécepteurs définit la tâche aveugle. Les nerfs optiques des deux yeux convergent au niveau du chiasme optique où ils se mélangent afin que l'ensemble des informations visuelles d'un hémichamp se projettent vers le cortex visuel ipsilatéral via les corps genouillés latéraux. C'est à dire que toutes les informations visuelles à gauche des yeux, se projetant sur la région à droite des deux rétines, convergent vers le cortex visuel droit.

En allant des cellules qui tapissent la rétine (photorécepteurs, cellules horizontales, bipolaires, amacrines, ganglionnaires) aux cellules des aires visuelles striées ou extrastriées (figure 3.13b), les champs récepteurs (qu'on définit par la région de l'espace visuel dans laquelle une stimulation lumineuse produit une réponse) modifient leurs propriétés (forme, sélectivité). Ainsi, les cellules horizontales ont des champs récepteurs circulaires pour lesquels on distingue deux régions : le centre et la périphérie. Ces deux régions sont antagonistes et on distingue principalement deux familles de cellules bipolaires :

- les cellules dont le potentiel membranaire augmente lorsque le centre est stimulé et diminue au fur et à mesure que le signal lumineux occupe la périphérie sont appelées ON-center/OFF-surround,
- au contraire, les cellules pour lesquelles le centre est inhibiteur et la périphérie excitatrice sont appelées OFF-center/ON-surround.

Les cellules ganglionnaires, qui sont hiérarchiquement les dernières cellules de la rétine avant le nerf optique, ont également des champs récepteurs centre/périphérie antagonistes dont le taux de décharge varie en fonction de la stimulation lumineuse. On distingue deux catégories de cellules ganglionnaires : magnocellulaires et parvocellulaires [Purves et al., 2005]. La périphérie et le centre des champs récepteurs des cellules magnocellulaires sont alimentés par le même type de cônes de telle sorte que ces cellules ne sont pas sensibles à des différences de longueur d'onde locales. Au contraire, le champ récepteur des cellules parvocellulaires ont un centre et une périphérie alimentée par des cônes différents. Par exemple, les cellules ganglionnaires R+/G- ont un centre excité par une stimulation de grande longueur d'onde (vers le rouge) et une périphérie inhibée par une stimulation de longueur d'onde moyenne (dans les verts).

La ségrégation entre cellules parvocellulaires et magnocellulaires se retrouve au sein des corps genouillés latéraux (LGN, *lateral geniculate nucleus*) sur lesquels les cellules ganglionnaires se projettent (figure 3.13c). On distingue nettement la ségrégation des couches magnocellulaires et parvocellulaires ainsi qu'une alternance entre une couche qui reçoit ses afférences des cellules ganglionnaires de l'œil gauche et une couche qui reçoit ses afférences des cellules ganglionnaires de l'œil droit. La ségrégation des voies magnocellulaires et parvocellulaires, ainsi que du traitement des informations visuelles de l'œil gauche et de l'œil droit (que l'on appelle bandes de dominance oculaire dans le cortex visuel primaire), reste très franche jusqu'au cortex visuel primaire (V1) sur lequel LGN se projette.

Cortex visuel primaire

Dans le cortex visuel primaire (V1), on trouve à la fois des cellules dites "simples" et des cellules "complexes" [Hubel, 1988]. Les cellules simples ont des champs récepteurs plutôt allongés selon un axe, avec un centre et une périphérie antagonistes (figure 3.14a). Ainsi, une cellule simple de centre excitateur et de périphérie inhibitrice décharge préférentiellement pour des stimuli allongés et orientés le long de la région excitatrice de leur champ récepteur. On trouve dans V1 des cellules simples pour toutes les positions du champ visuel et pour toutes les orientations. L'élongation selon un axe privilégié du champ récepteur des cellules simples pourrait s'expliquer par des afférences de cellules centre/périphérie dont les centres sont disposés le long de la bande centrale. Les cellules complexes ont, comme les cellules simples, une région privilégiée de l'espace

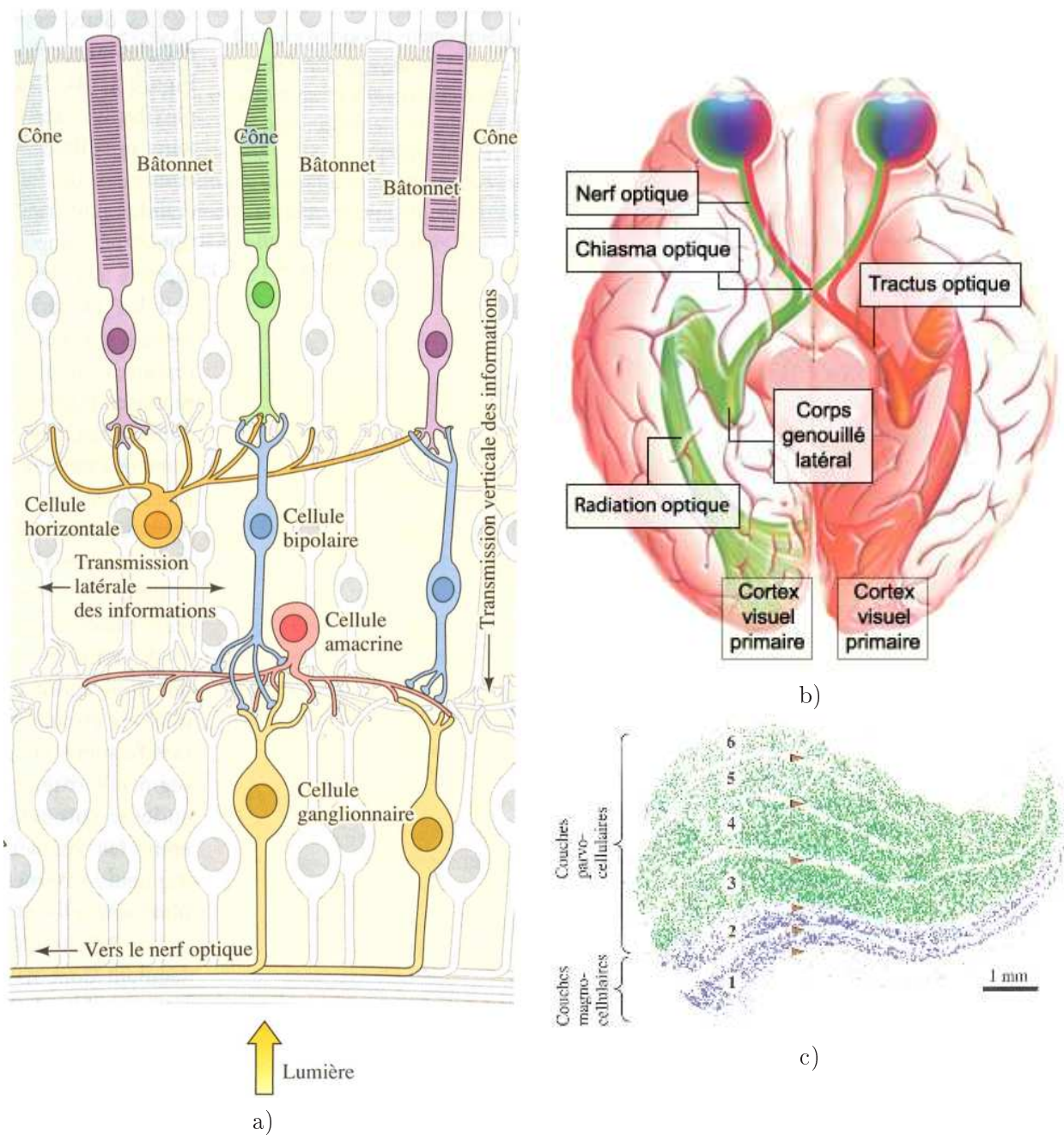


FIG. 3.13 – a) L'information visuelle transmise par les photorécepteurs qui tapissent la rétine est traitée par une succession de cellules horizontales, bipolaires, amacrines et ganglionnaires avant d'être transmise aux structures efférentes via le nerf optique. Adapté de [Purves et al., 2005] b) Les nerfs optiques des deux yeux se rejoignent au niveau du chiasma optique où l'information visuelle des deux yeux est répartie de telle sorte que les informations du champ visuel gauche et du champ visuel droit se projettent respectivement sur le cortex visuel droit et le cortex visuel gauche via les corps genouillés latéraux. c) Les corps genouillés latéraux sont formés d'une succession de couches, ayant alternativement des afférences d'un ou l'autre des deux yeux. On observe également une nette séparation entre les efférences des cellules ganglionnaires magnocellulaires et parvocellulaires. Adapté de <http://lecerveau.mcgill.ca/>

visuel dans laquelle elles répondent préférentiellement qui définit leur champ récepteur. Néanmoins, elles se distinguent des cellules simples par le fait qu'elles répondent à un contraste orienté, quelle que soit sa position dans le champ récepteur, tandis qu'une cellule simple ne répond que si le contraste possède la bonne orientation à une position caractéristique (figure 3.14b). Certaines cellules complexes sont également sélectives à la direction du mouvement. La sélectivité des cellules complexes sur tout leur champ récepteur pourrait s'expliquer par des afférences de cellules simples, de même orientation privilégiée, qui formeraient un pavage de ce champ récepteur.

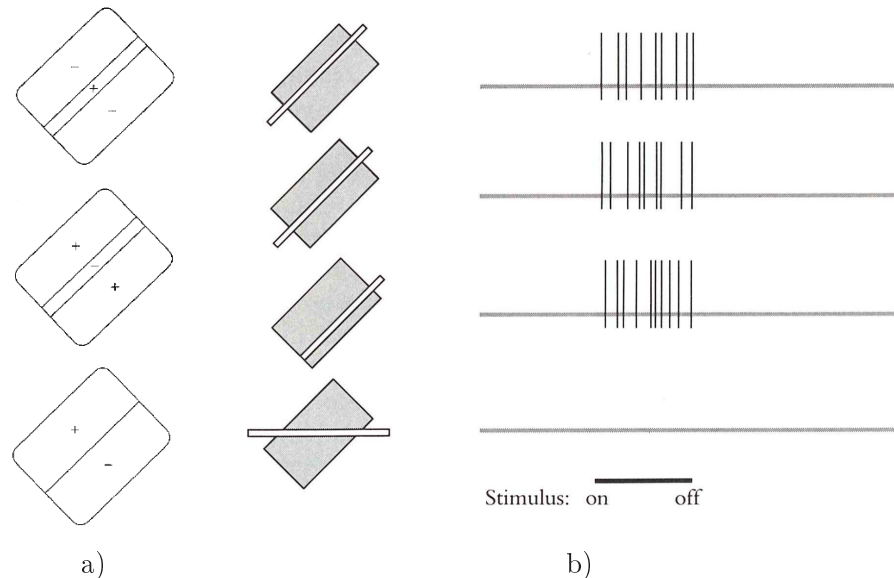


FIG. 3.14 – a) Trois exemples de champ récepteur d'une cellule simple. La cellule décharge d'autant plus que la région excitatrice (+) est illuminée et que la région inhibitrice (-) ne l'est pas b) Pour une cellule complexe, une stimulation longue et fine (rectangle blanc) évoque une réponse forte (le diagramme à droite représente la décharge d'une cellule en fonction de la présence ou non de la stimulation), quelle que soit sa position dans le champ récepteur (rectangle gris) et évoque une réponse faible ou nulle dès lors qu'elle a une orientation non optimale. Adapté de [Hubel, 1988]

Les cellules de V1 sélectives à la couleur se regroupent en taches où on trouve des cellules centre/périphérie antagonistes avec un centre inhibiteur ou excitateur sélectif à un type de cône (à une longueur d'onde) et une périphérie inhibitrice ou excitatrice sélective à un autre type de cône. Il existe par exemple des cellules dont le centre est excité par une stimulation de grande longueur d'onde (vers le rouge) et dont la périphérie est inhibée par une stimulation de moyenne longueur d'onde (vert), qu'on note plus facilement r+g-. En plus des cellules d'opposition simple (e.g. r+g-, r-g+), on trouve également des cellules de double opposition pour lesquelles le centre et la périphérie sont tous les deux sélectifs à des oppositions simples antagonistes. Ainsi V1 contient un ensemble de cellules sélectives à des contours orientés, au mouvement de ces contours ou à des contrastes de couleur. Nous allons voir dans le prochain paragraphe que ces cellules sont organisées selon une topographie particulière : elles sont regroupées de telle sorte que toutes ces propriétés visuelles sont extraites, pour une région de l'espace visuel, par des cellules voisines.

La colonne corticale

Le cortex est une structure laminaire dont l'épaisseur est faible (3-4 mm) relativement à sa superficie (environ 2600 cm²); on peut l'assimiler à une feuille chiffonnée occupant la boîte crânienne en formant des sillons et des circonvolutions. On distingue généralement six couches, distinguées par rapport aux cellules qu'elles contiennent (taille, densité, forme) ainsi que les afférences et efférences qu'elles ont (e.g. avec d'autres structures corticales, avec le thalamus ou des structures sous-corticales). Ces critères permettent ainsi de diviser le cortex en aires cytoarchitectoniques [Brodman, 1909]. Si on enregistre l'activité de cellules du cortex visuel primaire à l'aide d'une électrode qu'on déplace perpendiculairement à sa surface, on observe que les cellules enregistrées répondent à des attributs locaux (contraste orienté de luminosité ou de couleur) identiques. Si on déplace l'électrode tangentielle à la surface du cortex, les attributs préférés évoluent continûment (figure 3.15a). Ces observations conduisent à considérer que le cortex visuel primaire est construit par un pavage de colonnes de cellules, que l'on appelle colonnes corticales (ou microcolonne) [Mountcastle, 1955, Mountcastle, 1957], chaque colonne étant sensible à un attribut visuel particulier qui évolue progressivement pour les colonnes voisines. Chaque colonne corticale, qui contient une centaine de cellules, est sensible aux stimuli présentés dans une région particulière de l'espace visuel qui définit son champ récepteur. L'ensemble des colonnes corticales qui sont sensibles au même champ récepteur forment une macrocolonne. Chaque macrocolonne représente ainsi l'ensemble des valeurs possibles pour un attribut visuel (contraste orienté de luminosité et de couleur) dans son champ récepteur.

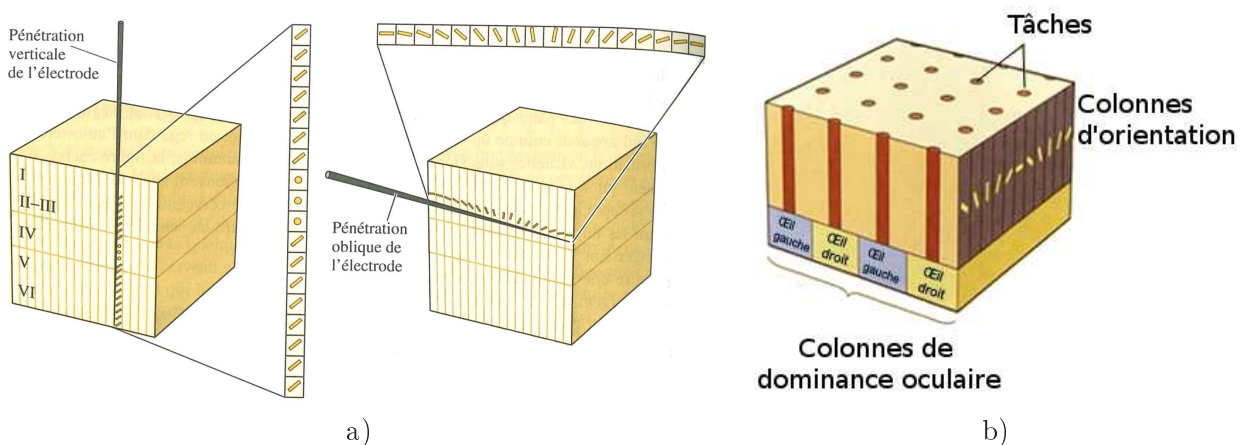


FIG. 3.15 – a) Lorsqu'une électrode est insérée perpendiculairement à la surface du cortex visuel, les cellules enregistrées ont la même sélectivité à un contour orienté. Quand l'électrode est insérée tangentielle, les orientations privilégiées changent au fur et à mesure de la pénétration de l'électrode, et s'inversent de 180° tous les millimètres. b) Les régions du cortex visuel qui traitent les informations de l'œil gauche et de l'œil droit sont alternées et forment des colonnes de dominance oculaire. Elles sont composées de colonnes d'orientation, et de taches. Les colonnes d'orientation sont formées de cellules simples ou complexes sensibles à l'orientation d'un contraste de luminosité, tandis que les taches contiennent des cellules sélectives à la longueur d'onde de la stimulation, donc à sa couleur, indépendamment de son orientation locale. Adapté de [Hubel, 1988]

Du cortex strié aux aires visuelles extrastriées : deux voies de traitement de l'information visuelle

Le cortex visuel primaire, ou aire striée, se projette sur les aires visuelles extrastriées parmi lesquelles on distingue les aires V2, V3, V4, MT(ou V5)/MST. V2 contient des cellules aux propriétés similaires à V1 avec néanmoins des champs récepteurs plus larges, des cellules binoculaires sensibles aux informations des deux yeux et une sélectivité à des formes plus complexes que des simples contours orientés [Hegde et VanEssen, 2000]. L'aire V3 contient des cellules dont les champs récepteurs sont plus larges que ceux de V2 et qui sont sélectives à la couleur, à l'orientation et à la direction du mouvement [Gegenfurtner et al., 1997, Felleman et VanEssen, 1987, Gattass et al., 1988]. Elle se projette principalement sur les aires MT et V4 dont les propriétés neuronales sont distinctes. MT (*Middle Temporal Area*, ou V5) [Dubner et Zeki, 1971] contient des cellules sélectives à la direction du mouvement dans leur champ récepteur ainsi qu'à sa vitesse, quelle que soit la couleur de la stimulation, et se projette sur MST (*Medial Superior Temporal Area*) qui elle-même se projette en particulier sur les aires du sillon intrapariétal. Au contraire, l'aire V4 [Zeki, 1973] contient des cellules sélectives à la couleur et à l'orientation des stimuli dans leur champ récepteur et non à la direction d'un mouvement, et se projette sur le cortex inférotemporal. On observe ainsi une ségrégation des informations visuelles traitées : d'une part des aires visuelles qui traitent progressivement des informations telles que la couleur et l'orientation et d'autre part des aires plutôt impliquées dans l'extraction du mouvement et le traitement des informations spatiales. Ces deux voies formeraient des extensions des voies parvocellulaires et magnocellulaires dont nous avons parlé précédemment. Cette ségrégation a conduit certains auteurs [Ungerleider et Mishkin, 1982, Goodale et Milner, 1992] (voir également [Merigan et Maunsell, 1993] qui nuance cette vision de deux systèmes parallèles indépendants) à proposer un traitement de l'information visuelle selon deux voies. La première, appelée voie ventrale (ou voie *What* dans la littérature anglo-saxonne), regroupe des aires allant du lobe occipital au lobe temporal (V1, V2, V4 et le cortex inférotemporal). Elle est plutôt impliquée dans l'extraction d'informations visuelles permettant la reconnaissance des objets. La seconde, la voie dorsale (ou voie *Where* dans la littérature anglo-saxonne), regroupant des aires allant du lobe occipital au lobe pariétal (V1, V2, V3, MT/MST et le cortex pariétal), est plutôt impliquée dans l'extraction d'informations spatiales (position, direction du mouvement, vitesse du mouvement). Cette distinction est renforcée par des syndromes spécifiques causés par des lésions localisées de l'une ou l'autre des deux voies. En particulier, des lésions du lobe pariétal d'un des hémisphères conduisent, par exemple, au syndrome d'héminégligence ou au syndrome de Balint qui se caractérisent par des déficits visuospatiaux (incapacité de diriger un geste vers une cible visuelle présentée dans le champ opposé au site lésé, négligence d'objets présentés dans le champ visuel opposé au site lésé). Des lésions localisées de la voie ventrale conduisent à des déficits appelés agnosies et qui se traduisent par l'incapacité de reconnaître des objets.

Dans le prochain paragraphe, nous nous concentrons sur la voie ventrale en explicitant les propriétés des cellules du cortex inférotemporal. Le reste du chapitre sera consacré aux aires impliquées dans le traitement de l'information visuelle spatiale conduisant à la sélection d'une cible pour une saccade oculaire, en incluant en particulier les aires de la voie dorsale.

Spécialisation de la voie ventrale dans la reconnaissance des objets visuels

Le cortex temporal inférieur (IT, *Inferior Temporal cortex*) est essentiel pour la perception et la reconnaissance des objets visuels [Gross, 1973] (voir également [Rolls, 1991, Miyashita, 1993, Tanaka, 1996] pour des revues récentes). Il est généralement divisé en deux aires : TEO (ou PIT, *Posterior Inferior Temporal*, figure 3.16) et TE (ou AIT, *Anterior Inferior Temporal*, figure 3.16).

TE reçoit des projections principalement des aires V4 et TEO. Ses neurones ont de larges champs récepteurs, entre 15 et 20 degrés, centrés dans l'hémichamp contralatéral mais qui peuvent contenir la fovéa et une partie de l'hémichamp ipsilatéral. Les neurones de TE sont sélectifs à des formes beaucoup plus complexes que les aires visuelles inférieures, comme par exemple les visages [Rolls, 2000]. L'une des méthodes employées pour étudier leur sélectivité consiste à présenter plusieurs formes afin d'en trouver une qui suscite une réponse forte de la part de la cellule puis à simplifier systématiquement cette forme pour extraire les caractéristiques essentielles des objets qui excitent la cellule [Desimone et al., 1984]. A partir de ces travaux, on observe que les neurones sont sélectifs à des formes géométriques complexes et leur réponse est invariante en translation (en déplaçant le stimulus dans leur large champ récepteur) et à des variations légères de la forme (rotation dans l'espace, occlusion) [Ito et al., 1995, Kovacs et al., 1995]. Les neurones de l'aire TE seraient organisés en colonnes, avec des neurones d'une même colonne sélectifs à une forme particulière et à des variations légères de cette forme (e.g. transformation géométrique, rotation dans l'espace) [Tanaka, 1996, Tsunoda et al., 2001]. Même si les cellules de TE répondent à des formes complexes, l'ensemble de ces formes privilégiées ne regroupent pas l'intégralité des stimuli naturels que l'on peut rencontrer. L'idée qu'il existe, pour chaque objet visuel, un neurone qui décharge maximale et que l'on appelle *cellule grand-mère* [Gross, 2002], est simplificatrice. Il est moins imagé de considérer qu'une population de cellules de TE représente un objet par ses différents attributs, aussi complexes soient-ils.

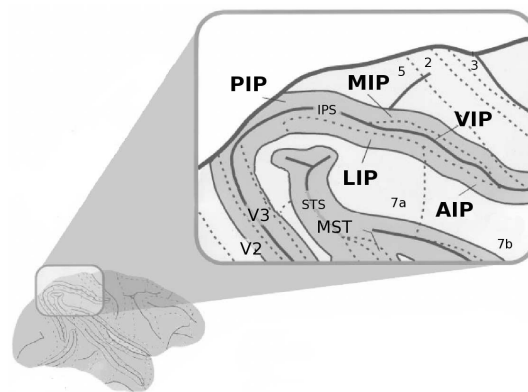


FIG. 3.17 – Illustration des principales aires visuelles avoisinant le sillon intrapariétal. On trouvera plus de détails sur ces aires dans le texte. VIP : Ventral Intraparietal area ; MIP : Medial Intraparietal area ; AIP : Anterior Intraparietal area. PIP : Posterior Intraparietal area ; LIP : Lateral Intraparietal area. Adapté de [Cavada, 2001]

Nous avons vu précédemment que l'information visuelle est traitée le long de deux voies : la voie ventrale plutôt spécialisée dans la reconnaissance des objets et la voie dorsale plutôt spécialisée dans le traitement des informations spatiales. Nous nous sommes concentrés pendant la thèse sur le déploiement de l'attention spatiale et sur l'extraction et la transformation des informations spatiales nécessaires au contrôle oculomoteur (nous reviendrons sur ce point au paragraphe 3.4), en utilisant un module de reconnaissance des objets qui pourrait très schématiquement s'apparenter aux traitements visuels le long de la voie ventrale ; c'est la raison pour laquelle nous ne détaillons pas plus la voie ventrale. On trouvera néanmoins d'avantage d'informations dans les références citées précédemment ainsi que dans les travaux de modélisation de l'équipe de Tomaso Poggio (voir par exemple [Serre et al., 2007]).

3.3.2 Cortex pariétal

Introduction

La première division corticale du cortex pariétal proposée par [Brodmann, 1909] ne contenait que deux aires : l'aire 5 et l'aire 7 séparées par le sillon intrapariétal. Cette division a été affinée par la suite selon des critères anatomiques et fonctionnels et on distingue au moins cinq aires dans le sillon intrapariétal, illustrées sur la figure 3.17 [Cavada, 2001, Lewis et VanEssen, 2000, Colby et Goldberg, 1999] (voir également [Klam, 2003]) :

- VIP (aire intrapariétale ventrale) contient une majorité de neurones bimodaux avec des champs récepteurs visuels et tactiles alignés, centrés sur le visage. Ces neurones déchargent lorsqu'on applique une stimulation tactile sur une région localisée de la face ou lorsqu'une cible visuelle s'approche de cette région [Colby et Duhamel, 1991, Duhamel et al., 1997, Maunsell et vanEssen, 1983]
- MIP (aire intrapariétale médiale) contient des cellules avec des champs récepteurs purement somatosensoriels, purement visuels ou bimodaux répartis des épaules à la main. Les neurones de cette aire semblent signaler lorsqu'une cible peut être atteinte par la main [Colby et Goldberg, 1999, Colby et Duhamel, 1991, Colby et al., 1988]
- AIP (aire intrapariétale antérieure) contient des neurones qui déchargent lorsqu'une cible visuelle peut être manipulée avec la main. Ils sont sélectifs à la forme et à l'orientation des objets et déchargent préférentiellement pour des mouvements spécifiques de la main [Sakata et al., 1995, Gallese et al., 1994]
- PIP (aire intrapariétale postérieure) contient des neurones visuels sélectifs à des formes élémentaires particulières (cylindre, cube, sphère) et participerait avec l'aire AIP, au sein d'un circuit pariéto-frontal, au codage de la position et des caractéristiques spatiales 3D des objets à saisir [Felleman et VanEssen, 1991, Sakata et al., 1997, Rizzolatti et al., 1998]
- LIP (aire intrapariétale latérale) est très impliquée dans la représentation des informations spatiales pour guider des saccades oculaires. Ses neurones ont des réponses visuelles, motrices, anticipatrices et modulées par l'attention [Colby et al., 1996]. Elle se projette sur le colliculus supérieur pour produire des saccades réflexes et elle est fortement interconnectée avec FEF (paragraphe 3.3.3) [Umeno et Goldberg, 1997] pour les saccades plus complexes à élaborer (saccade vers une cible mémorisée, séquence de saccade, choix d'une cible) [Gaymard et al., 2003].

Nous nous limitons par la suite à détailler l'aire LIP puisqu'elle est la principale aire du cortex pariétal à avoir des réponses permettant de guider les saccades oculaires. On pourrait

également mentionner l'aire 7a, adjacente à LIP, dont les activités visuelles, mnésiques et liées aux saccades soulignent son implication dans le contrôle oculomoteur. Néanmoins, LIP se projette plus densément sur FEF et SC que 7a [Andersen et al., 1990a] et les propriétés de ces deux aires sont très similaires. C'est la raison pour laquelle nous ne détaillerons pas plus l'aire 7a et que le prochain paragraphe se concentre exclusivement sur l'aire LIP.

Propriétés physiologiques des neurones de LIP

Les neurones de LIP répondent de manière phasique (la réponse dure moins longtemps que la présentation de la cible) lorsqu'on présente un stimulus stable dans leur champ récepteur, que ce stimulus soit ou non la cible d'un mouvement oculaire ce qui tend à suggérer que leur réponse peut être visuelle sans être motrice. Dans [Colby et al., 1996], le singe exécute plusieurs fois la même saccade vers une cible présentée à la même position. Puis, les essais suivants requièrent de la part du singe d'exécuter cette même saccade mais cette fois, sans présenter la cible visuelle, le signal d'exécution de la saccade étant toujours l'extinction du point de fixation. Les auteurs observent que même dans ce cas, lorsque le singe exécute une saccade apprise sans cible visuelle, des neurones de LIP déchargent, ce qui suggère que les neurones de LIP peuvent avoir une réponse motrice sans réponse visuelle. Par ailleurs, si une cible est présentée brièvement pendant qu'un singe fixe un point de fixation et que la tâche consiste à exécuter une saccade après l'extinction du point de fixation, en direction de la cible présentée (tâche de saccade vers une cible mémorisée, figure 3.5B), on observe que les neurones de LIP déchargent toniquement pendant le délai. Ainsi les neurones de LIP ont des réponses visuelles, motrices, liées à la mémorisation d'une cible et comme nous le verrons par la suite, modulées par la pertinence comportementale des stimuli dans leur champ récepteur.

Nous sommes restés jusqu'à maintenant relativement vague sur les propriétés des champs récepteurs des neurones de LIP. [Andersen et al., 1990b, Andersen et al., 1985] sont probablement les premiers à avoir observé que les neurones de LIP ont des champs récepteurs rétinotopiques modulés par la position de l'œil. La figure 3.18, adaptée de [Andersen et al., 1985], est une illustration de cette propriété appelée champ de gain (*gain field*) dans l'aire 7a, que l'on retrouve également dans LIP [Andersen et al., 1990b]. La figure 3.18a illustre le protocole expérimental utilisé pour observer cette propriété : pendant que le singe maintient son regard, tête fixée, sur un point de fixation \mathbf{f} , on enregistre l'activité d'un neurone lorsqu'on présente un stimulus visuel dont on fait varier la position. La figure 3.18b représente l'activité du neurone en fonction de la position verticale de l'œil \mathbf{ey} et de la position verticale de la cible par rapport à la tête \mathbf{hy} . On observe que le neurone décharge le plus lorsque la cible est environ située à 20° vers le bas et que le singe regarde également à 20° vers le bas. Si on maintient la position de la cible constante par rapport au point de fixation, c'est à dire si on se déplace suivant la droite en pointillée représentée sur la figure, l'activité du neurone décroît progressivement. Cette activité peut être modélisée, d'après les auteurs, par le produit d'un champ récepteur rétinotopique gaussien par une fonction linéaire de la position de l'œil. Dans ce cas, on parle de champ de gains planaire. On notera néanmoins que [Pouget et Sejnowski, 1997] propose que cette propriété de champs de gain pourrait également être modélisée en utilisant une fonction sigmoïdale de la position de l'œil. Ces mêmes auteurs suggèrent que les champs de gains permettent de decoder de la représentation de LIP n'importe quelle combinaison linéaire de la position de la projection d'une cible sur la rétine et de la position de l'œil. Ainsi, les champs de gains pariétaux pourraient par exemple servir de

substrat pour décoder une représentation centrée-tête. Nous reviendrons plus en détails sur ce point dans le paragraphe 3.4 dédié aux transformations sensorimotrices.

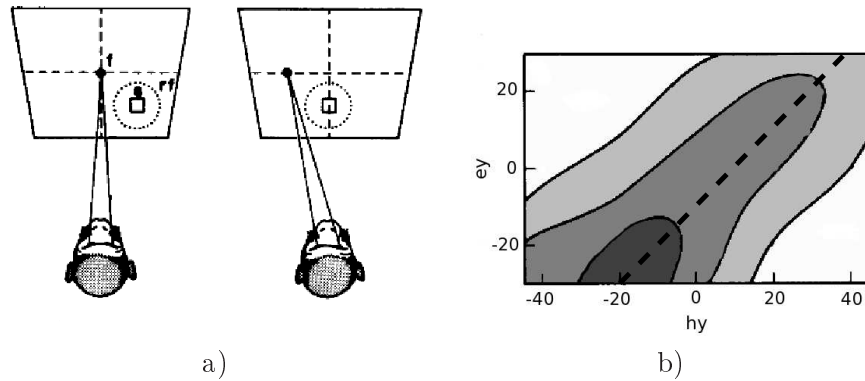


FIG. 3.18 – a) Pour observer les propriétés de champ de gains, on présente à un singe, qui maintient son regard sur le point de fixation f , un stimulus en périphérie. Pendant cette tâche, l'activité des neurones de 7a ou LIP est enregistrée. b) Lorsqu'on trace l'activité (en niveau de gris, du blanc au noir pour des activités croissantes) des cellules enregistrées en fonction de la position verticale de l'œil ey et de la position verticale de la cible par rapport à la tête hy , on observe deux propriétés : les champs récepteurs sont rétinotopiques (il y a une région de l'espace visuel, fixe par rapport à l'œil, dans laquelle les cellules déchargent maximale) et leur activité est modulée par la position de l'œil. Adapté de [Andersen et al., 1985]

Une seconde propriété très intéressante des champs récepteurs des neurones de LIP, observée par [Duhamel et al., 1992, Colby et al., 1995], tient au fait que ces derniers se comportent comme si, quelques millisecondes avant de déclencher une saccade, ils se déplaçaient par anticipation sur leur future position. Dans [Duhamel et al., 1992], le singe doit exécuter une saccade d'une position initiale à une position finale, indiquées par des points lumineux. Pendant la tâche, une cellule de LIP, dont le champ récepteur a été préalablement identifié, est enregistrée. On présente alors un stimulus visuel avant, pendant et/ou après la saccade, dans le champ récepteur de la cellule enregistrée. La figure 3.19 illustre deux des situations étudiées par les auteurs. Sur la figure 3.19a, le stimulus visuel est allumé avant d'exécuter la saccade. Les auteurs observent dans ce cas que la cellule décharge 80 ms avant que la saccade ne soit exécutée. Dans un cas contrôle dans lequel on présente un stimulus dans le champ récepteur de la cellule, sans exécuter de saccade, sa latence est de 70ms. Ainsi, la cellule décharge comme si, peu de temps avant d'exécuter la saccade, son champ récepteur couvrait déjà la future position qu'il va occuper après la saccade. Le cas illustré sur la figure 3.19b est légèrement différent : le stimulus visuel est brièvement allumé à la future position du champ récepteur de la cellule. La présentation du stimulus est tellement brève (il ne dure que 50ms et est éteint au moins 150 ms avant le début de la saccade) que le stimulus n'occupe jamais réellement le champ récepteur classique de la cellule. Néanmoins, les auteurs observent que la cellule décharge même si le stimulus n'a jamais occupé le champ récepteur de la cellule. Les auteurs interprètent ces résultats en suggérant qu'il existerait un mécanisme grâce auquel la représentation spatiale dans LIP est mise à jour, à partir d'une décharge corollaire d'une saccade qui va être exécutée, avant qu'elle ne le soit réellement. Ce mécanisme pourrait contribuer à l'illusion de la perception d'un monde stable malgré les fréquentes saccades oculaires [Merriam et Colby, 2005, Merriam et al., 2007, Nakamura et Colby, 2002].

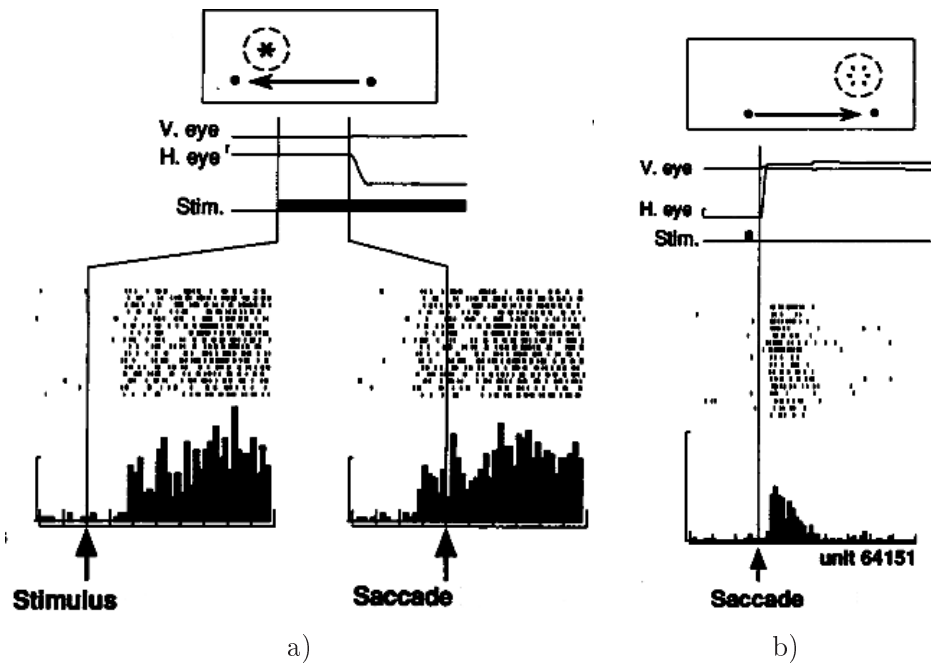


FIG. 3.19 – a) Lorsqu'un stimulus est présenté peu de temps avant d'exécuter une saccade et à la position que va occuper le champ récepteur d'une cellule que l'on enregistre, elle répond par anticipation, de l'ordre de 80 ms avant l'exécution du mouvement. b) Cette réponse par anticipation s'observe également lorsque le stimulus est éteint avant l'exécution du mouvement. Dans ce cas, aucun stimulus n'est jamais présent dans le champ récepteur classique de la cellule. Adapté de [Duhamel et al., 1992]

Interprétation fonctionnelle des activités des neurones de LIP

[Gottlieb et al., 1998] réalisent un ensemble d'expériences montrant que l'activité des neurones de LIP dépend de la pertinence comportementale des stimuli qui occupent leur champ récepteur. Trois des conditions utilisées par les auteurs sont illustrées sur les figures 3.20 et 3.21. Dans le cas illustré sur la figure 3.20, le singe exécute une saccade du point de fixation FP1 vers le point de fixation FP2 (le point de fixation FP2 est allumé en même temps que le point de fixation FP1 est éteint). Dans la première condition *stimulus stable*, la scène visuelle est stable tout au long de l'expérience. Les auteurs observent alors que la réponse de la cellule enregistrée est faible comparée au cas contrôle (non représenté) où un stimulus est présenté brutalement dans le champ récepteur de la cellule. Dans la deuxième condition *stimulus apparu récemment*, la cible qui va occuper le champ récepteur de la cellule enregistrée n'est allumée que quelques millisecondes avant que le singe n'exécute la saccade vers FP2. Dans ce cas, les auteurs observent deux choses : d'une part la réponse de la cellule a une latence plus faible que dans le cas contrôle, en accord avec les réponses par anticipation évoquées dans le paragraphe précédent. D'autre part, la réponse est beaucoup plus forte que dans la condition *stimulus stable*. Les résultats de ces expériences suggèrent que la nouveauté est un facteur important pour l'activité des cellules de LIP.

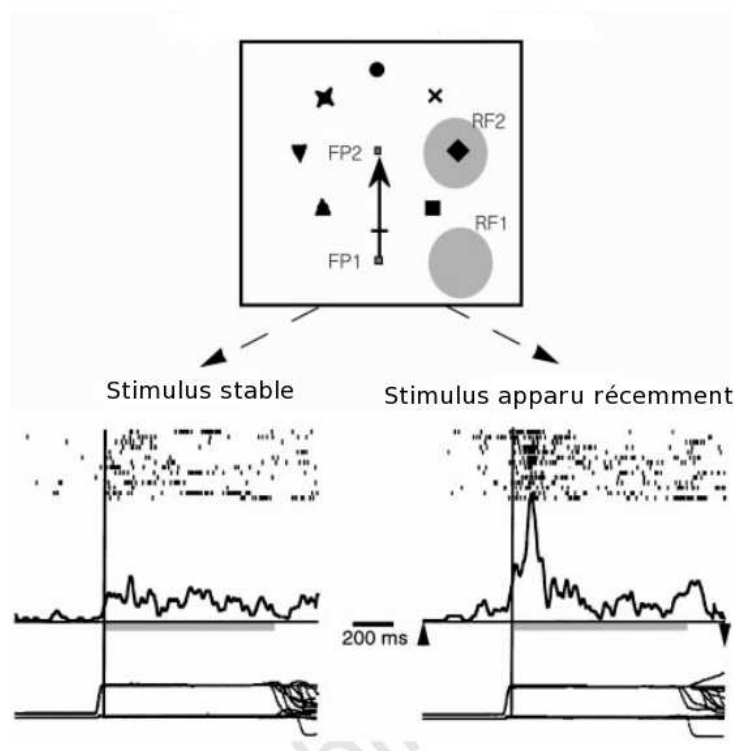


FIG. 3.20 – Lorsqu’une saccade amène un stimulus dans le champ récepteur d’une cellule dont on enregistre l’activité, cette cellule ne décharge que faiblement si ce stimulus était présent bien avant d’exécuter la saccade (condition *stimulus stable*). Elle décharge au contraire fortement lorsque le stimulus apparaît peu de temps avant d’exécuter une saccade qui amène le champ récepteur d’une cellule sur celui-ci (condition *stimulus apparu récemment*). Adapté de [Gottlieb et al., 1998]

La nouveauté est un facteur qui peut capturer l’attention visuelle comme nous l’avons mentionné au chapitre 1. Nous avons également vu que les buts internes constituent un autre facteur permettant de guider l’attention puisqu’ils modifient la pertinence comportementale des stimuli. Nous allons voir qu’il en est de même pour l’activité de LIP.

Dans une deuxième série d’expériences (figure 3.21), les auteurs fournissent au singe un indice (en haut à droite du point de fixation FP) qui lui indique quelle séquence de saccades exécuter. La première saccade est toujours la même : le singe doit déplacer son regard au centre de la scène visuelle. L’indice détermine la cible (unique) de la seconde saccade. Les deux conditions illustrées sur la figure 3.21 requièrent de la part du singe d’exécuter les séquences de saccades suivantes :

- sur la figure 3.21a : une saccade au centre de la scène puis une saccade vers la cible qui occupe le champ récepteur de la cellule enregistrée ;
- sur la figure 3.21b : une saccade au centre de la scène puis une saccade vers une cible qui n’occupe pas le champ récepteur de la cellule enregistrée.

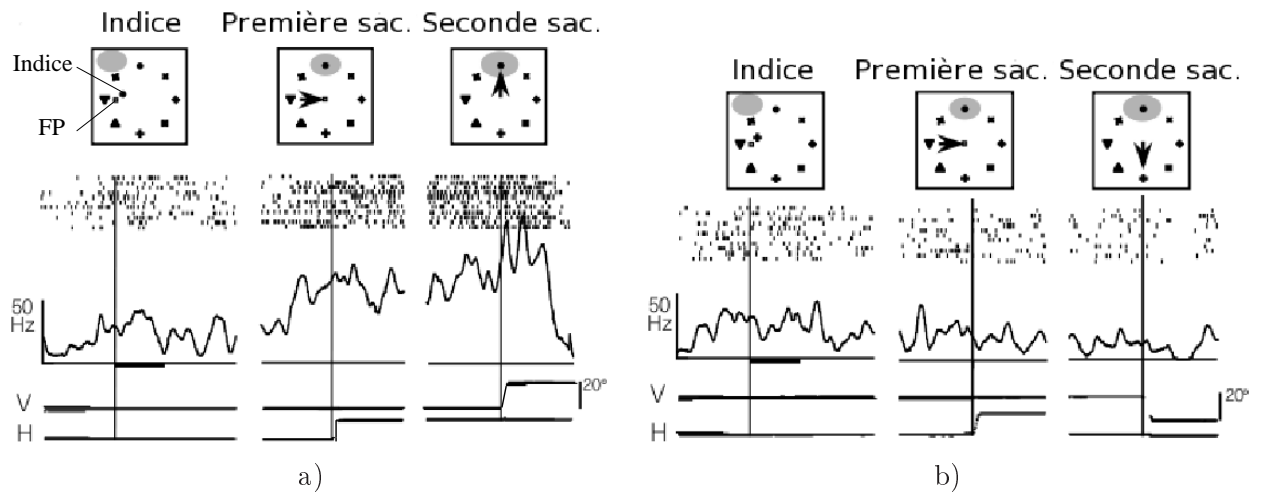


FIG. 3.21 – Le singe fixe initialement le point de fixation FP. Il doit exécuter deux saccades, une fois que le point de fixation est éteint : une première vers le centre de l’affichage puis une seconde vers une cible, définie de manière unique par l’indice présenté pendant la fixation. La cellule décharge plus fortement lorsque la cible amenée dans le champ récepteur de la cellule est pertinente pour la tâche (figure a) que lorsqu’elle ne l’est pas (figure b). Adapté de [Gottlieb et al., 1998]

Les auteurs observent que lorsque le stimulus qui occupe le champ récepteur de la cellule enregistrée sera la cible de la seconde saccade, la cellule décharge beaucoup plus, même avant la première saccade (c’est à dire avant la saccade qui amène le stimulus dans son champ récepteur), comparé au cas où le stimulus amené dans le champ récepteur de la cellule n’est pas la cible de la seconde saccade. Cette expérience permet de souligner que l’activité des cellules enregistrées n’est pas purement visuelle mais ne permet pas encore de distinguer entre une composante motrice et une composante attentionnelle. En effet, la différence d’activité pourrait s’expliquer par la préparation du singe à exécuter la saccade vers la cible. Pour distinguer une réponse attentionnelle d’une réponse purement motrice, les auteurs proposent une deuxième expérience qui consiste à présenter alternativement au singe un affichage qui contient la cible et un affichage qui ne la contient pas, le singe étant récompensé s’il réalise la même saccade dans les deux cas. Dans cette expérience, les auteurs observent que la cellule décharge lorsque le stimulus dans son champ récepteur est présent et qu’il est la cible d’une saccade mais ne décharge que faiblement lorsque le stimulus n’est pas présent, même si la saccade est dirigée vers son champ récepteur. La réponse de ces cellules n’est donc pas purement motrice mais visuelle et modulée par la pertinence comportementale du stimulus qui occupe leur champ récepteur. Tout ces résultats suggèrent que l’activité des neurones de LIP est fortement dépendante de la pertinence comportementale des stimuli qui occupent leur champ récepteur. Elle est définie, comme dans le cas de l’attention visuelle, par des facteurs exogènes (comme la nouveauté) ou des facteurs endogènes (comme la cible d’une saccade définie par ses attributs). Cette interprétation de l’activité des neurones de LIP est très semblable à la notion de carte de saillance, définie comme une représentation spatiale de la pertinence comportementale des stimuli, dont nous avons parlé au paragraphe 1.4 du chapitre sur l’attention visuelle.

3.3.3 Cortex frontal

Champs oculomoteurs frontaux (FEF)

On connaît l'implication dans le contrôle oculomoteur des champs oculomoteurs frontaux (FEF, *Frontal Eye Field*), localisés sur le banc antérieur du sillon arqué, depuis les travaux fondateurs par stimulation électrique de David Ferrier [Ferrier, 1875]. Des études ultérieures, par lésion ou inactivation temporaire, montrent que FEF n'est pas essentiel pour produire des saccades réflexes mais critique pour des saccades plus cognitives impliquant par exemple la mémorisation d'une cible, le choix récompensé ou l'inhibition d'une saccade [Deng et al., 1986, Pierrot-Deseilligny et al., 1991a, Pierrot-Deseilligny et al., 2002, Schiller et Chou, 1998, Dias et Segraves, 1999]. Ces résultats sont à comparer aux lésions ou inactivations temporaires du pariétal postérieur qui ont des conséquences beaucoup plus significatives sur les saccades réflexes aussi bien chez le singe [Lynch et McLaren, 1989] que chez l'homme [Pierrot-Deseilligny et al., 1991a, Braun et al., 1992].

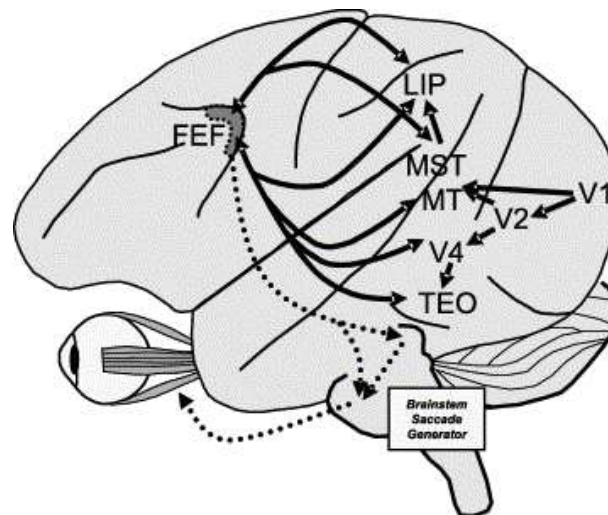


FIG. 3.22 – FEF est interconnecté avec les aires visuelles postérieures impliquées dans le traitement de l'information spatiale (LIP), le traitement du mouvement (MT, MST) et la reconnaissance des objets visuels (V4, TEO). Il se projette sur les structures du tronc cérébral impliquées dans l'exécution d'une saccade. Les projections impliquant des aires frontales (SEF, dlPFC) ne sont pas représentées ainsi que l'interaction de FEF avec les ganglions de la base. Adapté de [Thompson et Bichot, 2005]

Les principales afférences de FEF proviennent des aires visuelles incluant V2, V3, V4, IT, MT et MST et de LIP : 50 % des afférences de FEF proviennent des aires visuelles associatives et 23% de LIP (figure 3.22) [Barbas et Mesulam, 1981, Schall, 1995]. Les aires V2, V3 et V4 sont spécialisées dans l'extraction de caractéristiques visuelles, IT contient des représentations visuelles plus complexes que les aires visuelles striées et extrastriées et enfin MT et MST sont des aires très impliquées dans l'extraction du mouvement visuel et pourraient fournir à FEF les signaux nécessaires à l'exécution de mouvement de poursuite [Krauzlis, 2005]. FEF reçoit également des projections d'aires frontales comme les champs oculomoteurs supplémentaires (SEF) [Schall

et al., 1993] et le cortex préfrontal dorsolatéral (dlPFC) [Tian et Lynch, 1996]. Comme nous le verrons par la suite, SEF est supposé être impliqué dans des tâches saccadiques complexes (séquences de saccades, antisaccade) [Isoda et Tanji, 2002] et ses cellules ont des activités corrélées à la prédiction et l'obtention d'une récompense [Amador et al., 2000]. dlPFC est quant à lui très impliqué dans des tâches de mémorisation, dans l'inhibition de saccades non désirées et, plus généralement, dans la régulation de commandes motrices [Goldman-Rakic, 1987].

La plupart des structures afférentes de FEF reçoivent en retour des projections de cette aire. [Stanton et al., 1995] montrent par exemple que FEF se projette massivement sur LIP et [Moore et Armstrong, 2003] montre l'influence d'une microstimulation de FEF sur l'activité V4 et le rôle potentiel de ces projections dans le déploiement de l'attention spatiale. FEF se projette également sur la formation réticulée [Stanton et al., 1988b, Segraves, 1992], le colliculus supérieur [Segraves et Goldberg, 1987, Schlag-Rey et al., 1992] et les ganglions de la base [Stanton et al., 1988a, Parthasarathy et al., 1992]. Les projections sur la formation réticulée expliqueraient pourquoi il est encore possible de susciter des saccades oculaires en stimulant électriquement FEF malgré des lésions du colliculus supérieur [Schiller et al., 1979] (voir également les travaux de [Hanes et Wurtz, 2001] qui comparent les conséquences de léser ou d'inactiver temporairement SC sur les projections entre FEF et la formation réticulée). Les projections sur le colliculus supérieur visent les couches intermédiaires et profondes qui regroupent les cellules motrices de SC, et sont organisées topographiquement. En d'autres termes, une région de FEF, dont la stimulation électrique produit une saccade d'une amplitude et d'une direction données, se projette sur une région de SC dont la stimulation électrique produit une saccade de la même amplitude et de la même direction [Sommer et Wurtz, 2000, Hanes et Wurtz, 2001]. Enfin, FEF se projette également sur les ganglions de la base. Tandis que les projections sur le colliculus supérieur contribueraient à fournir les cibles d'un mouvement oculaire, les projections via les ganglions de la base pourraient contribuer à la sélection de la cible la plus pertinente et à la détermination du moment de son exécution [Segraves et Goldberg, 1987, Stanton et al., 1988a, Parthasarathy et al., 1992].

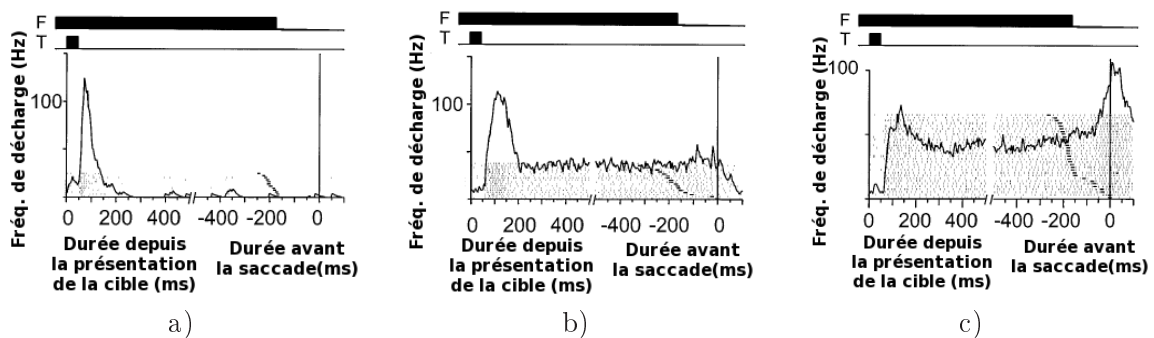


FIG. 3.23 – Quelques exemples d'activités de cellules de FEF. a) Une cellule visuelle phasique qui décharge à la présentation d'un stimulus dans son champ récepteur puis est silencieuse b) Une cellule visuelle tonique qui décharge une bouffée d'activité à la présentation d'un stimulus dans son champ récepteur suivie d'une décharge tonique pendant le délai précédant l'exécution d'une saccade vers cette cible mémorisée c) Une cellule dont l'activité croît significativement à l'exécution d'une saccade vers une cible mémorisée. Adapté de [Hanes et al., 1998]

[Bruce et Goldberg, 1985, Bruce et al., 1985] ont montré que les champs oculomoteurs frontaux contiennent un ensemble de cellules ayant des réponses visuelles, motrices présaccadiques ou visiomotrices. Les cellules visuelles déchargent lorsqu'un stimulus visuel est présenté dans une région localisée de l'espace visuel, qui définit son champ récepteur. Les cellules motrices déchargent peu de temps avant l'exécution d'une saccade d'une amplitude et d'une direction données. Cette région de l'espace moteur définit leur champ de mouvement. Par ailleurs ces cellules sont organisées topographiquement : la région ventrale de FEF contient des cellules motrices qui représentent des saccades de petite amplitude tandis que les saccades de grande amplitude sont représentées par des cellules de la région dorsale. Les cellules visiomotrices ont les deux types de réponses et déchargent à la fois à la présentation d'une cible dans leur champ récepteur et à l'exécution d'une saccade dans leur champ de mouvement. Enfin, certaines cellules de FEF exhibent également des activités liées à la mémorisation d'une cible [Umeno et Goldberg, 2001, Lawrence et al., 2005] et à la fixation [Hanes et al., 1998]. La figure 3.23 illustre certaines de ces réponses. En particulier, la figure 3.23a illustre l'activité d'une cellule visuelle phasique qui répond à la présentation d'une cible et reste silencieuse par la suite. La figure 3.23b représente l'activité d'une cellule visuelle qui a une composante phasique à la présentation de la cible puis tonique pendant le délai précédant l'exécution d'une saccade vers la cible. Enfin, la figure 3.23c illustre l'activité d'une cellule motrice qui croît significativement au moment d'exécuter la saccade.

Certaines cellules ayant une réponse visuelle répondent également par anticipation, lorsqu'une saccade va amener un stimulus dans leur champ récepteur visuel ; c'est la raison pour laquelle ces cellules sont appelées quasivisuelles [Umeno et Goldberg, 1997], au même titre que les cellules prédictives observées dans le colliculus supérieur [Mays et Sparks, 1980] ou dans le cortex pariétal [Duhamel et al., 1992]. La figure 3.24a correspond au paradigme expérimental de double saccade dans lequel on flashe brièvement une cible (carré rouge) permettant d'observer les activités quasivisuelles. Comme pour une tâche de double saccade classique (figure 3.5d), le singe doit réaliser deux saccades successives en direction de T1 et T2 présentées séquentiellement. L'astuce, pour observer les réponses quasivisuelles, consiste à flasher brièvement une cible (carré rouge) dans le champ récepteur visuel (en pointillés sur la figure) courant (indiqué par RF) ou futur (indiqué par FF). Le futur champ récepteur correspond à la région de l'espace qui sera couverte par le champ récepteur de la cellule après la saccade oculaire. On distingue quatre cas :

- conditions RF, haut : la cible est présentée dans le champ récepteur courant, pendant la fixation et avant que la cible T1 de la première saccade n'apparaisse ;
- conditions RF, bas : la cible est présentée dans le champ récepteur courant, alors que la cible T1 est allumée mais que le singe n'a pas encore exécuté la première saccade ;
- conditions FF, haut : la cible est présentée dans le futur champ récepteur, pendant la fixation et avant que la cible T1 ne soit allumée ;
- conditions FF, bas : la cible est présentée dans le futur champ récepteur, alors que la cible T1 est allumée mais que le singe n'a pas encore exécuté la première saccade.

Les enregistrements de l'activité d'une cellule de FEF dans ces différentes conditions, adaptés de [Sommer et Wurtz, 2006], sont présentés sur la figure 3.24b et sont organisés dans le même ordre que les paradigmes expérimentaux. On observe que la cellule répond lorsqu'on présente un stimulus dans son champ récepteur "suffisamment" tôt avant l'exécution de la saccade (voir les deux illustrations en haut de la figure 3.24b) mais que peu de temps avant d'exécuter la saccade, le neurone est insensible à un stimulus présenté dans son champ récepteur courant. Certaines cellules exhibent une propriété intéressante : elles sont sensibles à leur futur champ récepteur peu de

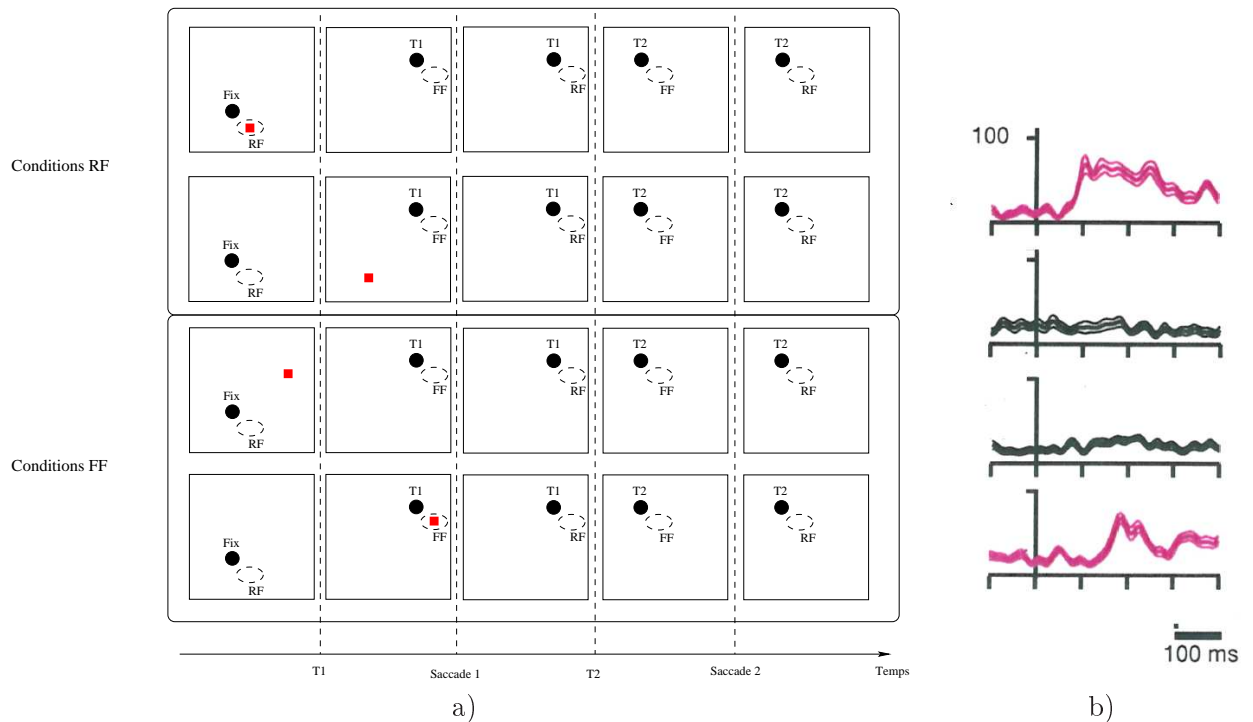


FIG. 3.24 – a) Pour faire apparaître les propriétés des cellules quasivisuelles, on entraîne un singe à exécuter deux saccades successives (vers T1 puis vers T2) pendant qu'un stimulus est présenté dans le champ récepteur (en pointillés) courant (RF) ou futur (FF). La seconde saccade sert exclusivement à dissuader le singe d'exécuter une saccade vers le stimulus présenté. Quatre conditions sont testées en fonction que le stimulus périphérique est présenté pendant la fixation ou peu de temps avant d'exécuter une saccade, et en fonction de la position à laquelle le stimulus est présenté (dans le champ récepteur courant ou futur) b) La cellule enregistrée est sensible à son champ récepteur courant pendant la fixation (première figure) mais ne l'est plus peu de temps avant d'exécuter une saccade (deuxième figure). Elle devient également sensible à son futur champ récepteur peu de temps avant d'exécuter une saccade (quatrième figure) alors qu'elle ne l'est pas pendant la fixation (troisième figure). Adapté de [Sommer et Wurtz, 2006]

temps avant d'exécuter une saccade mais demeurent également sensibles à leur champ récepteur courant. De plus, la cellule décharge lorsqu'un stimulus est présenté dans son futur champ récepteur, peu de temps avant d'exécuter la saccade. Cette réponse est d'autant plus surprenante qu'aucun stimulus n'est jamais présent dans son champ récepteur classique étant donné que le stimulus est éteint lorsque le singe exécute la saccade. La cellule se comporte comme si, au moment d'exécuter une saccade, son champ récepteur se déplaçait de manière prédictive de sa position courante à la future position qu'il va occuper.

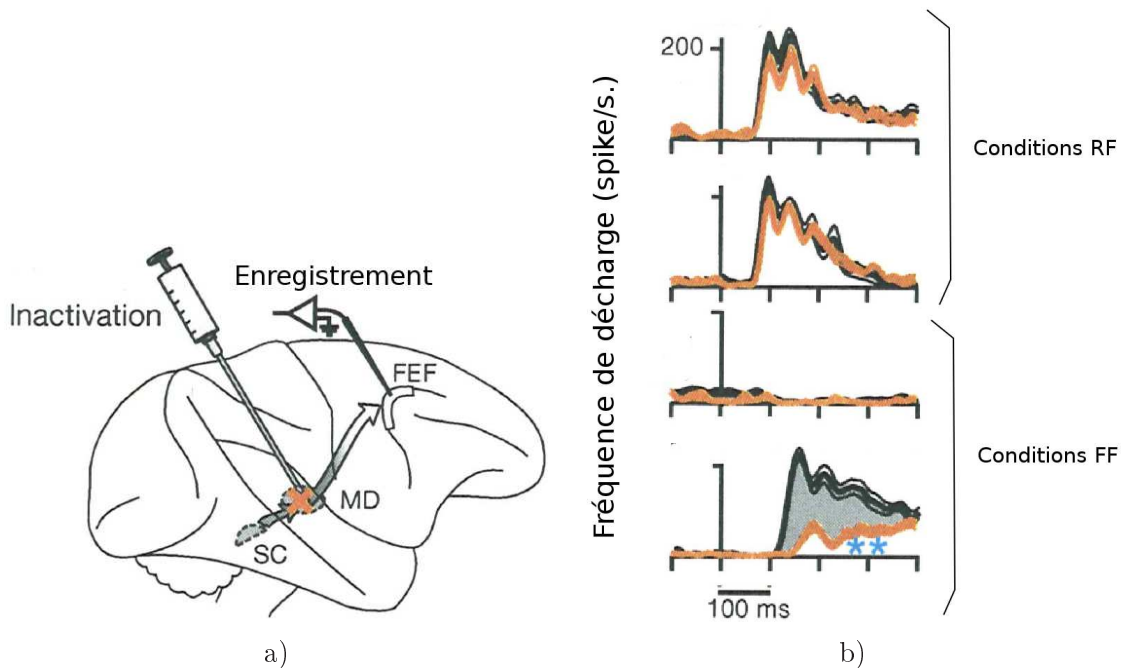


FIG. 3.25 – a) On enregistre l'activité de cellules de FEF, qui reçoivent des afférences de SC via le noyau médiodorsal du thalamus (MD), pendant qu'un singe exécute les mêmes tâches que celles présentées sur la figure 3.24a. Deux conditions sont étudiées : avec ou sans inactivation des cellules relais de MD. b, Conditions RF) La cellule enregistrée est sensible à son champ récepteur durant la fixation et peu de temps avant d'exécuter une saccade, dans les deux conditions (avec ou sans inactivation). b, Conditions FF) Avant inactivation, la cellule est sensible à son futur champ récepteur peu de temps avant d'exécuter une saccade mais son activité prédictive est significativement réduite lorsque MD est inactivé (figure du bas). Adapté de [Sommer et Wurtz, 2006]

[Sommer et Wurtz, 2006, Sommer et Wurtz, 2004a, Sommer et Wurtz, 2004b] vont plus loin qu'identifier des réponses quasivisuelles puisqu'ils montrent également qu'elles seraient la conséquence d'une décharge corollaire d'un mouvement saccadique envoyée par le colliculus supérieur aux champs oculomoteurs frontaux et relayée par le noyau médiodorsal du thalamus. Pour ce faire, ils commencent par identifier des neurones relais du noyau médiodorsal du thalamus situés sur le chemin des projections entre le colliculus supérieur et FEF, puis ils inactivent temporairement ces cellules relais (figure 3.25a) et étudient les conséquences de cette inactivation sur les performances dans une tâche de double saccade et sur les réponses quasivisuelles de FEF. La

figure 3.25b représente les activités enregistrées, dans les mêmes conditions que précédemment (figure 3.24a), avant et après injection de l'agent pharmacologique dans le noyau thalamique. On observe essentiellement que la réponse de la cellule à un stimulus présenté dans son futur champ récepteur peu de temps avant l'exécution d'une saccade est significativement diminuée. Par ailleurs, cette diminution significative d'activité ne s'observe que pour des saccades contraversives¹⁶, en accord avec les résultats de [Sommer et Wurtz, 2004a] selon lesquels les projections SC-MD-FEF dans un même hémisphère ne transportent que des signaux pour des saccades contraversives.

Pour résumer, les champs oculomoteurs frontaux contiennent des cellules dont les activités sont corrélées à la présentation d'un stimulus dans leur champ récepteur, à l'exécution d'une saccade volontaire dans leur champ de mouvement ou répondent dans les deux conditions. Les champs récepteurs et de mouvement de ces cellules sont définis par rapport à la position courante de l'œil, dans un référentiel centré-œil, avec la propriété intéressante qu'ils occuperaient par anticipation leur future position peu de temps avant d'exécuter une saccade. Certaines cellules déchargent également toniquement lorsqu'il est nécessaire de mémoriser la position d'une cible. Ces propriétés suggèrent que FEF fait partie d'un circuit impliqué dans la mémorisation d'une information spatiale, via par exemple des projections du cortex préfrontal dorsolatéral qui est impliqué dans la mémorisation à court-terme d'informations spatiales [Constantinidis et al., 2001]. Or, si FEF représente cette information dans un référentiel centré-œil, on peut imaginer au moins deux possibilités pour mettre à jour cette mémoire. La première est que les propriétés liées à la mémoire des cellules de FEF soient dues à des projections d'un circuit impliqué dans la mémorisation dans un référentiel plus abstrait (par exemple indépendant de la position de l'œil, comme un référentiel centré-tête). Nous verrons en particulier dans le prochain paragraphe que FEF est interconnecté avec les champs oculomoteurs supplémentaires dans lesquels l'information spatiale semble être codée dans un continuum de référentiels centré-tête, centré-œil et centré-corps. La seconde possibilité est que cette mémoire n'implique que des structures qui représentent l'information dans un référentiel centré-œil. Auquel cas, la décharge corollaire de SC qui semble être la cause des modifications de sélectivité des champs récepteurs de FEF, pourrait contribuer à la mise à jour de cette mémoire en anticipant les conséquences de la saccade oculaire qui va être exécutée.

Champs oculomoteurs supplémentaires (SEF)

Les champs oculomoteurs supplémentaires, situés sur la surface dorsomédiale du lobe frontal [Schlag et Schlag-Rey, 1987, Schlag et Schlag-Rey, 1985], sont impliqués dans le contrôle oculomoteur comme le montrent des études par microstimulation, lésion ou inactivation réversible ; on trouvera dans [Tehovnik et al., 2000] une revue des études par microstimulation, lésion et inactivation réversible de SEF. SEF est interconnecté avec FEF, dlPFC, LIP, 7a, MT, MST, SMA, le cortex cingulaire [Huerta et Kaas, 1990, Schall et al., 1993, Lynch et Tian, 2005]. Il se projette également sur les couches intermédiaires et profondes du colliculus supérieur et sur la formation réticulée, deux structures dont on connaît l'implication dans le contrôle oculomoteur [Huerta et Kaas, 1990].

¹⁶Une saccade contraversive par rapport à une structure se trouvant dans un hémisphère est dirigée vers l'hémisphère opposé. Par exemple, quand on parle de FEF droit, une saccade contraversive est dirigée vers la gauche et une saccade ipsiversive vers la droite

A partir des premières études par microstimulation, on pensait que SEF encodait la cible d'une saccade dans un référentiel centré-tête, de sorte qu'en stimulant électriquement un site de SEF, l'œil se retrouvait dans une position indépendante de la position de départ. Des études ultérieures ont montré qu'il existe une ségrégation entre les sites de SEF dont la stimulation produit une saccade d'amplitude et de directions fixées (donc dans un référentiel centré-œil, comme dans FEF ou SC). Ces sites se trouvent sur la région rostrale, et ceux dont la stimulation produit une saccade pour amener l'œil dans une position dans l'orbite caractéristique du site stimulé (donc dans un référentiel centré-tête), qui se trouvent dans la région caudale [Park et al., 2006]. [Martinez-Trujillo et al., 2004] vont même plus loin en observant qu'une stimulation de SEF produit une saccade dont la cible est codée par rapport à la position de la tête, à la position de l'œil et par rapport à la position du corps, classé par ordre de fréquence d'observation.

Comme FEF, SEF contient des cellules qui déchargent avant l'exécution d'une saccade. Par contre, les activités de SEF peuvent également être corrélées à un signal de récompense dans des tâches où les saccades sont conditionnées. Par exemple, dans des *countermanding task* où le singe doit exécuter une saccade simple vers une cible périphérique et l'inhiber si le point de fixation réapparaît, on observe que des cellules de SEF déchargent sur les essais réussis, d'autres sur les essais ratés [Stuphorn et al., 2000, Schall et al., 2000] et que leur réponse n'est pas corrélée à l'inhibition de la saccade mais indiquerait plutôt que la saccade a été ou non correctement annulée. Ces résultats sont en accord avec les travaux de [Amador et al., 2000] qui enregistrent des cellules dans SEF dont l'activité est corrélée à l'obtention d'une récompense (ces neurones déchargent lorsque le singe reçoit une récompense) et d'autres dont l'activité prédit une récompense (ces neurones déchargent progressivement jusqu'à ce que le singe reçoive une récompense, auquel cas, les cellules arrêtent brutalement de décharger). SEF semble également être impliqué dans l'apprentissage et la reconstitution de séquences de saccades [Isoda et Tanji, 2002, Tobler et Müri, 2002] et dans les tâches d'antisaccade [Amador et al., 1998, Amador et al., 2004].

Cortex préfrontal dorsolatéral (dlPFC)

Le cortex préfrontal dorsolatéral (dlPFC, *dorsolateral prefrontal cortex*, figure 3.11b), une région entourant le sillon principal dans la partie antérieure du cortex cérébral, semble être impliqué dans au moins trois fonctions liées au contrôle oculomoteur [Pierrot-Deseilligny et al., 2004, Pierrot-Deseilligny et al., 2003, Constantinidis et Wang, 2004, Goldman-Rakic, 1987, Fuster, 1997] :

- l'inhibition de saccades non désirées dans les tâches d'antisaccade où le sujet doit exécuter une saccade en direction d'une cible imaginaire symétrique de la cible visuelle présentée, et ne doit pas exécuter de saccade vers la cible visuelle ;
- les processus décisionnels ;
- la mémoire spatiale à court terme (jusqu'à une vingtaine de secondes).

Le cortex préfrontal dorsolatéral est connecté avec les autres champs oculomoteurs, les aires visuelles associatives et les structures limbiques [Goldman-Rakic, 1987, Selemon et Goldman-Rakic, 1988]. Il est effectivement densément interconnecté avec FEF dont on a vu le rôle dans

le contrôle oculomoteur précédemment [Barbas et Mesulam, 1985, Huerta et al., 1987], avec SEF dont on a explicité le rôle dans le codage de séquences de saccades oculaires et avec LIP et 7a [Barbas et Mesulam, 1985, Goldman-Rakic, 1987], les deux principales aires du cortex pariétal postérieur impliquées dans le contrôle oculomoteur. Selon [Goldman-Rakic, 1987], les projections récurrentes organisées topographiquement entre dlPFC et le pariétal postérieur pourraient contribuer à la mémorisation à court terme d'une information spatiale. Il est également interconnecté avec le cortex parahippocampique et la formation hippocampique, deux structures impliquées dans la mémorisation à long terme [Pierrot-Deseilligny et al., 2004].

Différents travaux portant sur des lésions, des perturbations temporaires ou des enregistrements cellulaires de dlPFC soulignent son implication dans la mémoire spatiale à court terme [Pierrot-Deseilligny et al., 2002, Ploner et al., 1999, Müri et al., 1996, Petrides, 1994, Funahashi et al., 1989]. En particulier, [Müri et al., 1996] étudient les conséquences d'une stimulation magnétique du cortex préfrontal dorsolatéral pendant une tâche de saccade dirigée vers une cible mémorisée pour laquelle le délai de mémorisation est de deux secondes. Ils montrent en particulier que lorsque la stimulation, qui perturbe temporairement l'aire ciblée, est appliquée entre 700 et 1500 ms après la présentation de la cible à mémoriser, la stimulation affecte l'amplitude de la saccade que le sujet effectue, lorsqu'elle est dirigée dans l'espace contralatéral à l'aire ciblée. Dans [Funahashi et al., 1989], les auteurs enregistrent l'activité de cellules dans dlPFC pendant qu'un singe réalise une tâche de saccade vers une cible mémorisée. L'animal doit conserver les yeux fixés sur un point de fixation pendant 1 à 6 secondes. Pendant ce temps, une cible est brièvement présentée en périphérie, pendant 500 ms. Les auteurs observent que 30% des cellules enregistrées ont une variation d'activité significative pendant le délai. De plus, 58% de ces cellules ont une activité plus forte pendant le délai et la plupart de ces cellules sont sélectives à la position de la cible mémorisée (figure 3.26). C'est à dire qu'elles déchargent toniquement pendant le délai si une cible à mémoriser se trouve dans une région localisée de l'espace visuel. Enfin, il est intéressant de noter que lorsqu'une récompense, ajoutée au paradigme standard de saccade vers une cible mémorisée, est prévisible, les activités soutenues pendant le délai dans dlPFC sont modulées par l'attente de la récompense [Kobayashi et al., 2002].

En plus de son rôle dans la mémorisation à court terme, dlPFC serait également impliqué dans le contrôle oculomoteur en biaisant le comportement par facilitation ou inhibition. En particulier, dlPFC a des projections inhibitrices sur le colliculus supérieur susceptibles d'expliquer son rôle dans l'inhibition de saccades non désirées (par exemple une saccade réflexe vers la cible visuelle présentée pendant une tâche d'antisaccade) [Pierrot-Deseilligny et al., 1991b, Goldman-Rakic, 1987]. dlPFC est également en interaction avec les ganglions de la base [Middleton et Strick, 2002]. Le circuit ainsi formé pourrait former une mémoire de travail [Frank et al., 2001] et contribuer à la sélection de l'action.

En conclusion, le cortex préfrontal dorsolatéral semble contribuer à la mémorisation à court terme ainsi qu'à la régulation de commandes motrices. Si on se restreint au contrôle oculomoteur, il joue un rôle particulièrement important dans les tâches de saccades dirigées vers des cibles mémorisées, d'antisaccade, de séquences de saccades, en un mot toutes les saccades volontaires reposant sur une information spatiale mémorisée récemment.

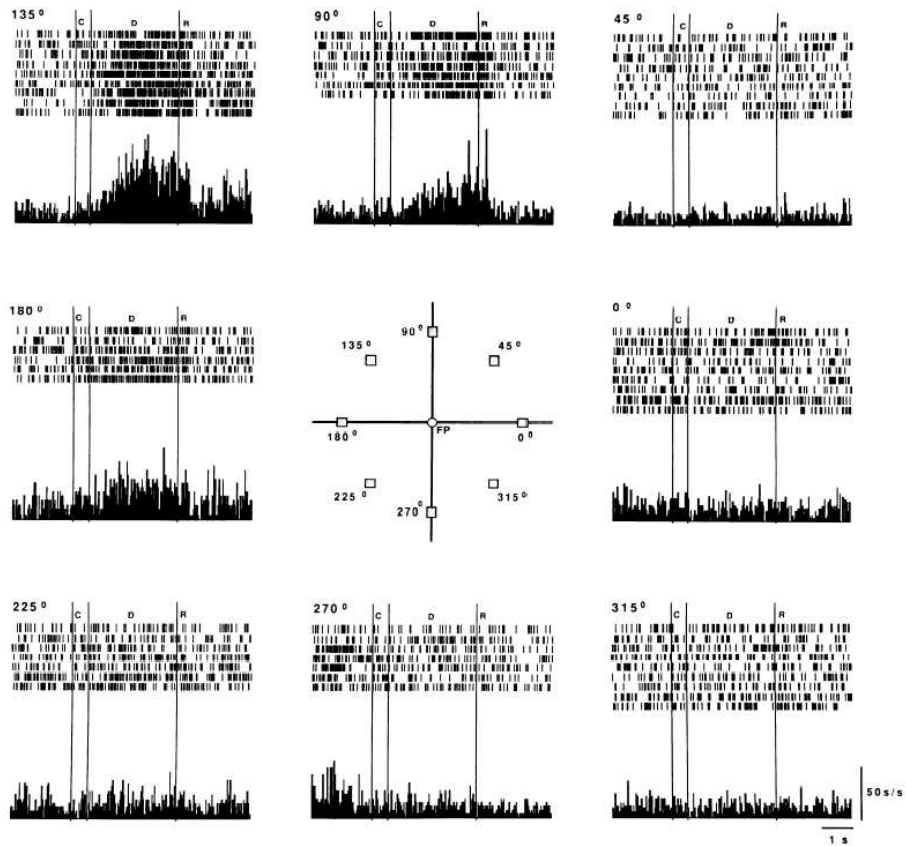


FIG. 3.26 – Activité d'un neurone du cortex préfrontal lors de la mémorisation de saccades dans huit directions. La période appelée 'C' correspond à la présentation de l'indice, celle appelée 'D' au délai de trois secondes et celle appelée 'R' à l'exécution de la saccade. Le neurone est actif pendant le délai lorsque la saccade à exécuter est dans les directions 90° , 135° et 180° , avec un maximum en 135° . Adapté de [Funahashi et al., 1989]

3.3.4 Ganglions de la base

Les ganglions de la base (BG, figure 3.27) sont constitués d'un ensemble de noyaux dont le rôle est central dans le contrôle moteur (sélection de l'action, mémoire de travail, récompense obtenue et attendue, apprentissage de séquences motrices) comme le soulignent de nombreux désordres moteurs associés à des maladies affectant les ganglions de la base (maladie de Parkinson, chorée de Huntington) [Hikosaka et al., 2000, Redgrave, 2007]. Certains noyaux ont des connexions intrinsèques aux ganglions de la base et d'autres interagissent avec une majeure partie des structures corticales et sous-corticales. On distingue ainsi deux principales structures d'entrée, le striatum (*striatum*, STR) et les noyaux sous-thalamiques (*subthalamic nucleus*, STN) et deux principales structures de sortie, la partie réticulée de la substance noire (*substantia nigra pars reticulata*, SNr) et le segment interne du globus pallidus (GPi). Le segment externe du globus pallidus (GPe) est la principale structure intrinsèque aux ganglions de la base en ce sens qu'il est connecté presque exclusivement avec les autres noyaux de BG. Enfin, la partie compacte de la substance noire (*substantia nigra pars compacta*, SNc) interagit fortement avec le système dopaminergique qui fournit une modulation liée à la récompense obtenue ou attendue.

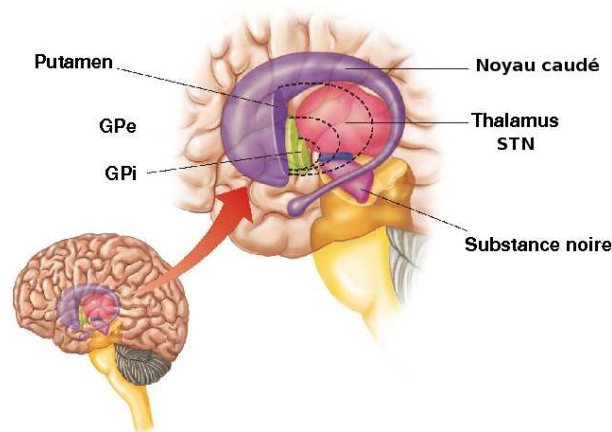


FIG. 3.27 – a) Noyaux des ganglions de la base. Le striatum est composé du noyau caudé et du putamen. Le Globus Pallidus (GP) est constitué de deux parties fonctionnellement distinctes : le segment interne (GPi) et le segment externe (GPe). La substance noire est composée de la partie compacte (SNc) et la partie réticulée (SNr). Les ganglions de la base comprennent également les noyaux sous thalamiques (STN).

Les afférences des aires corticales motrices, oculomotrices, préfrontales et associatives se projettent sur des parties distinctes du striatum et forment un des maillons de boucles impliquant le cortex, les ganglions de la base et le thalamus, partiellement indépendantes [Alexander et al., 1986]. Nous avons vu dans les paragraphes précédents que les champs oculomoteurs pariétaux et frontaux se projettent directement sur le colliculus supérieur. Ces aires se projettent également indirectement sur le colliculus supérieur via les ganglions de la base (figure 3.28). Ces projections sont supposées avoir un rôle fonctionnel complètement différent des projections corticales directes. En effet, tandis que les projections directes fourniraient la ou les cibles d'un mouvement oculaire, les ganglions de la base, par l'intermédiaire de projections inhibitrices toniques de SNr sur le colliculus supérieur [Hikosaka et Wurtz, 1985b], permettraient de favoriser une

cible au détriment des autres. Les neurones de SNr, qui est considéré comme l'un des principaux noyaux de sortie des ganglions de la base en ce qui concerne les mouvements oculaires, ont un taux de décharge spontanée élevé. Ils exercent ainsi une inhibition active sur les couches intermédiaires et profondes du colliculus supérieur. Ils sont inhibés sélectivement par le noyau caudé de telle sorte qu'une excitation du noyau caudé produit une inhibition de SNr permettant de relâcher sélectivement l'inhibition sur le colliculus supérieur, autorisant ainsi l'excitation des neurones colliculaires qui sont la cible des projections nigrales. Ce principe d'influence par inhibition/désinhibition semble être un principe général de l'action des ganglions de la base [Hikosaka et al., 2000, Penney et Young, 1983, Deniau et Chevalier, 1985]. SNr a également des afférences excitatrices des noyaux sous-thalamiques qui, comme nous le verrons par la suite, ont un effet opposé aux projections du noyau caudé. Nous proposons par la suite de nous concentrer sur les trois structures mentionnées précédemment (noyau caudé, noyaux sous-thalamiques, substantia nigra pars reticulata) en soulignant les propriétés physiologiques, liées au contrôle oculomoteur, des neurones de ces structures. On trouvera une revue récente de l'implication des ganglions de la base dans le contrôle des saccades dans [Hikosaka, 2007].

On distingue dans le noyau caudé une région oculomotrice qui a des afférences de FEF, SEF et du cortex préfrontal dorsolatéral [Parthasarathy et al., 1992, Stanton et al., 1988a, Shook et al., 1991, Selemon et Goldman-Rakic, 1985, Yeterian et Pandya, 1991]. Cette région contient des cellules dont les réponses peuvent être liées à la présentation d'un stimulus visuel, à l'exécution d'une saccade dans leur champ de mouvement centré dans l'hémichamp contralatéral ou à la mémorisation d'une cible [Hikosaka et Wurtz, 1985a]. En particulier, les cellules qui déchargent lorsqu'une saccade est dirigée vers leur champ de mouvement ne déchargent pas pour des saccades spontanées (ou réflexes). Ces neurones ont une activité également très dépendante de la récompense et l'activité de certains de ces neurones change, même avant la présentation d'une stimulation visuelle, lorsqu'une récompense associée à une position dans l'hémichamp contralatéral est prévisible [Lauwereyns et al., 2002, Hikosaka et al., 2006].

Le noyau caudé a des projections inhibitrices directes sur SNr ; ces projections forment la *voie directe* des ganglions de la base qui émanent des cellules du noyau caudé ayant des récepteurs dopaminergiques D1 [Albin et al., 1989]. Il se projette également sur les noyaux sous-thalamiques (STN) via le segment externe du globus pallidus par une double inhibition, STN ayant des projections excitatrices sur SNr ; ces projections forment la *voie indirecte* des ganglions de la base qui émanent des cellules du noyau caudé possédant des récepteurs dopaminergiques D2 [Albin et al., 1989]. La dopamine n'a pas la même influence sur le noyau caudé en fonction du type de récepteur : elle a globalement un effet excitateur pour les récepteurs D1 et un effet inhibiteur pour les récepteurs D2. Nous reviendrons à la fin de cette section sur le rôle respectif supposé des deux "circuits" dopaminergiques, après la description de l'architecture des ganglions de la base.

SNr est l'une des principales structures de sortie des ganglions de la base et projette des connexions inhibitrices sur les couches intermédiaires du colliculus. Ses neurones ont un taux de décharge spontanée élevé, et sont silencieux avant les saccades volontaires dirigées vers une cible visuelle ou mémorisée [Hikosaka et Wurtz, 1983a, Hikosaka et Wurtz, 1983b, Hikosaka et Wurtz, 1983c, Hikosaka et Wurtz, 1983d]. L'un des rôles supposés des projections nigrales sur le colliculus supérieur serait d'empêcher l'exécution d'une saccade, leur inhibition devant être

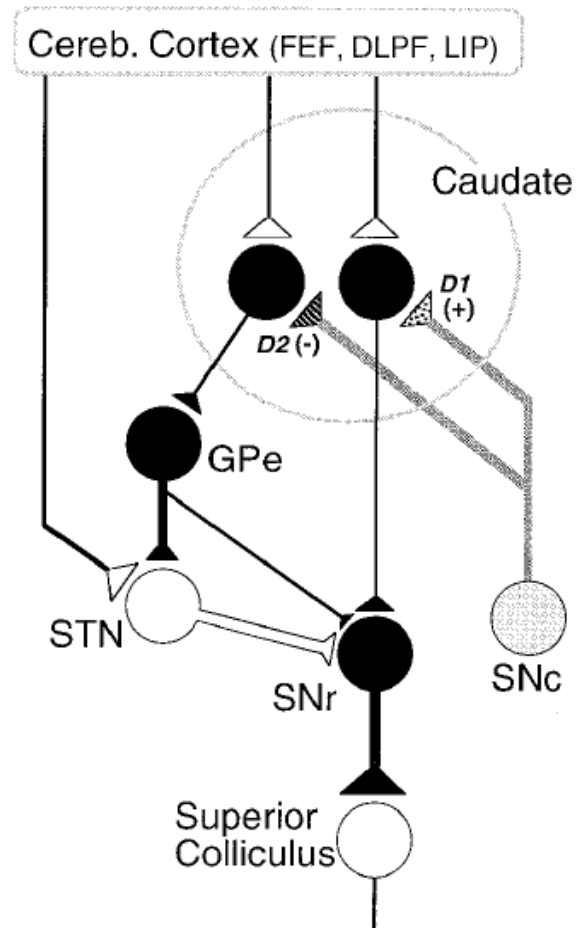


FIG. 3.28 – On distingue principalement deux mécanismes parallèles et antagonistes quant à leurs effets sur SNr : la voie directe entre le noyau caudé et SNr, modulée par la dopamine qui agit sur les récepteurs D1, et la voie indirecte entre le noyau caudé et SNr, par l'intermédiaire de deux inhibitions successives sur GPe puis sur STN. Les projections dont la terminaison est noire sont inhibitrices et celles dont la terminaison est blanche sont excitatrices. SNc : partie compacte de la substance noire ; SNr : partie réticulée de la substance noire ; STN : noyaux sous-thalamiques ; GPe : segment externe du globus pallidus. Adapté de [Hikosaka et al., 2000]

supprimée sélectivement pour qu'une saccade volontaire puisse être exécutée. En faveur de cette hypothèse, [Hikosaka et Wurtz, 1985b] observent que la suppression, à l'aide d'un agent pharmacologique, de l'influence inhibitrice de SNr sur SC conduit à des saccades spontanées et inévitables en direction d'un stimulus présenté dans l'hémichamp contralatéral au site lésé, même si la tâche requiert de la part du singe de maintenir son regard sur un point de fixation.

En plus des projections inhibitrices du noyau caudé, SNr est également sous l'influence excitatrice des noyaux sous thalamiques (STN), eux-mêmes sous l'influence du noyau caudé par une succession de deux inhibitions, via le segment externe du globus pallidus (GPe) (figure 3.28). Comme les neurones de GPe ont un taux de décharge spontané élevé, une stimulation du noyau caudé conduit à une inhibition de GPe qui relâche alors son inhibition sur STN. STN ayant des afférences corticales [Monakow et al., 1978, Nambu et al., 1996], il a ainsi la possibilité d'exciter SNr et d'avoir une influence opposée aux projections directes du noyau caudé sur SNr. Ainsi, le noyau caudé peut avoir une influence soit inhibitrice soit facilitatrice sur SNr via ses projections directes et indirectes, ce qui est cohérent avec des expériences de microstimulation du noyau caudé [Hikosaka et al., 1993].

D'après [Hikosaka et al., 2000], sachant que le colliculus supérieur a des afférences de différentes modalités (e.g. auditive, visuelle) et de diverses aires corticales (e.g. LIP, FEF), les projections inhibitrices de SNr sur le colliculus supérieur offrent un moyen privilégié de contrôler, de sélectionner quelle cible, parmi plusieurs cibles potentielles, sera la cible d'un mouvement saccadique, et ce sur la base d'informations mémorisées ou de récompense attendue. La modulation de ces projections inhibitrices (par facilitation ou par inhibition) est la conséquence d'un réseau complexe impliquant des projections intrinsèques aux ganglions de la base et des projections provenant de structures extérieures aux ganglions de la base, qui rend leur compréhension difficile. [Hikosaka, 2007] propose une interprétation des voies directes et indirectes dans le contrôle des saccades oculaires. Il observe que le blocage sélectif de l'un ou l'autre des récepteurs dopaminergiques conduit à des comportements différents dans des tâches saccadiques récompensées, en fonction de l'amplitude de la récompense, et suggère que cette différence peut s'expliquer par une modification de l'influence de la voie directe ou de la voie indirecte sur la substance noire réticulée SNr. En particulier, l'auteur observe que le temps de réaction est significativement plus long pour des récompenses importantes lorsque le récepteur D1 est bloqué, et pour des récompenses plus faibles lorsque le récepteur D2 est bloqué. L'interprétation de l'auteur est la suivante : puisque la dopamine a un effet excitateur sur le noyau caudé via le récepteur D1, son blocage diminue l'excitation du noyau caudé d'autant plus que la récompense est importante, tout en maintenant l'inhibition sur les cellules du noyau caudé à récepteur D2. La diminution de l'excitation du noyau caudé conduit à une réduction de l'inhibition de SNr, via le chemin direct, et donc à une réduction de la désinhibition de SC. Le colliculus supérieur étant inhibé plus fortement lorsque les récepteurs D1 sont bloqués, le temps de réaction est plus long. Au contraire, la dopamine a un effet inhibiteur sur le noyau caudé via le récepteur dopaminergique D2. En le bloquant, le noyau caudé étant moins inhibé, il inhibe plus fortement GPe. Comme GPe a un taux de décharge spontané élevé, son inhibition permet de relâcher l'inhibition sur STN qui favorise ainsi l'excitation de STN sur SNr. Puisque l'activité de SNr est plus importante, elle maintient plus fortement son inhibition sur le colliculus, ce qui expliquerait à nouveau l'augmentation du temps de réaction. Le fait que le blocage du récepteur D1 n'a une influence que pour les fortes récompenses tandis que le blocage du récepteur D2 n'a d'influence que pour

les faibles récompenses pourrait s'expliquer par la sensibilité à la récompense des deux récepteurs dopaminergiques. Ces résultats suggèrent que dans leur fonctionnement normal, la sélection par désinhibition des ganglions de la base est le résultat d'un équilibre entre l'influence inhibitrice de la voie directe sur SNr modulée par les récepteurs dopaminergiques D1 et l'influence excitatrice de la voie indirecte modulée par les récepteurs dopaminergiques D2.

On trouvera également dans [Brown et al., 2004] une étude de modélisation récente donnant une interprétation fonctionnelle aux différentes projections internes aux ganglions de la base en terme de sélection, de rétention (pendant un délai) et de maintien (pendant son exécution) d'un plan d'action, en considérant non seulement les voies directes et indirectes mais également une voie hyperdirecte impliquant les entrées corticales des ganglions de la base par l'intermédiaire des noyaux sous-thalamiques. On trouvera également dans [Girard, 2003] une revue des différents modèles computationnels impliquant les ganglions de la base dans des mécanismes cognitifs tels que la sélection de l'action, la mémorisation à court terme ou l'apprentissage de séquences motrices.

3.4 Mécanismes de traitement de l'information spatiale : transformation et mise à jour

Le début de ce chapitre est exclusivement consacré à l'étude des mouvements oculaires et à la description anatomique et fonctionnelle des principales structures corticales et sous-corticales impliquées dans le contrôle de ces mouvements, tout en essayant de mettre en avant les interactions entre ces structures. Dans cette dernière partie, on considère ces structures dans leur ensemble en s'intéressant au problème qu'est la représentation des informations spatiales, des structures sensorielles aux structures motrices. On présente pour ce faire des modèles qui traitent des transformations sensorimotrices dans le paragraphe 3.4.1 puis des modèles de mise à jour de représentations spatiales pendant l'exécution d'une saccade dans le paragraphe 3.4.2.

3.4.1 Transformations sensorimotrices : champs de gain

Les informations sensorielles sont traitées en premier lieu par des senseurs et représentées dans des référentiels qui leur sont liés. En particulier, l'information visuelle est représentée dans un référentiel lié à la position courante de l'œil, ou rétinotopique. Une stimulation auditive, traitée par le système auditif, est représentée dans un référentiel lié à la tête, tout comme une stimulation tactile sur le visage. Les commandes motrices sont quant à elles dépendantes de l'effecteur engagé dans l'action. En particulier, la mise en mouvement des yeux nécessite de contrôler les muscles extraoculaires et il est donc nécessaire de fournir aux muscles extraoculaires, via les motoneurones qui les innervent, une commande adéquate à leur contraction ou à leur relâchement et nous avons vu dans le paragraphe 3.2.1 que les noyaux de la formation réticulée étaient chargés de produire cette commande temporelle. Les mouvements du bras nécessitent de coordonner la contraction ou le relâchement des muscles qui contrôlent ses articulations, donc de produire des commandes dans un référentiel complètement différent de celui dans lequel est représenté le stimulus vers lequel le bras est dirigé. On peut ainsi définir le problème général des

transformations sensorimotrices comme la nécessité de convertir un signal sensoriel représenté dans un référentiel propre au senseur, en un signal moteur représenté dans un référentiel propre à l'effecteur.

Les champs de gains (*gain field*) sont un des mécanismes essentiels qui semblent intervenir dans les transformations sensorimotrices. Un champ de gain est caractérisé par le changement de l'amplitude de la réponse d'un neurone, indépendamment de sa sélectivité ou des caractéristiques de son champ récepteur [Salinas et Thier, 2000]. On trouve cette propriété de champs de gains dans plusieurs aires du cortex [Graziano, 2006, Salinas et Sejnowski, 2001] le long de la voie dorsale (aires 7a et LIP du pariétal postérieur [Andersen et al., 1985, Andersen et al., 1990b], aire MIP du pariétal postérieur [Batista et al., 1999], cortex prémoteur dorsal [Jouffrais et Boussaoud, 1999], cortex prémoteur ventral [Graziano et al., 1997, Graziano et Gross, 1998]). La première observation de cette propriété est due à l'équipe de Vernon Mountcastle [Andersen et Mountcastle, 1983, Andersen et al., 1985]. Les auteurs ont observé que, dans le pariétal postérieur, les neurones ont des champs récepteurs visuels rétinotopiques (c'est à dire qu'ils répondent lorsqu'un stimulus est présenté dans une région du champ visuelle relative à la position des yeux) dont la réponse est modulée par la position des yeux, un phénomène dénommé *gaze effect*.

Les premiers travaux de modélisation de [Zipser et Andersen, 1988] ont révélé que les champs de gains pouvaient être obtenus par apprentissage d'un perceptron à trois couches. La couche d'entrée contient deux représentations qui codent respectivement un stimulus visuel dans une carte 2D rétinotopique et la position des yeux par quatre populations sigmoïdales (deux populations, à pente négative ou positive, par composante, verticale ou horizontale). La couche cachée combine additivement les entrées par des poids initialement aléatoires. Une non-linéarité est introduite par la fonction de transfert sigmoïdale des neurones de la couche cachée. Enfin la couche de sortie représente la position du stimulus visuel dans un référentiel centré-tête, c'est à dire indépendamment de la position des yeux. Après l'apprentissage par rétropropagation d'un ensemble de couples entrées-sortie, les auteurs observent que les neurones de la couche cachée ont des réponses similaires aux champs de gains observés dans le pariétal.

Cette même démarche consistant à utiliser un perceptron multicouches et à apprendre une transformation par rétropropagation du gradient est également utilisée par [Xing et Andersen, 2000]. Les auteurs vont néanmoins plus loin, en utilisant systématiquement différentes architectures et en analysant les propriétés des neurones de la couche cachée. Les architectures se distinguent à la fois par leurs entrées sensorielles (stimulus auditif dans un référentiel centré-tête noté A , stimulus visuel dans un référentiel centré-œil noté V) et proprioceptives (position de l'œil par rapport à la tête notée E , position de la tête par rapport au corps notée H) et leurs sorties (représentation centrée-œil notée ME , centrée-tête notée HE , centrée-corps notée BE). Les transformations apprises sont définies par les équations 3.1.

$$\begin{aligned}
ME &= \begin{cases} V & \text{pour un stimulus visuel} \\ A - E & \text{pour un stimulus auditif} \end{cases} \\
HE &= \begin{cases} V + E & \text{pour un stimulus visuel} \\ A & \text{pour un stimulus auditif} \end{cases} \\
BE &= \begin{cases} V + E + H & \text{pour un stimulus visuel} \\ A + H & \text{pour un stimulus auditif} \end{cases} \quad (3.1)
\end{aligned}$$

Le principal résultat de ces modèles est qu'en cherchant à réaliser des transformations sensorimotrices, apprises par rétropropagation du gradient de l'erreur sur un perceptron à trois couches, les champs de gains émergent de l'apprentissage, ce qui suggère que les champs de gain sont efficaces pour calculer ces transformations.

Pour montrer l'efficacité des champs de gains dans le calcul des transformations sensorimotrices, [Pouget et Sejnowski, 1997] les représentent explicitement en utilisant des fonctions à bases radiales (*radial basis functions*) qui combinent multiplicativement une entrée sensorielle et une entrée motrice. Leur architecture, illustrée sur la figure 3.29a, se compose également de trois niveaux : les couches d'entrée représentent respectivement l'information visuelle avec des cellules à champs récepteurs rétino-topiques gaussiens et la position de l'œil avec des cellules à champs récepteurs sigmoïdaux. Chaque neurone de la couche cachée reçoit en entrée le produit de l'activité de deux neurones : l'un de la couche visuelle, l'autre de la couche proprioceptive. Le modèle étudié par les auteurs considère des représentations 1D de telle sorte que la couche cachée est 2D afin de représenter tous les couples de deux neurones d'entrée. Enfin, la couche de sortie est constituée d'une ou plusieurs représentations dont les cellules prennent leurs entrées sur la couche cachée, les liens entre la couche cachée et la couche de sortie étant déterminés par apprentissage. L'un des principaux résultats des auteurs est que les champs de gain (construits explicitement dans leur modèle) permettent de décoder en sortie n'importe quelle combinaison linéaire des variables codées en entrée (position d'un stimulus visuel, position de l'œil par rapport à la tête), donc en particulier une représentation centrée-œil et une représentation centrée-tête. Il est important de noter que ces représentations sont décodées à partir du même substrat.

Les auteurs vont même plus loin en proposant une architecture dans laquelle les projections entre les couches ne sont plus simplement unidirectionnelles mais récurrentes [Pouget et al., 2002]. Dans ce cas, il y a une interaction bidirectionnelle entre la représentation centrée-œil et la représentation centrée-tête, de telle sorte qu'il est possible de déterminer la position d'un stimulus auditif par rapport à la position courante de l'œil et la position d'un stimulus visuel par rapport à la position courante de la tête, avec le même substrat neuronal. Ces résultats conduisent les auteurs à proposer que les fonctions radiales de bases pourraient constituer l'interface entre des modalités sensorielles (vision, audition, toucher), des modalités proprioceptives (position de l'œil par rapport à la tête, position de la tête par rapport au corps) et des modalités motrices (mouvement de l'œil, mouvement de la tête, mouvement de saisie du bras) (figure 3.29b).

Puisque les champs de gains semblent être particulièrement adaptés pour réaliser des transformations sensorimotrices et qu'ils peuvent être modélisés par une interaction multiplicative

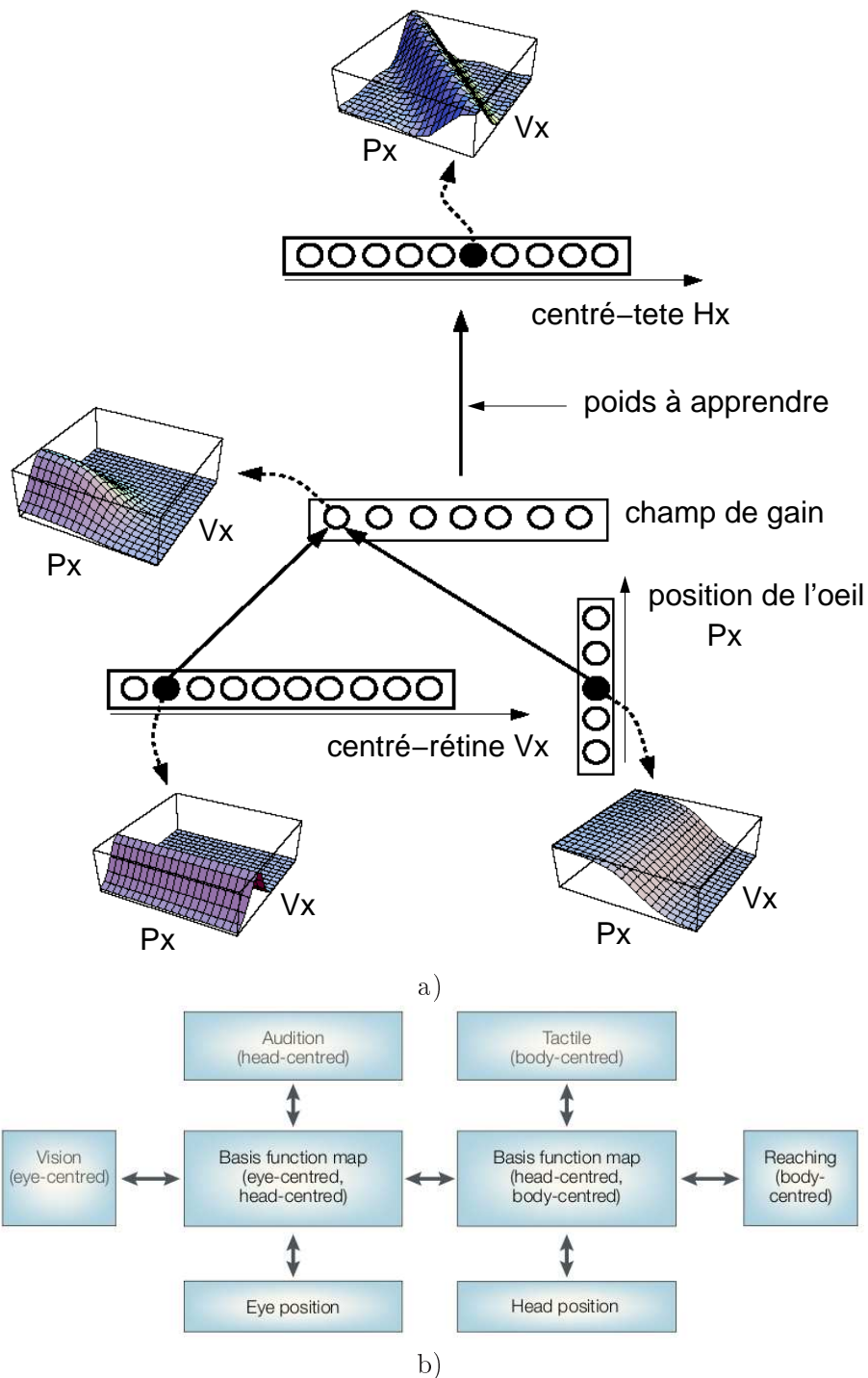


FIG. 3.29 – a) Dans le cadre des fonctions radiales de base utilisées pour calculer des transformations sensorimotrices, chaque neurone du champ gain reçoit en entrée le produit de l'activité de deux neurones : un neurone sensoriel ayant un champ récepteur gaussien rétinotopique et un neurone proprioceptif codant la position de l'œil avec une réponse sigmoïdale. La transformation sensorimotrice, qui peut être n'importe quelle combinaison linéaire des deux informations codées en entrée (position du stimulus visuel et position de l'œil), est apprise par l'intermédiaire des poids entre le champ de gains et la couche de sortie. Adapté de [Pouget et Sejnowski, 1997] b) Les champs de gains modélisés par des fonctions radiales de base pourraient constituer des interfaces entre des informations sensorielles (stimulus visuel, auditif), proprioceptives (position de l'œil, position de la tête) et motrices (mouvement de l'œil, mouvement de la tête, mouvement du bras) . Adapté de [Pouget et al., 2002]

entre des signaux sensoriels, moteurs et proprioceptifs, une question encore ouverte concerne les corrélats biophysiques de cette modulation non-linéaire. Certains auteurs suggèrent que cette interaction multiplicative peut être la conséquence d'interactions latérales (excitation locale, inhibition distale) [Salinas et Abbott, 1995, Salinas et Abbott, 1996]. D'autres auteurs se penchent sur des modèles biophysiques détaillés [Chance et al., 2002, Mehaffey et al., 2005, Tiesinga et al., 2004] comme par exemple [Mel, 1993, Mel, 1999] qui étudient l'influence des synapses dans l'apport de cette non-linéarité.

3.4.2 Mettre à jour une mémoire de travail spatiale

Plusieurs mécanismes ont été proposés pour mettre à jour une mémoire de travail spatiale étant donnée une saccade à exécuter. [Pouget et Sejnowski, 2002] distinguent deux familles de mécanismes qui se différencient par la nature de l'information utilisée pour mettre à jour la mémoire : l'amplitude et la direction du déplacement ou la vitesse du déplacement. La première famille nécessite de connaître l'amplitude et la direction du mouvement oculaire, une information que l'on trouve par exemple dans le colliculus supérieur ou les cellules motrices de FEF, et l'utilise pour mettre à jour la mémoire en calculant une translation des profils mémorisés. Nous avons vu dans le paragraphe précédent différents mécanismes pour calculer la position d'un stimulus visuel dans un référentiel centré-tête étant donnée sa position dans un référentiel centré-œil et la position des yeux. Ces mécanismes calculent ce changement de représentation en approximant cette transformation par une translation des profils d'activité. C'est typiquement l'opération requise pour la mise à jour de la mémoire étant donnée la direction et l'amplitude d'une saccade. De par leurs capacités de représentations, les mécanismes *feedforward* reposant sur un perceptron multicouches [Zipser et Andersen, 1988, Xing et Andersen, 2000] peuvent réaliser cette opération, au même titre que les réseaux utilisant des fonctions à bases radiales [Pouget et Sejnowski, 2001].

La seconde famille utilise la vitesse de déplacement des yeux et met à jour la mémoire continûment avec le déplacement des yeux. Cette vitesse est utilisée pour moduler des connexions latérales asymétriques qui déplacent les profils d'activités représentant les cibles mémorisées. Dans [Zhang, 1996], l'auteur considère des connexions latérales avec une première composante en différence de gaussiennes et une seconde composante qui est la dérivée de la différence de gaussiennes modulée par le signal de vitesse (figure 3.30a). Les connexions latérales en différence de gaussiennes permettent de maintenir une activité même en l'absence de stimulus puisque les cellules dans un voisinage local s'excitent mutuellement, l'inhibition distante permettant d'éviter que l'activité ne se propage sur toute la carte. Les connexions asymétriques apportent un biais directionnel qui, modulé par la vitesse de déplacement des yeux, permet de translater les profils mémorisés, sans altérer sa forme (figure 3.30b, voir également la section 2.3). La principale difficulté imposée par l'utilisation de ces mécanismes est qu'il faut disposer de la vitesse de déplacement des yeux. [Dominey et Arbib, 1992] proposent de la dériver de deux positions successives des yeux fournies directement par les signaux proprioceptifs ou indirectement, par exemple via les champs de gains pariétaux. Dans [Droulez et Berthoz, 1991], les auteurs utilisent un mécanisme similaire au précédent, et soulignent que l'on peut imaginer différentes sources pour fournir, ou à partir desquelles calculer, le signal de vitesse. Il pourrait s'agir du flot optique, causé par le déplacement des yeux. Néanmoins, cette solution ne semble pas viable pour les saccades oculaires étant données la vitesse de déplacement des yeux et la latence du traite-

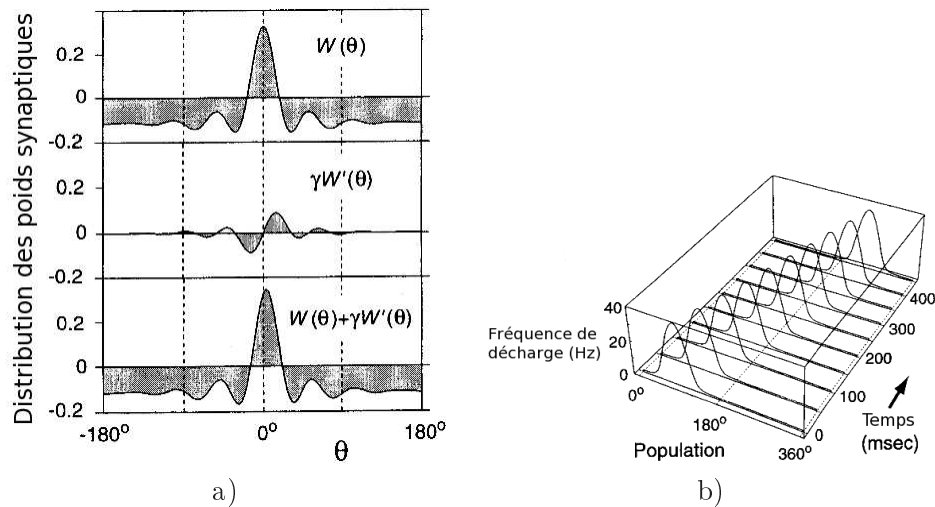


FIG. 3.30 – a) Les connexions latérales ont deux composantes, une composante symétrique à laquelle s’ajoute une composante asymétrique, dérivée de la composante symétrique, et modulée par un terme proportionnel à la vitesse désirée de déplacement du profil d’activité. b) En choisissant la composante impaire dérivée de la composante paire, le déplacement du profil d’activité n’altère pas sa forme. Adapté de [Zhang, 1996]

ment des informations visuelles. Une autre solution est ce que signal pourrait être obtenu par les noyaux de la formation réticulée dont nous avons parlé au paragraphe 3.2.1, étant donné qu’une des informations codées par les afférences des neurones oculomoteurs est justement la vitesse de déplacement des yeux.

Conclusion

Nous avons vu dans ce chapitre les différents mouvements oculaires ainsi qu’une majeure partie des structures corticales et sous-corticales impliquées dans le contrôle de l’un de ces mouvements en particulier, les saccades oculaires. Le terme “contrôle” reste néanmoins vague et comprend un ensemble de processus. Tout d’abord les informations sensorielles sont traitées pour en extraire à la fois les caractéristiques visuelles qui conduisent au choix d’une cible et les informations spatiales qui, transformées d’un référentiel sensoriel en un référentiel moteur, permettent de guider une saccade. Nous avons présenté succinctement les aires de la voie ventrale impliquées dans la reconnaissance des stimuli visuels avant de s’intéresser plus spécifiquement au traitement, à la représentation et aux transformations de l’information spatiale le long de la voie dorsale. Le colliculus supérieur qui est la dernière structure à coder spatialement la cible d’une saccade oculaire (en omettant les projections directes des structures corticales sur la formation réticulée) reçoit de ses afférences un ensemble de cibles potentielles. Ces cibles peuvent être des cibles suffisamment saillantes pour l’organisme pour susciter un mouvement d’orientation réflexe. Elles peuvent également faire partie d’un plan plus complexe d’exploration volontaire. Les ganglions de la base ont un rapport privilégié avec le colliculus supérieur par l’intermédiaire de leur inhibition tonique qui doit être levée sélectivement pour qu’un mouvement oculaire puisse être exécuté. Ce

rôle dans la sélection de la prochaine cible saccadique est renforcé par leurs nombreuses afférences corticales ainsi que par leurs interactions avec le système dopaminergique, lié quant à lui à la récompense obtenue ou attendue. Enfin, on a également mis en avant la contribution du cortex frontal dans les tâches saccadiques complexes. Ces aires contribuent à l'élaboration d'un plan d'actions (comme SEF qui est impliqué dans la mémorisation et la reconstitution de séquences de saccades) ainsi que la mémorisation à court terme. Bien qu'ayant orienté notre discours sur la mémorisation d'informations spatiales, en soulignant le rôle du cortex préfrontal dorsolatéral, la région ventrale adjacente est également impliquée dans la mémorisation à court terme de caractéristiques visuelles non spatiales. Cette aire pourrait constituer un maillon reliant les deux voies de traitement de l'information visuelle, ventrale et dorsale.

Tout au long de ce chapitre, nous avons également identifié les référentiels dans lesquels les principales aires oculomotrices encodent une information spatiale, qu'elle soit sensorielle ou motrice. En particulier, le colliculus supérieur, les champs oculomoteurs frontaux, le cortex pariétal ainsi que le cortex visuel contiennent des cellules dont le champ récepteur ou de mouvement est défini relativement à la position courante des yeux. Nous avons également vu que des structures comme les champs oculomoteurs supplémentaires ou le cortex préfrontal dorsolatéral semblent utiliser un encodage plus complexe. En particulier, on trouve dans SEF des neurones dont le champ récepteur est défini dans un référentiel centré-tête, centré-œil ou centré-corps. Quel peut être l'intérêt d'avoir recours à des représentations plus évoluées si les structures sensorielles et motrices encodent une information spatiale dans un référentiel centré-œil. L'un des intérêts pourraient résider dans la mémorisation d'une information spatiale, qui, comme nous l'avons vu, impliquerait le cortex préfrontal dorsolatéral, tout au moins pour la mémorisation à court terme. Si une information spatiale est mémorisée dans un référentiel centré-œil, cette représentation doit nécessairement être mise à jour avec l'exécution d'un mouvement oculaire pour qu'elle reste cohérente. C'est une première possibilité. Une deuxième repose sur une mémoire spatiale dans un référentiel indépendant de la position des yeux. Ce référentiel a l'avantage de ne pas nécessiter de mise à jour avec le déplacement des yeux. La manière dont le cerveau transforme et mémorise une information spatiale reste un problème difficile. Nous avons présenté, à la fin du chapitre, deux modèles qui apportent une réponse à ces deux problèmes : les mécanismes à base de champs de gain ainsi qu'un mécanisme de mise à jour continue d'une mémoire spatiale.

Dans le prochain chapitre, nous proposons d'étudier plus en détails des mécanismes de transformation et de mémorisation d'une information spatiale. Nous étudions en particulier la possibilité qu'il existe une mémoire spatiale à court-terme dans un référentiel centré-tête. Nous étudierons également un mécanisme qui repose sur une mémoire spatiale centrée-œil, mise à jour par anticipation. Comme nous l'avons vu dans ce chapitre, différentes aires corticales ont des activités anticipatrices. C'est notamment le cas de LIP ou de FEF. Concernant FEF, les travaux récents de [Sommer et Wurtz, 2006] montrent qu'il a des afférences du colliculus supérieur. Le colliculus supérieur pourrait ainsi transmettre à FEF une copie d'un mouvement oculaire sur le point d'être exécuté. Du point de vue fonctionnel, cette copie efférente permet de prédire les conséquences sensorielles du mouvement oculaire. C'est un des principes que nous utiliserons dans le prochain chapitre. Enfin nous verrons également comment il est possible d'étendre le mécanisme d'attention visuelle présenté dans le chapitre 2 pour réaliser une tâche d'exploration visuelle avec ou sans saccade.

Chapitre 4

Anticiper pour assurer la cohérence d'une mémoire de travail spatiale

Anticipatory system : A system containing a predictive model of itself and/or its environment, which allows it to change state at an instant in accord with the model's predictions pertaining to a latter instant.

Robert Rosen
Anticipatory Systems, 1985

Introduction

La vision est une extraordinaire illusion. Nous exécutons en moyenne trois saccades par seconde, à des vitesses pouvant atteindre 1000 degrés par seconde. Ceci a au moins deux conséquences sur l'information visuelle qui se projette sur la rétine. D'une part, ces mouvements oculaires conduisent à un déplacement de l'information rétinienne. D'autre part, étant donnée la vitesse de ces mouvements, l'image rétinienne est floue durant la saccade. Pour être plus exact, il faudrait également prendre en considération la distribution non-uniforme des photorécepteurs sur la rétine, l'absence de photorécepteurs dans la région où le nerf optique quitte la rétine (la fameuse tâche aveugle) ou encore la magnification corticale (qui décrit la proportion de cellules du cortex visuel dont le champ récepteur couvre une région de l'espace visuel). Nous avons malgré tout l'expérience d'un monde visuel stable, continu et perçu dans ses moindres détails. Dans ce chapitre, nous allons présenter des modèles des mécanismes sous-jacents à l'apparente stabilité du monde visuel. C'est un problème intéressant car il permet d'appréhender la manière dont le cerveau représente l'environnement.

Dans le cas des saccades oculaires, seules trois informations peuvent contribuer à la stabilité du monde visuel [Wurtz, 2008] :

- la réafférence visuelle [vonHolst et Mittelstaedt, 1950], qui correspond aux conséquences perceptives d'une action ;
- la proprioception, qui informe sur la position actuelle des yeux ;
- la décharge corollaire ou copie efférente d'un mouvement oculaire imminent. [vonHolst et Mittelstaedt, 1950, Sperry, 1950]

La réafférence visuelle correspond aux informations visuelles qui indiquent le déplacement de la scène visuelle. En particulier, le flot optique, duquel peuvent être extraites la vitesse et la direction du déplacement de la scène visuelle, pourrait contribuer à l'illusion d'un monde stable. Il est probable que ces signaux visuels interviennent dans le contrôle des mouvements de poursuite lente. Néanmoins, dans le cas des saccades oculaires, puisque ces mouvements sont très rapides, le temps d'intégration et la latence de l'information visuelle sont probablement trop importants pour que la réafférence visuelle puisse influencer sur la stabilité apparente de l'expérience visuelle.

La seconde information est la proprioception qui indique la position courante des yeux. De nombreuses études montrent son influence sur la réponse de cellules dans le cortex visuel [Galletti et Battaglini, 1989], le cortex pariétal postérieur [Andersen et al., 1990b] ou encore le cortex frontal [Cassanello et Ferrera, 2007, Schlag et al., 1992, Funahashi et Takeda, 2002], ce qui conduit aux propriétés de champs de gains qu'on a mentionnées au chapitre précédent. La proprioception pourrait ainsi contribuer aux transformations et aux représentations de l'information visuelle spatiale en participant à la construction de représentations indépendantes de la position des effecteurs (e.g. une représentation centrée-tête intégrant les informations visuelles obtenues dans un référentiel centré-œil).

Néanmoins, puisque la proprioception n'est disponible qu'une fois le mouvement oculaire exécuté, elle ne peut pas expliquer les réponses par anticipation observées par exemple dans le cortex pariétal postérieur [Duhamel et al., 1992], les champs oculomoteurs frontaux [Umeno et Goldberg, 1997] ou le colliculus supérieur [Walker et al., 1995], qui pourraient jouer un rôle dans la construction de représentations spatiales cohérentes malgré les saccades oculaires.

Enfin, la dernière source d'information provient d'une décharge corollaire ou d'une copie de la commande qui met en mouvement les yeux. Selon ce principe, introduit par [vonHolst et Mittelstaedt, 1950, Sperry, 1950] (les premiers auteurs introduisent le terme de copie efférente et les seconds le terme de décharge corollaire), une copie de la commande motrice est envoyée à d'autres régions du cerveau pour les informer de l'exécution d'un mouvement. Récemment, les travaux de [Sommer et Wurtz, 2006] ont permis d'identifier des projections entre le colliculus supérieur et les champs oculomoteurs frontaux qui transportent un tel signal. En inactivant des cellules relais du noyau médiodorsal du thalamus, les auteurs observent une baisse des performances dans des séquences de saccades mémorisées ainsi qu'une diminution des propriétés anticipatrices des cellules de FEF enregistrées. Cette diminution n'est pas totale, et comme le soulignent les auteurs, puisque que les cellules motrices de FEF déchargent sélectivement peu de temps avant l'exécution d'une saccade, elles peuvent également avoir accès localement aux paramètres de la saccade qui va être exécutée.

Dans ce chapitre, on commence par introduire un mécanisme d'intégration non-linéaire des

entrées des neurones qu'on applique au calcul de transformations d'information spatiale dans deux contextes. On commence par explorer la possibilité que la proprioception puisse être combinée avec une mémoire spatiale centrée-tête, de plus haut niveau qu'une mémoire centrée-œil, pour réaliser des tâches de séquences de saccades mémorisées. Puis on étudie un mécanisme qui n'utilise qu'une mémoire centrée-œil pour réaliser un parcours oculaire, mise à jour à l'aide d'une copie efférente d'une saccade sur le point d'être exécutée. Enfin, dans la dernière partie, le mécanisme de mise à jour de la mémoire centrée-œil est intégré au mécanisme de déploiement de l'attention visuelle présenté au paragraphe 2.4.3 pour obtenir un mécanisme qui utilise à la fois l'attention guidée par les attributs visuels et l'attention spatiale avec ou sans saccade oculaire.

4.1 Mécanisme sigma-pi pour les transformations de représentations spatiales

Nous avons vu dans le chapitre précédent comment réaliser des transformations sensorimotrices en utilisant le principe des champs de gain. Dans le calcul de transformations sensorimotrices, le champ de gain est une couche cachée, dans un réseau à trois couches. Cette couche cachée peut être relativement grande. En principe, pour deux entrées de taille $N \times N$, le champ de gain doit contenir N^4 cellules pour que chaque cellule de cette couche représente un couple de deux entrées. Cette contrainte sur les dimensions de la couche cachée peut conduire à un coût très important en ressources de calcul. Dans ce paragraphe, on présente un mécanisme qui permet de calculer une transformation sensorimotrice (à comparer aux champs de gains grâce auxquels on peut calculer n'importe quelle combinaison linéaire des informations codées en entrée) en se passant d'une couche intermédiaire. La transformation sensorimotrice est ici directement *câblée* et déterminée par la connectivité des couches d'entrées avec la couche de sortie. Nous allons considérer des transformations non-linéaires qui ne peuvent pas être calculées par un réseau sans couche cachée avec des neurones qui intègrent linéairement leurs entrées. Nous allons donc utiliser des neurones qui intègrent non-linéairement leurs entrées : des neurones sigma-pi [Rumelhart et al., 1987], dont le principe est illustré sur la figure 4.1.

Ces neurones intègrent leurs entrées en calculant la somme de produits des activités des cellules afférentes. C'est à dire que le résultat I_i de l'intégration des entrées du neurone sigma-pi i est de la forme :

$$I_i = \sum_j w_{ij} \prod_{k \in E_j} u_k \quad (4.1)$$

où u_k est l'activité du neurone k et E_j l'ensemble des unités impliquées dans la conjonction du j -ième lien.

Nous nous limitons par la suite à une définition plus restrictive qui n'implique que des conjonctions de deux unités. Dans le paragraphe 4.1.1 on propose une étude analytique d'une forme de connectivité que nous allons utiliser par la suite. Cette étude analytique suppose que l'espace est continu, ce qui facilite les calculs. Dans le paragraphe 4.1.2 on présente une illustration de l'utilisation de ce mécanisme mais cette fois-ci dans le cas discret.

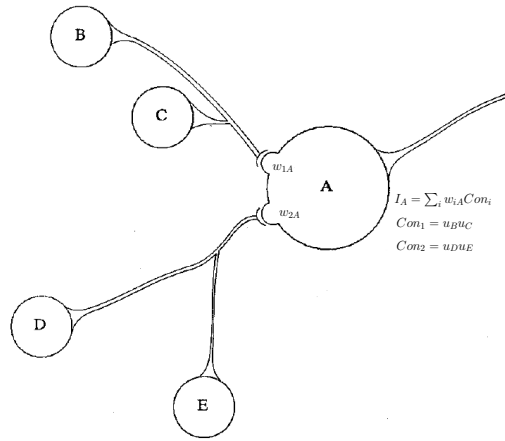


FIG. 4.1 – Schéma de principe d'un neurone sigma-pi. Un neurone sigma-pi intègre non-linéairement ses afférences. L'intégration de ses entrées est en effet réalisée en calculant la somme de la contribution de chaque synapse, où la contribution de chaque synapse est le produit de l'activité des cellules présynaptiques. Seules des synapses conjonctives à deux afférences sont représentées sur cette illustration mais on pourrait imaginer des conjonctions impliquant plus de cellules. L'équation 4.1 constitue, à cet égard, une définition plus générale d'unité sigma-pi. Adapté de [Rumelhart et al., 1987]

4.1.1 Etude analytique du cas 2D continu

Soient α, β, δ des réels non nuls, I, i, c des fonctions de \mathbb{R}^2 dans \mathbb{R} avec I définie par :

$$\forall \vec{x} \in \mathbb{R}^2, I(\vec{x}) = \int i(\vec{y}) c(\alpha \vec{x} + \beta \vec{y} + \delta) d\vec{y} \quad (4.2)$$

On suppose que les fonctions se comportent suffisamment bien pour que cette intégrale existe. Pour simplifier les notations, on note :

$$\iint f(x, y) dx dy = \int f(\vec{x}) d\vec{x} \quad (4.3)$$

Faisons l'hypothèse que la fonction c est symétrique en $\vec{x}_0 \in \mathbb{R}^2$, de telle sorte qu'on peut définir une fonction d symétrique en 0 :

$$\forall \vec{x} \in \mathbb{R}^2, d(\vec{x}) = c(\vec{x} + \vec{x}_0) \quad (4.4)$$

On peut alors réécrire l'équation 4.2 :

$$\begin{aligned}
\forall \vec{x} \in \mathbb{R}^2, I(\vec{x}) &= \int i(\vec{y}) c(\alpha \vec{x} + \beta \vec{y} + \delta) d\vec{y} \\
&= \int i(\vec{y}) d(\alpha \vec{x} + \beta \vec{y} + \vec{\delta} - \vec{x}_0) d\vec{y} \\
&= \int i(\vec{y}) d\left(\alpha \left(\vec{x} + \frac{\beta}{\alpha} \vec{y} + \frac{\vec{\delta} - \vec{x}_0}{\alpha}\right)\right) d\vec{y} \\
&= \frac{\alpha^2}{\beta^2} \int i\left(-\frac{\alpha}{\beta} \vec{u} - \frac{\vec{\delta} - \vec{x}_0}{\beta}\right) d(\alpha(\vec{x} - \vec{u})) d\vec{u} \quad (4.5)
\end{aligned}$$

Le changement de variable est $\vec{u} = -\frac{\beta}{\alpha}\vec{y} - \frac{\vec{\delta}-\vec{x}_0}{\alpha}$. On va maintenant chercher à décomposer le calcul 4.5 pour comprendre ce qu'il réalise.

L'équation 4.5 est un produit de convolution entre les fonctions i_0 et d_0 définies par :

$$\begin{aligned} i_0(\vec{x}) &= i\left(-\frac{\alpha}{\beta}\vec{x} - \frac{\vec{\delta}-\vec{x}_0}{\beta}\right) \\ d_0(\vec{x}) &= d(\alpha\vec{x}) \end{aligned}$$

Si la fonction d est une gaussienne (ce que nous utiliserons par la suite), alors d_0 est cette même gaussienne dont la variance est divisée par $\frac{1}{\alpha}$.

La transformation $\vec{u} \rightarrow -\frac{\alpha}{\beta}\vec{u} - \frac{\vec{\delta}-\vec{x}_0}{\beta}$ est une transformation du plan. On considère le cas particulier $-\frac{\alpha}{\beta} = 1$. Dans ce cas, la transformation du plan est simplement une translation de vecteur $\frac{\vec{\delta}-\vec{x}_0}{\beta}$.

Ainsi, le calcul 4.5 peut se résumer en trois étapes :

- Translater dans l'espace la fonction i ;
- Convolver le résultat par une gaussienne ;
- Multiplier le résultat par un facteur d'échelle $\frac{\alpha^2}{\beta^2} = 1$ puisque nous avons fait l'hypothèse que $-\frac{\alpha}{\beta} = 1$.

sous les conditions $-\frac{\alpha}{\beta} = 1$ et c est une fonction symétrique en un certain x_0 .

Nous allons maintenant voir le lien avec les transformations sensorimotrices. Dans les applications que nous allons considérer, la fonction i représente une mémoire spatiale 2D, la fonction c représente une consigne motrice de déplacement oculaire. Nous faisons l'hypothèse que la commande motrice est une gaussienne centrée sur le vecteur de déplacement des yeux. Nous utilisons alors l'équation 4.5 pour définir les poids qui lient les neurones d'entrées qui représentent i et c avec la couche de sortie. D'après le calcul, il apparait qu'avec ce profil de connexion, les activités dans la couche de sortie sont les activités de la couche d'entrée i déplacées d'une consigne définie par une gaussienne c . Il faut légèrement nuancer ce résultat puisqu'il ne faut pas oublier qu'en plus de la translation, il reste la convolution des deux signaux qui peut altérer le résultat. Or la convolution d'un signal par une gaussienne atténuée les hautes fréquences du signal convolué ; c'est une forme de filtre passe bas. Plus la variance de la gaussienne est petite, moins le filtre passe bas est sélectif, ce qui revient à dire que plus la gaussienne définissant la consigne est étroite, moins le signal d'entrée sera altéré par la convolution. Nous utiliserons ainsi ce mécanisme pour translater un profil d'activité étant donnée une consigne. Pour l'illustrer, on propose dans le paragraphe suivant un exemple concret de transformation.

4.1.2 Illustration dans le cas 1D discret

On considère dans ce paragraphe une architecture constituée de trois cartes :

- une carte d'entrée de N neurones qui représentent un profil d'entrée ;
- une carte de consigne de N neurones qui représentent une consigne définie par une gaussienne centrée en x_0 (qui varie au cours du temps) ;
- une sortie de N neurones qui contient des neurones sigma-pi.

Les neurones sigma-pi intègrent leurs entrées d'après l'équation 4.2 avec $\alpha = 1, \beta = -1, \delta = \frac{N}{2}$ ce qui conduit à une translation de vecteur $\frac{N}{2} - x_0$.

La figure 4.2 illustre le mécanisme. Les activités des trois cartes sont représentées dans des graphiques espace/temps. L'entrée est constante et constituée de trois gaussiennes. La consigne est représentée par une gaussienne dont le centre x_0 décrit une sinusoïde autour du centre de la population en $\frac{N}{2}$. On constate que les activités de sortie sont les activités d'entrée (on retrouve bien les trois gaussiennes) translattées d'un vecteur défini par la consigne.

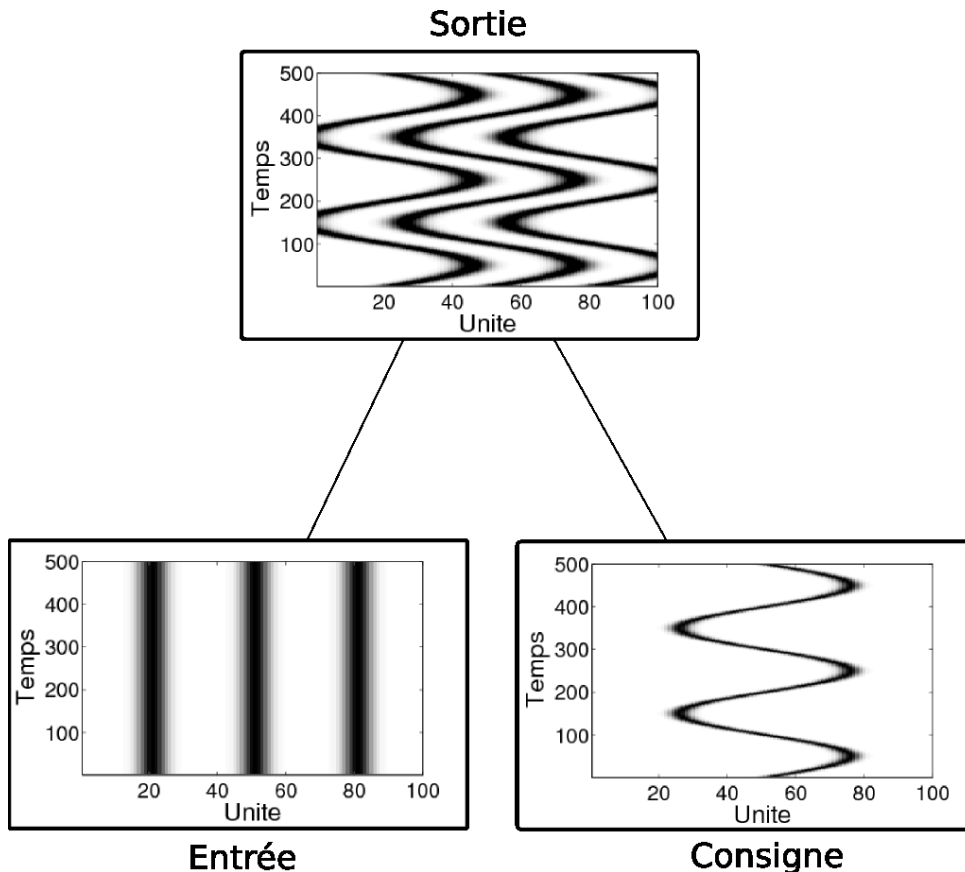


FIG. 4.2 – Les activités des trois cartes de 100 cellules sont représentées dans des graphiques espace/temps. L'entrée est constante et constituée de trois gaussiennes. La consigne est une gaussienne dont le centre décrit une sinusoïde. La carte de sortie contient des neurones sigma-pi qui intègrent leurs entrées selon l'équation 4.2 avec $\alpha = 1, \beta = -1, \delta = \frac{N}{2}$. Le profil d'activité de la sortie est alors le profil d'entrée translatté d'un vecteur défini par la consigne.

4.1.3 Conclusion

En conclusion, nous venons de présenter un mécanisme qui permet de réaliser des transformations d'information spatiale. Il réalise des opérations similaires aux champs de gain présentés dans le chapitre précédent, à savoir la combinaison linéaire d'informations codées en entrée. Il est néanmoins important de remarquer que même si la transformation est linéaire par rapport aux variables codées en entrée, le calcul sous-jacent est, lui, non-linéaire. Il implique en effet de combiner des représentations de ces variables. Puisque les variables sont en général codées par des populations dont la réponse est une fonction non-linéaire de la variable à coder (une gaussienne, une sigmoïde), la transformation linéaire dans l'espace des variables codées est non-linéaire dans l'espace neuronal.

Les mécanismes à base de champs de gain nécessitent une couche cachée, à savoir le champ de gain à proprement parler, dont les dimensions peuvent être importantes. De cette représentation, qui combine tous les neurones des couches d'entrée, on peut décoder n'importe quelle combinaison des variables d'entrée, ce qui conduit à une certaine forme de généralité quant à la transformation calculée. On peut voir le mécanisme que nous venons de présenter comme une restriction des champs de gain à une transformation en particulier. En se limitant aux cellules du champ de gain qui sont impliquées dans la transformation recherchée, on peut directement définir la transformation dans les projections entre les couches d'entrée et la couche de sortie. L'avantage est bien évidemment de s'abstraire d'une représentation intermédiaire coûteuse mais le revers de la médaille est que l'architecture considérée est plus ad-hoc que les champs de gain. Néanmoins c'est une restriction qui n'est pas forcément un inconvénient.

Ce mécanisme sigma-pi est au coeur de toutes les transformations sensorimotrices utilisées par la suite que ce soit pour calculer des changements de référentiel centré-œil vers centré-tête ou l'inverse (paragraphe 4.2), ou bien anticiper les conséquences d'un mouvement oculaire sur l'état d'une mémoire de travail spatiale rétinotopique (paragraphe 4.3).

4.2 Saccades vers des cibles mémorisées dans un référentiel extra-rétinien

Il existe différents paradigmes qui soulignent l'importance de la mémoire spatiale dans la sélection d'une cible saccadique. En particulier, dans les séquences de saccades *double-step task*, deux cibles sont présentées brièvement et le sujet doit exécuter deux saccades vers les cibles mémorisées, dans le même ordre que leur présentation (cf figure 3.5d). Il y a au moins deux référentiels dans lesquels les cibles d'une saccade peuvent être mémorisés : dans un référentiel oculo-centré ou dans un référentiel extra-rétinien. Si on considère que les cibles sont mémorisées par rapport à la position courante de l'œil, sa mise à jour est nécessaire après chaque mouvement oculaire et nous verrons dans le paragraphe 4.3 comment cela peut être réalisé. Nous allons considérer dans un premier temps que les cibles sont mémorisées dans un référentiel extra-rétinien, centré par rapport à la tête, et ainsi indépendant de la position actuelle des yeux. Cette représentation étant indépendante de la position courante de l'œil (nous reviendrons sur cette hypothèse à la fin du paragraphe), il n'y a pas besoin de la mettre à jour avec les mouvements

oculaires. Néanmoins, puisque la sélection semble s'opérer dans un référentiel centré-œil (FEF et SC codent la cible d'une saccade oculaire dans un référentiel oculocentré), il est nécessaire de disposer d'un mécanisme qui puisse transformer la position des cibles mémorisées dans un référentiel centré-tête, dans le référentiel centré-œil de la sélection. Le mécanisme présenté dans le paragraphe précédent peut réaliser cette opération en combinant la mémoire centrée-œil avec la position actuelle de l'œil fournie par la proprioception.

4.2.1 Architecture du mécanisme proposé

Le mécanisme proposé pour réaliser la tâche de double saccades mémorisées est présenté sur la figure 4.3. On suppose qu'il existe une représentation spatiale de la pertinence comportementale de chaque position de l'espace visuel dans un référentiel rétinotopique. Comme nous l'avons vu au chapitre précédent, cette représentation pourrait correspondre aux représentations observées dans le cortex pariétal postérieur. Par ailleurs, nous avons également vu que les neurones des aires LIP et 7a du pariétal postérieur ont la propriété de champs de gain (champ récepteur rétinotopique, modulé par la position de l'œil) qui permet de décoder de cette représentation n'importe quelle combinaison linéaire de la position de l'œil et de la position rétinotopique d'un stimulus visuel. Pour le mécanisme proposé, on considère que deux représentations sont décodées : une transformation centrée-œil vers centrée-tête par l'intermédiaire des projections de LIP¹⁷ sur SEF et une transformation centrée-œil vers centrée-œil par l'intermédiaire des projections de LIP sur FEFV. Les neurones de la carte SEF possèdent des connexions récurrentes avec la carte DLPFC. Ces projections constituent pour le modèle une mémoire de travail spatiale dans un référentiel centré-tête.

La carte FEFVM reçoit deux afférences. D'une part les afférences de FEFV lui fournissent la position rétinotopique des cibles visuelles. D'autre part, les projections de SEF lui fournissent la position des cibles mémorisées dans un référentiel centré-tête. Puisque les cibles sont mémorisées dans un référentiel centré-tête et puisque FEFVM travaille dans un référentiel centré-œil, il est nécessaire de réaliser une transformation centré-tête vers centré-œil. Pour ce faire, on utilise un mécanisme sigma-pi, similaire à celui présenté au paragraphe 4.1, modulé par la position courante de l'œil fournie par la proprioception qui est représentée dans la carte EYE POSITION. Nous utilisons dans ce mécanisme une représentation explicite en 2D de la position de l'œil. Il existe pour le moment peu de résultats concernant la représentation de la position de l'œil dans le cortex bien qu'on connaisse son influence notamment pour les propriétés de champs de gain. L'étude récente de [Wang et al., 2007] montre qu'une région du cortex somatosensoriel pourrait représenter la position de l'œil. Le codage utilisé par les cellules identifiées est un codage monotone en fonction de l'excentricité de l'œil. Dans l'annexe D nous montrons empiriquement qu'il est possible de construire une représentation explicite 2D de la position de l'œil étant donné un codage monotone 1D. La transformation est *presque* linéaire en ce sens que, dans l'architecture considérée en annexe, les neurones qui représentent explicitement la position de l'œil intègrent linéairement la réponse des cellules des populations monotones. Seule une non-linéarité de la fonction de transfert est introduite pour que la transformation fonctionne. Ce résultat présenté

¹⁷On notera LIP lorsqu'on fait référence à une carte dans le mécanisme et LIP lorsqu'on fait référence à l'aire corticale

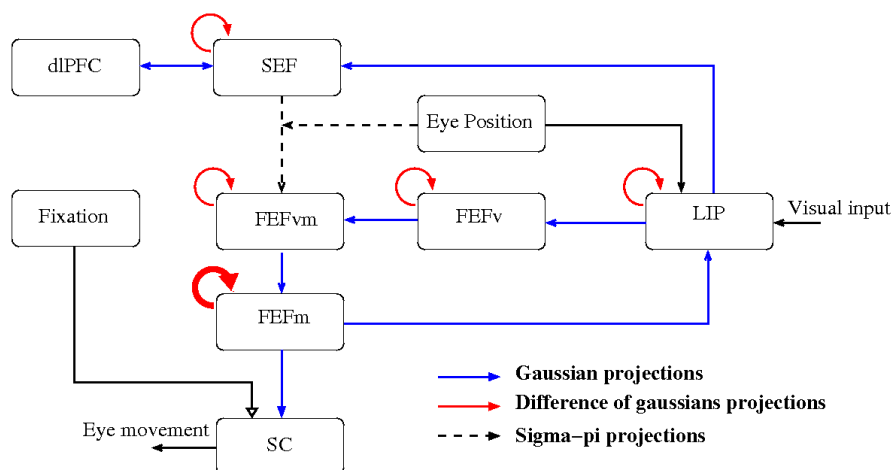


FIG. 4.3 – L’information visuelle spatiale, représentée dans la carte LIP se projettent directement sur FEF et indirectement via la mémoire de travail construite à l’aide du circuit récurrent SEF-DLPFC. La mémoire de travail, en centré-tête, fournit à FEFVM la position mémorisée des cibles, en combinant sa représentation avec la position actuelle de l’œil représentée dans la carte EYE POSITION. La sélection de la prochaine cible saccadique s’opère dans la carte FEFM grâce à des connexions latérales en différence de gaussiennes qui induisent de la compétition entre les unités. La cible sélectionnée se projette sur la carte motrice SC, qui est également sous le contrôle inhibiteur des cellules de FIXATION. Les paramètres de la simulation sont disponibles dans l’annexe A.3

en annexe contribue à l’idée qu’il peut être possible de substituer la représentation 2D explicite à des représentations 1D monotone sans changer l’architecture.

La sélection de la prochaine cible saccadique a lieu dans FEFVM par l’intermédiaire de connexions latérales excitatrices localement et inhibitrices sur des plus longues distances. Nous avons vu au chapitre 2 que ce mécanisme permet d’obtenir un comportement similaire à un *winner-take-all*, mais de façon distribuée en mettant les unités en compétition. Lorsqu’une cible est sélectionnée dans FEFVM, elle se projette sur FEFM. Les projections efférentes de FEFM ciblent deux cartes : LIP et SC. Les premières projections, sur LIP, modélisent l’influence de l’attention spatiale sur les premiers niveaux de traitement de l’information spatiale. Ces projections n’ont pas de rôle fonctionnel dans le mécanisme mais permettent simplement de reproduire la modulation observée lors du déploiement couvert ou ouvert de l’attention spatiale. Enfin la carte SC représente la carte motrice du modèle. Elle est excitée par FEFM, qui lui fournit la cible d’une saccade, et inhibée par des cellules de fixation qui déterminent quand la saccade doit être exécutée.

Les projections entre les cartes sont des gaussiennes et la plupart des cartes ont des connexions latérales en différence de gaussiennes locales. La seule exception pour la localité des connexions latérales concerne la carte FEFM pour laquelle les connexions latérales couvrent toute la carte pour induire une compétition entre les unités. Enfin, la plupart des unités utilisées sont des unités qui intègrent linéairement leurs entrées à l’exception de la carte FEFVM qui a des afférences sigma-pi des cartes SEF et EYE POSITION afin de calculer la transformation centré-tête vers centré-œil.

Tous les paramètres du mécanisme sont disponibles dans l'annexe A.3

4.2.2 Simulations et résultats

Les figures 4.4 et 4.5 représentent l'évolution du modèle pendant l'exécution d'une tâche de double saccades mémorisées. Deux cibles sont présentées et leur positions spatiales se propagent le long des projections directes de LIP sur FEF et le long des projections indirectes, via la mémoire de travail formée par le circuit récurrent SEF-DLPFC (figures 4.4a,b). Initialement, les yeux sont centrés, ce qui est représenté par la bulle d'activité centrée dans la carte EYE POSITION. Puisque la tâche nécessite initialement de fixer les yeux, les cellules de fixation déchargent toniquement et inhibent activement la carte motrice SC.

Lorsque les cibles sont éteintes (figure 4.4c), le sujet doit exécuter deux saccades successives vers les deux cibles mémorisées. Puisque les cibles visuelles ne sont plus présentes, seule la contribution de SEF permet la sélection dans FEFVM de la prochaine cible saccadique. La cible sélectionnée se propage sur la carte motrice qui n'est plus sous l'inhibition tonique des cellules de fixation, ce qui a pour conséquence de déclencher une saccade vers cette cible.

Lorsque le mouvement est terminé (figure 4.4d), la nouvelle position des yeux, représentée dans la carte EYE POSITION, permet de réaliser la mise à jour des positions centrées-œil des cibles. Elle est en effet combinée avec la mémoire centrée-tête, indépendante de la position des yeux, pour fournir à FEFVM la position des cibles mémorisées relativement à la nouvelle position des yeux (figure 4.5a). La prochaine cible saccadique est sélectionnée dans FEFVM et se propage sur la carte motrice, ce qui déclenche un mouvement saccadique.

Une fois le mouvement terminé, la nouvelle position des yeux, combinée avec la mémoire de travail, permet à nouveau de mettre à jour la position des cibles mémorisées dans la représentation centrée-œil de FEFVM (figures 4.5b,c).

Afin de mettre l'accent sur l'évolution temporelle des activités au sein des différentes cartes, nous proposons de réaliser la même expérience tout en enregistrant l'activité de cellules aux cinq positions qui sont potentiellement occupées par un stimulus pendant l'expérience. La figure 4.6a représente les cinq positions spatiales couvertes par les champs récepteurs des cellules enregistrées ainsi que la position qu'occuperaient les cibles visuelles, pendant l'exécution de la tâche, si elles étaient visibles. La figure 4.6b représente l'évolution de l'activité pour cinq cellules enregistrées dans les six cartes INPUT, LIP, FEFVM, FEFM, SEF et SC. Les barres verticales qui apparaissent sur les graphiques représentent respectivement :

- l'instant de l'extinction des cibles ;
- l'instant de l'exécution de la première saccade ;
- l'instant de l'exécution de la seconde saccade.

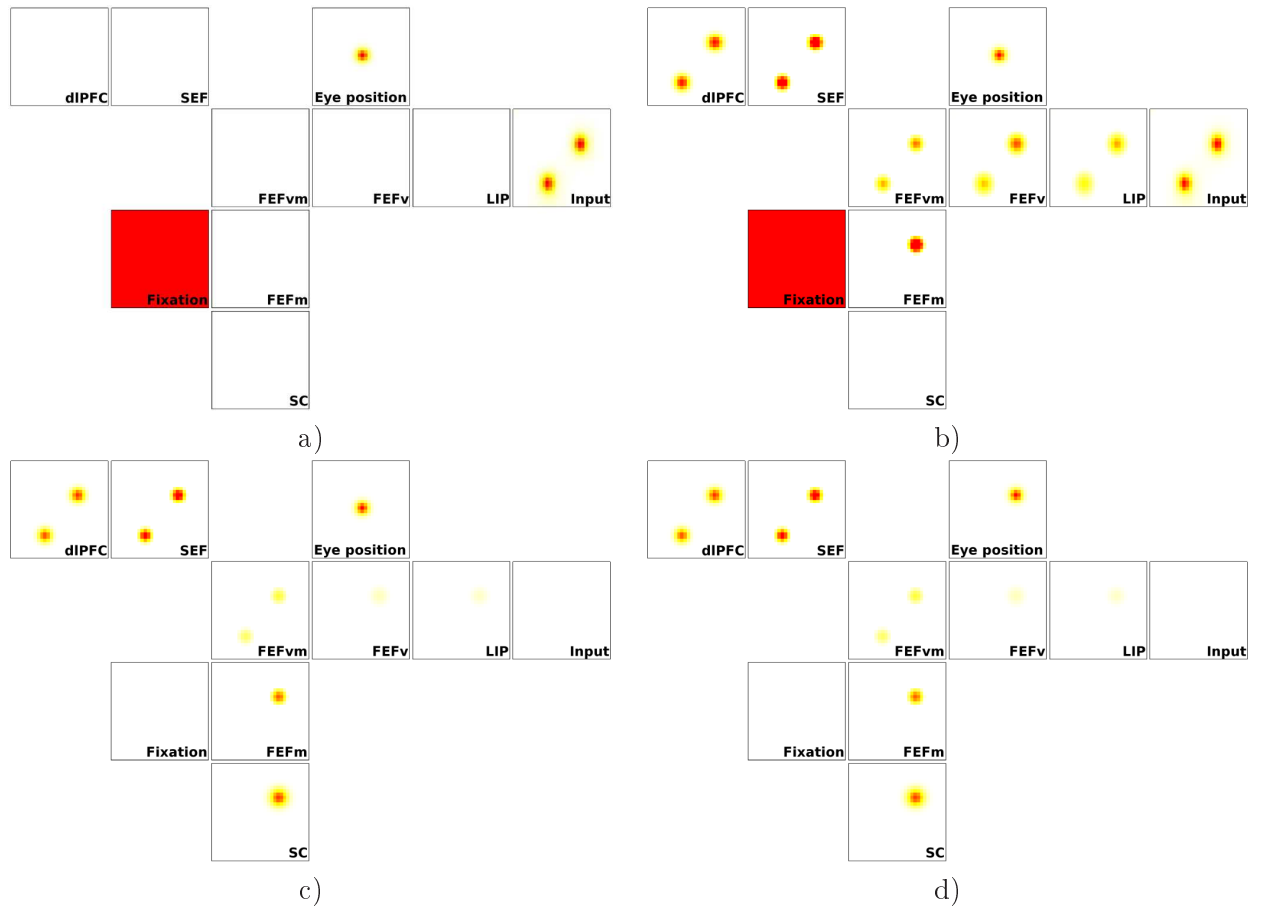


FIG. 4.4 – a) Pendant la fixation, les cellules de fixation déchargent toniquement et inhibent la carte motrice, les yeux sont fixés (caractérisé par la bulle d'activité dans la carte EYE POSITION) et deux cibles visuelles sont présentées. b) L'information visuelle se propage vers FEF et la mémoire de travail formée par le circuit récurrent SEF-DLPFC. Puisque les yeux sont centrés, les positions centrées-œil et centrées-tête des cibles sont alignées. c) Lorsque les cibles sont éteintes, le sujet doit exécuter la tâche. Les cellules de fixation deviennent silencieuses et relâchent leur inhibition tonique sur la carte motrice qui permet la production d'une saccade. d) Immédiatement après la fin de la saccade, la nouvelle position des yeux est représentée dans la carte EYE POSITION mais la position centrée-œil des cibles n'est pas encore mise à jour. La suite de l'évolution du système est représentée sur la figure 4.5

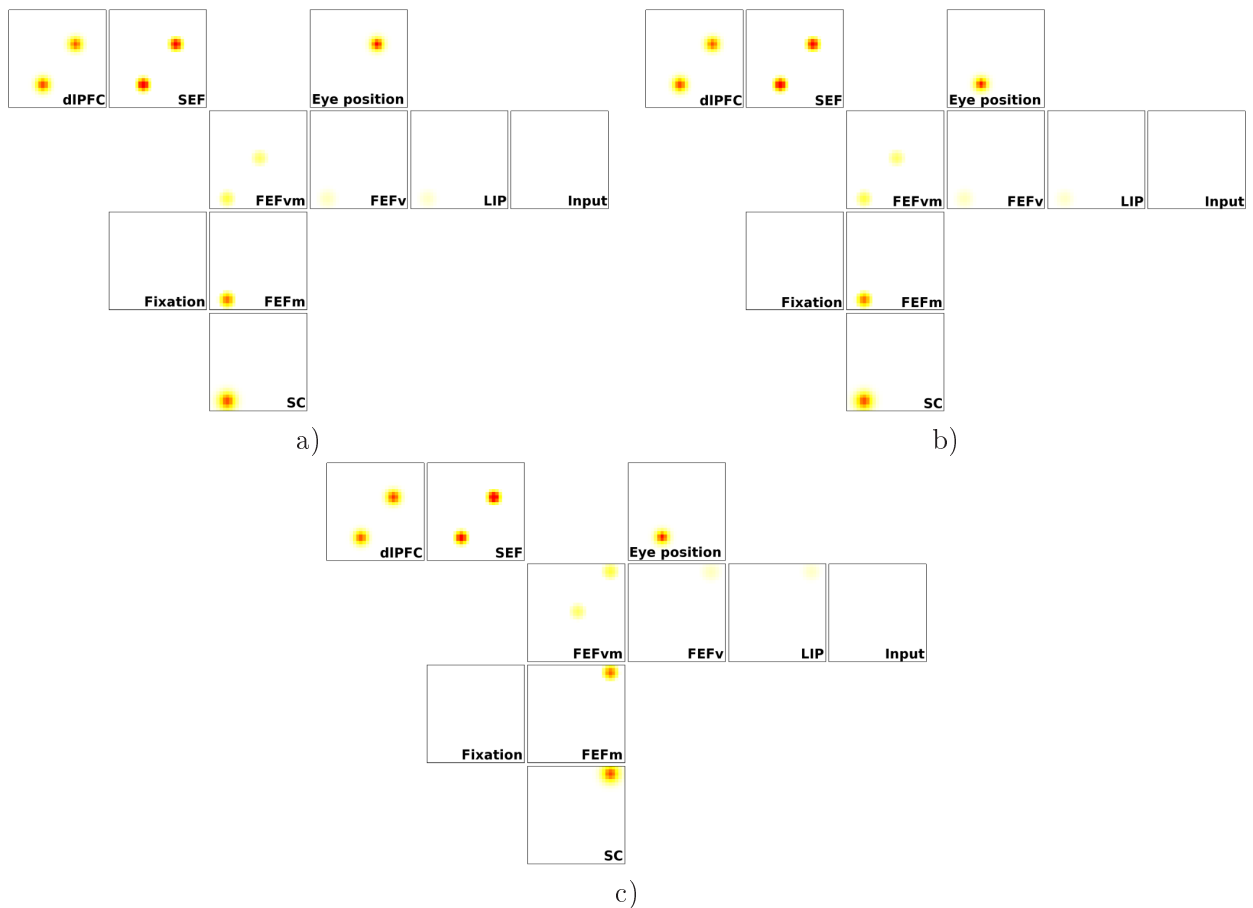


FIG. 4.5 – Suite des illustrations de la figure 4.4. a) Après quelques pas de temps, la représentation rétinotopique de FEFVM est à jour, la seconde cible a gagné la compétition dans FEFM qui excite SC, ce qui a pour conséquence de produire une saccade vers cette cible. b) Une fois la saccade exécutée, la nouvelle position des yeux est disponible. c) La mise à jour de la proprioception conduit à la mise à jour de la représentation rétinotopique de FEFVM.

Le lien avec les illustrations des figures 4.4 et 4.5 est comme suit :

- les figures 4.4a-b correspondent à la simulation précédant l’extinction des cibles ;
- la figure 4.4c correspond à la période entre l’extinction des cibles et l’exécution de la première saccade ;
- les figures 4.4d et 4.5a correspondent aux instants suivant l’exécution de la première saccade et précédant l’exécution de la seconde saccade ;
- les figures 4.5b-c correspondent aux instants suivant l’exécution de la seconde saccade.

Les cibles sont présentées, pendant la fixation, aux positions x_1 et x_3 . L’activité liée à la représentation de ces deux cibles se propage de LIP vers FEFVM et SEF. Dans FEFM, on observe que l’activité en x_3 croît au détriment de l’activité en x_1 , ce qui caractérise la sélection de la cible en x_3 . Initialement les deux cibles sont également représentées dans LIP. Puisque FEFM a des projections sur LIP, la sélection de la cible en x_3 dans FEFM produit un biais en faveur de cette cible dans LIP. On observe ce biais à la fois lorsque les cibles visuelles sont présentes et lorsqu’elles sont éteintes. La décroissance de l’activité dans la voie directe LIP, FEFVM, FEFM, suivant l’extinction des cibles, est la conséquence de l’absence de l’excitation visuelle.

Après l’exécution du mouvement oculaire, la combinaison de la mémoire centrée-tête (représentée en partie dans SEF) avec la nouvelle position de l’œil, fournit à FEFVM les positions mémorisées des cibles relatives à la nouvelle configuration de l’œil, d’où les activités en x_0 et x_2 dans FEFVM. Les afférences excitatrices de FEFVM sur FEFM excitent les cellules dont le champ récepteur couvre les positions des deux cibles, qui entrent alors en compétition. Les cellules qui ont gagné la compétition excitent la carte motrice SC à la position x_0 , ce qui conduit à l’exécution d’une saccade vers cette cible.

Nous sommes restés émusifs sur le mécanisme de compétition. En effet, la compétition pour la sélection d’une cible dans FEFM, qui est la conséquence des connexions latérales au sein de cette carte, n’est pas biaisée pour favoriser la sélection de la seconde cible lorsque la première a déjà été fixée. La compétition telle qu’elle est réalisée dans cette architecture n’est que le fruit du hasard, avec néanmoins quelques effets de bord. En particulier, la victoire de la seconde cible saccadique est la conséquence d’un biais inhibiteur vers la première cible saccadique. En effet, l’activité dans la carte FEFM à la position de la première cible saccadique, sous la perte d’afférence excitatrice, met un certain temps à décroître. Par conséquent, les unités qui étaient excitées continuent à avoir une influence sur les unités voisines par l’intermédiaire des connexions latérales, excitatrices localement et inhibitrices sur de plus longues distances. Or, la première cible saccadique, qui occupe le centre du champ visuel à la suite du mouvement oculaire est justement sous une influence inhibitrice plus forte, rémanente de la sélection de cette cible, comparée à la seconde cible potentielle.

Une solution plus élégante pour la sélection de la prochaine cible saccadique pourrait reposer sur une mémoire des cibles qui ont été fixées. Cette mémoire est a priori distincte de la mémoire centrée-tête que nous utilisons dans le mécanisme. En effet, le circuit récurrent que nous utilisons permet de mémoriser tous les stimuli qui seront la cible d’un mouvement oculaire. Au contraire, la mémoire qu’il faudrait ajouter pour améliorer le système de sélection mémoriserait uniquement les stimuli qui ont déjà été fixés.

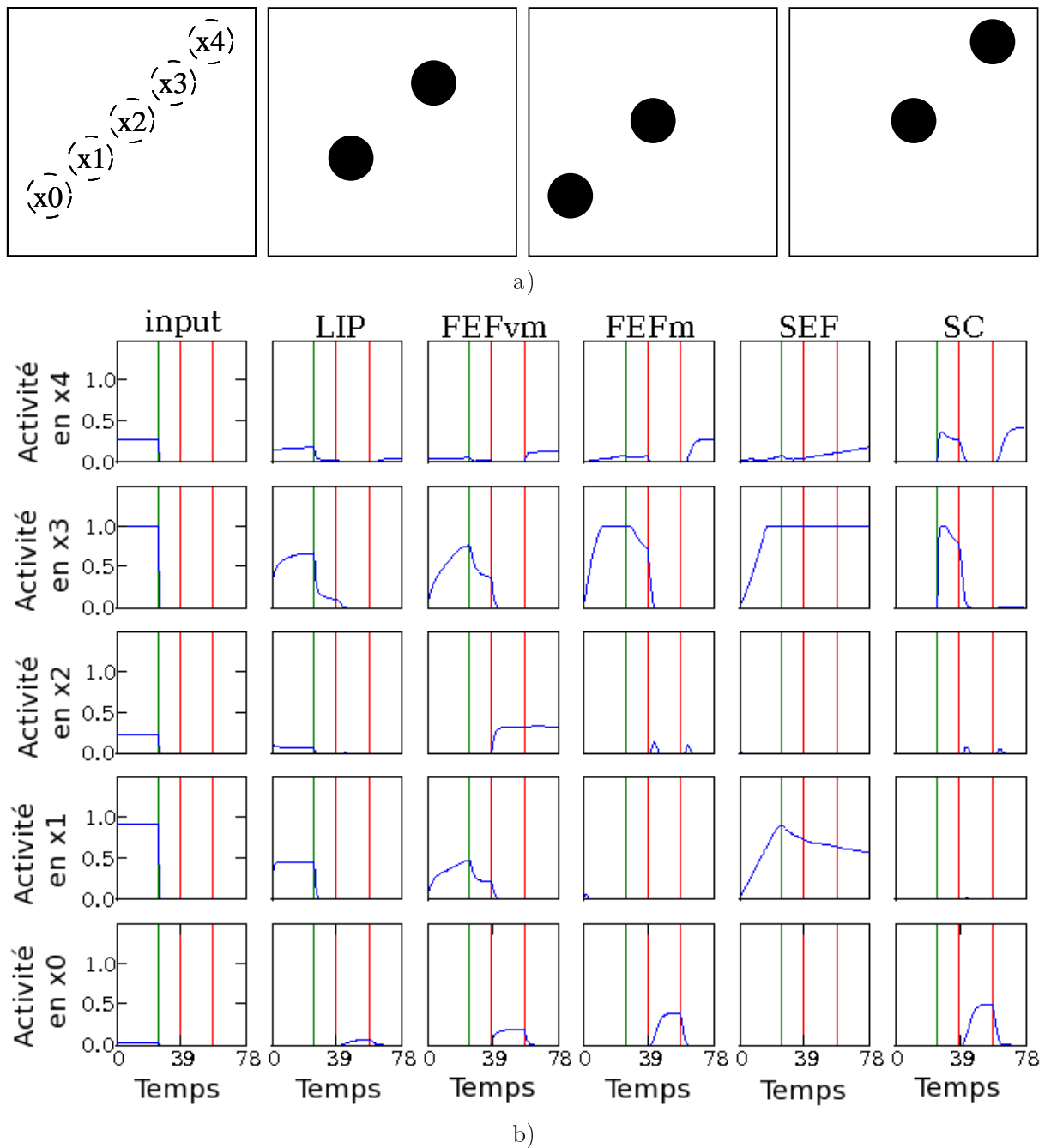


FIG. 4.6 – a) Pendant l'exécution de la tâche de saccades vers deux cibles mémorisées, l'activité de cellules dont le champ récepteur couvre l'une des cinq positions x_0, x_1, x_2, x_3, x_4 sont enregistrées. Ces positions particulières sont potentiellement occupées par une cible à chaque instant du déroulement de la tâche, comme illustré sur la figure de droite qui montre la position que les cibles visuelles occuperaient si elles étaient visibles b) L'activité de ces cellules particulières est enregistrée simultanément dans les six cartes INPUT, LIP, FEFVM, FEFM, SEF et SC. Les barres verticales qui apparaissent sur les graphiques représentent respectivement les instants d'extinction des cibles puis d'exécution de la première et de la seconde saccade. La tâche réalisée est la même que pour les figures 4.4, 4.5 de telle sorte que l'extinction des cibles correspond à l'illustration 4.4c, l'exécution de la première saccade correspond à l'illustration 4.4d et l'exécution de la seconde saccade correspond à l'illustration 4.5b. Pour plus de détails, se référer au texte.

4.2.3 Conclusion

On vient d'illustrer les performances du mécanisme dans la réalisation de la tâche de saccades vers deux cibles mémorisées. Au delà des performances du mécanisme, on aimerait revenir sur la plausibilité biologique de ce mécanisme et sur ses implications en terme de circuits neuronaux impliqués dans le contrôle des saccades oculaires.

La première hypothèse sur laquelle repose le mécanisme est l'existence d'une représentation spatiale rétinotopique de la pertinence comportementale des cibles visuelles. Nous avons vu dans le chapitre précédent que chez les primates, LIP et 7a pourraient être de bons candidats pour une telle représentation. En effet, l'activité de leurs cellules semble dépendre à la fois de facteurs exogènes comme la nouveauté d'un stimulus visuel, et de facteurs endogènes comme la position d'une cible pour un mouvement oculaire.

De plus, les neurones des aires LIP et 7A exhibent la propriété d'un champ de gains avec des champs récepteurs rétinotopiques modulés par la position actuelle de l'œil. Comme nous l'avons vu dans le chapitre précédent, cette propriété de champ de gain permet de décoder toutes les combinaisons linéaires des variables encodées, à savoir la position rétinotopique des stimuli visuels et la position des yeux. Nous avons utilisé cette propriété pour décoder de l'aire LIP modélisée une représentation centrée-œil et une représentation centrée-tête. Pour des raisons de simplicité, cette transformation centré-œil vers centré-tête est réalisée dans le mécanisme avec le même mécanisme sigma-pi utilisé pour calculer la transformation centré-tête vers centré-œil.

Il est montré que les cellules de FEF peuvent être réparties en cellules visuelles, visiomotrices et motrices. Dans l'architecture proposée, on distingue trois couches pour FEF. Les cellules de FEFV ont des réponses purement visuelles, les cellules de FEFM déchargent lorsque la cible d'une saccade oculaire occupe leur champ de mouvement, et les cellules de FEFVM constituent une couche intermédiaire, ayant des afférences visuelles et des afférences liées à la mémorisation, qui fournit à la couche motrice FEFM les cibles potentielles pour un mouvement oculaire.

Différentes études ont montré l'existence d'un continuum de référentiels dans lesquels les neurones de SEF encodent l'information spatiale. Par ailleurs, SEF semble être impliqué à la fois dans l'apprentissage et dans la reconstitution de séquences de saccades oculaires. Dans le mécanisme proposé, l'aire modélisée SEF fournit la position des cibles mémorisées pour une tâche qui implique une séquence de saccades. Bien évidemment, la proposition que nous faisons est simpliste. Tout d'abord, on ne considère qu'une représentation centrée-tête. De plus, puisque SEF est important dans le déroulement d'une séquence de saccades, il est possible que les neurones de cette aire encodent plus que la position spatiale des cibles, en ayant des activités corrélées à l'ordre de la séquence motrice. Ses connexions récurrentes avec dIPFC l'impliquant dans la mémorisation à court terme sont hypothétiques; même si dIPFC est fortement impliqué dans la mémorisation spatiale à court terme, rien ne suggère qu'un circuit récurrent entre dIPFC et

SEF forme le substrat de cette mémoire. Néanmoins, puisque d'une part dIPFC est impliqué dans la mémoire spatiale à court terme et que d'autre part SEF est impliqué dans l'encodage de séquences oculaires, le parti pris du mécanisme est de considérer un circuit récurrent qui code spatialement la position des cibles d'une séquence oculaire.

FEF se projette par deux voies sur le colliculus supérieur. La première par des projections directes et la seconde par des projections indirectes via les ganglions de la base. Nous avons pris le parti de modéliser les projections indirectes, qui pourraient contribuer à la fois à la sélection de la prochaine action motrice et au déclenchement de cette action par désinhibition de la part des ganglions de la base, par des cellules de fixation hypothétiques, qui déchargent toniquement tant que les yeux doivent rester fixes.

Ce mécanisme, qui a fait l'objet d'une publication [Fix, 2007], permet d'illustrer comment la proprioception pourrait contribuer au contrôle des saccades oculaires. Nous avons mentionné en introduction qu'une autre proposition repose sur l'existence d'une décharge corollaire qui permet d'assurer le lien entre les perceptions pré et post-saccadiques. Dans le prochain paragraphe, on propose d'exploiter le fait qu'une saccade est un mouvement propre dont on peut prédire certaines conséquences.

4.3 Saccades vers des cibles mémorisées dans un référentiel centré-œil : rôle de l'anticipation

4.3.1 Introduction

Nous avons vu dans le paragraphe précédent qu'une solution pour réaliser des saccades vers des cibles mémorisées consiste à mémoriser la position des cibles dans un référentiel centré-tête et à combiner cette mémoire avec la position courante des yeux pour fournir aux aires oculomotrices la position des cibles mémorisées par rapport à la position des yeux. On propose dans ce paragraphe d'explorer une autre solution qui consiste à utiliser une mémoire centrée-œil, mise à jour à l'aide d'une "copie" de la commande motrice.

Il y a plusieurs raisons qui nous poussent à explorer cette possibilité. La première est qu'étant donné que les principales aires oculomotrices ont des cellules à champ récepteur, ou champ de mouvement, rétinotopiques, il est plus simple de leur fournir la position des cibles relativement à la position courante des yeux. Par ailleurs, nous ne travaillons pour le moment qu'avec des mouvements oculaires en supposant le reste du corps fixe. Dans ce contexte, la position des cibles dans un référentiel centré-tête ne bouge pas, de telle sorte qu'une simple combinaison avec la position courante des yeux nous permet de retrouver facilement la position des cibles par rapport à la position courante des yeux. Le problème est plus compliqué si la tête est également en mouvement. En effet, dans ce cas, la représentation centrée-tête doit également être mise à jour. Si on conserve le principe exposé précédemment (selon lequel la mise à jour de la représentation de cibles par rapport à un effecteur est réalisée à partir d'une représentation indépendante de la position de cet effecteur), cette mise à jour pourrait être réalisée à partir d'un référentiel

centré-corps. Et si le corps est en mouvement et que l'on suit toujours le même principe, on en arrive à devoir définir un référentiel centré-monde en passant de représentations égocentrées à une représentation allocentrée pour mémoriser les cibles et de la représentation allocentrée aux représentations égocentrées pour définir la cible des mouvements des effecteurs. Une telle représentation mnésique allocentrée semble être utilisée dans l'hippocampe, en particulier pour la navigation [Rougier, 2000]. Néanmoins cette structure semble plus impliquée dans la mémoire à long terme que dans la mémoire à court terme. L'objectif de ce paragraphe n'est pas d'exclure la possibilité que le cerveau ait accès à une représentation allocentrée mais plutôt d'explorer la faisabilité d'une mémoire, dans un référentiel lié à l'effecteur, qui puisse être mise à jour avec les mouvements exécutés.

Le principe du mécanisme repose sur l'anticipation des conséquences d'une action sur la représentation des cibles mémorisées. La proposition que l'anticipation puisse jouer un rôle dans la mise à jour de la mémoire de travail spatiale est confortée par les travaux de [Sommer et Wurtz, 2006, Sommer et Wurtz, 2004a, Sommer et Wurtz, 2004b] qui soulignent l'importance d'un signal colliculaire, transmis aux champs oculomoteurs frontaux, dans les tâches de saccades vers des cibles mémorisées et que nous avons mentionné au paragraphe 3.3.3 du chapitre précédent.

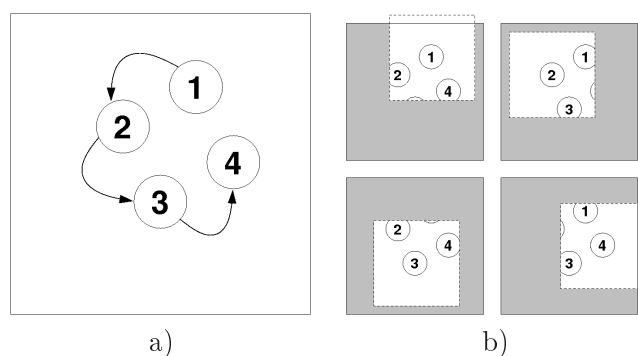


FIG. 4.7 – a) La tâche consiste à exécuter successivement quatre saccades, une vers chaque cible, sans jamais exécuter deux saccades vers la même cible. Les chiffres n'apparaissent sur les cibles que pour les besoins de l'explication de telle sorte que le seul critère permettant de discriminer les cibles est leur position spatiale. b) Pendant l'exécution de la tâche, la position des cibles change radicalement. Une des difficultés dans cette situation est d'être capable de se souvenir quelles cibles ont déjà été fixées.

La figure 4.7 illustre l'une des difficultés de mémoriser les cibles dans un référentiel centré-œil. La tâche requiert de la part du mécanisme d'exécuter quatre saccades, une en direction de chaque objet, peu importe l'ordre, mais il ne doit pas exécuter deux saccades vers la même cible. Les chiffres qui apparaissent sur les cibles de la figure 4.7 ne sont là que pour indiquer l'une des séquences motrices possibles. Toutes les cibles étant identiques, la seule information qui permet de les discriminer est leur position spatiale. La figure 4.7b représente le déplacement du champ visuel causé par chaque saccade. En particulier, après avoir fixé la cible 2 et si on considère que le champ visuel est restreint aux trois cibles 1, 2, 3, il y a deux cibles potentielles pour le prochain mouvement, exceptée la cible 2 : la cible 1 qui vient d'être fixée et la cible 3 qui ne l'a pas encore été. Bien évidemment, puisque la cible 1 vient d'être fixée, c'est la cible 3 qui doit

être sélectionnée. Néanmoins, si on considère que les cibles qui ont été fixées sont mémorisées dans un référentiel centré-œil, la position de la cible 1 a changé pendant le mouvement oculaire. Mais puisque ce changement est causé par un mouvement propre, il est prévisible.

4.3.2 Mécanisme proposé

Le mécanisme proposé pour réaliser la tâche de parcours oculaire illustrée précédemment est présenté sur la figure 4.8. Il est constitué de quatre cartes fonctionnelles. Tout d'abord, l'information visuelle spatiale est représentée dans la carte de saillance rétinotopique. Elle se projette sur deux cartes : la carte FOCUS et la mémoire de travail. La carte FOCUS réalise la sélection d'une cible pour un mouvement oculaire grâce à des connexions latérales en différence de gaussiennes qui induisent une compétition entre les unités. La mémoire de travail, rétinotopique, est constituée de deux cartes connectées récursivement (représentées sur la figure par une seule carte avec une connexion réentrante), qui reçoivent trois afférences :

- la carte de saillance fournit la position spatiale de l'ensemble des cibles,
- la carte FOCUS fournit la position de la prochaine cible saccadique,
- la carte ANTICIPATION fournit la prédiction de la position que les cibles mémorisées vont occuper après le mouvement oculaire vers la cible représentée dans la carte FOCUS.

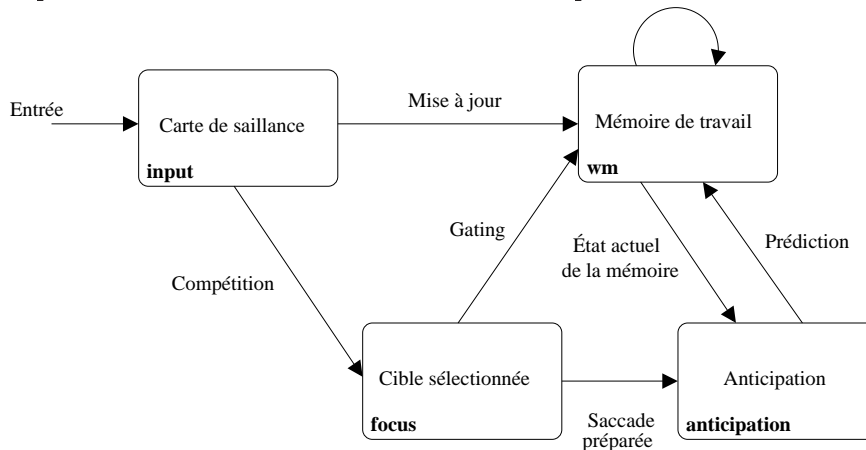


FIG. 4.8 – L'entrée visuelle représentée dans la carte de saillance excite à la fois la carte de sélection FOCUS et la mémoire de travail WM. La sélection d'une cible s'opère par des connexions latérales dans FOCUS qui induisent de la compétition entre les unités. L'excitation conjointe de la carte de saillance et de la carte FOCUS permet de faire émerger une cible dans la mémoire de travail, tandis que l'excitation de la carte de saillance suffit à entretenir la mémoire. La carte d'anticipation combine les afférences des cartes FOCUS et WM pour prédire l'état de la mémoire de travail après l'exécution de la saccade en direction de la cible sélectionnée. Les paramètres des neurones et des projections sont disponibles dans l'annexe A.4

Les paramètres des cellules de la mémoire de travail sont ajustés de telle sorte que l'entrée d'une cible en mémoire de travail dépend de la présence conjointe de cette cible dans la carte de saillance et dans la carte FOCUS (les paramètres sont disponibles dans l'annexe A.4). Ainsi, lorsqu'une cible est sélectionnée dans la carte FOCUS, elle entre en mémoire de travail puisqu'elle est également présente dans l'entrée visuelle. La mémorisation de cette cible permet de s'assurer qu'elle ne sera pas sélectionnée ultérieurement.

Pour qu'une cible mémorisée demeure en mémoire de travail, il suffit qu'elle soit excitée par la carte de saillance. Ainsi, dès lors qu'une cible est entrée en mémoire de travail, une autre cible peut être sélectionnée dans la carte FOCUS sans que la première ne soit oubliée. Enfin, la troisième afférence provenant de l'anticipation permet de mettre à jour la mémoire de travail. Au même titre que l'entrée en mémoire de travail d'une cible est conditionnée par sa présence conjointe dans la carte de saillance et la carte FOCUS, une cible peut également être mémorisée par l'activation conjointe de la carte ANTICIPATION et de la carte de saillance. Ainsi, lorsqu'un mouvement oculaire est exécuté, la carte d'anticipation prédit la position future des cibles mémorisées et cette prédiction est combinée avec la nouvelle entrée visuelle, une fois le mouvement exécuté, afin de mettre à jour la mémoire. Une fois la saccade exécutée, une nouvelle compétition est engagée entre les unités de la carte FOCUS. Il peut arriver qu'une cible déjà fixée soit sélectionnée étant donné qu'aucun biais n'est apporté en faveur des cibles non encore fixées. Néanmoins, lorsqu'une cible déjà fixée est sélectionnée, un mécanisme, non représenté sur la figure, permet de changer la sélection. Un exemple d'un tel mécanisme a déjà été proposé dans le chapitre 2 mais comme ce n'est pas le point essentiel discuté dans ce paragraphe, nous ne le détaillerons pas plus.

Nous n'avons pas encore détaillé le mécanisme qui réalise l'anticipation. Néanmoins il est basé sur l'idée présentée tout au long de ce chapitre qui consiste à utiliser un mécanisme sigma-pi, combinant l'état actuel de la mémoire de travail et la cible d'un mouvement oculaire fournie par la carte FOCUS. Tous les paramètres du mécanisme sont disponibles dans l'annexe A.4.

4.3.3 Simulations et résultats

Fonctionnement normal du mécanisme

Le fonctionnement du mécanisme est illustré sur la figure 4.9. Le mécanisme doit réaliser trois saccades successives vers les trois cibles visuelles présentées, sans jamais exécuter deux saccades vers la même cible. L'entrée visuelle, représentée dans la carte INPUT excite la carte FOCUS ainsi que la carte WM, de telle sorte que trois régions sont excitées dans ces deux cartes (figure 4.9a). La compétition latérale dans la carte FOCUS conduit à la sélection d'une des trois cibles. L'excitation conjointe de la carte FOCUS et de la carte INPUT conduit à l'entrée de la cible sélectionnée en mémoire de travail, qui est formée par le circuit récurrent WM-THAL_WM. Les unités de la carte ANTICIPATION combinent l'état actuel de la mémoire de travail (fourni par ses afférences de WM) avec la position de la cible saccadique (fournie par ses afférences de FOCUS) pour prédire la position occupée par cette cible après la saccade exécutée pour la fixer.

Pendant l'exécution d'une saccade, nous faisons l'hypothèse qu'il n'y a pas d'entrée visuelle (figure 4.9b). Même si cette hypothèse semble de plus en plus remise en cause d'un point de vue biologique, elle est pratique du point de vue de l'implémentation, en particulier parce qu'elle permet de s'assurer que les cibles dans la mémoire de travail sont "oubliées" avant que la nouvelle entrée visuelle n'arrive. Si les anciennes cibles mémorisées ne sont pas oubliées et qu'une stimulation visuelle occupe, après la saccade, la même position rétinotopique qu'une cible mémorisée, alors, ces cibles mémorisées demeureront en mémoire de travail. Ainsi, des cibles pourraient être

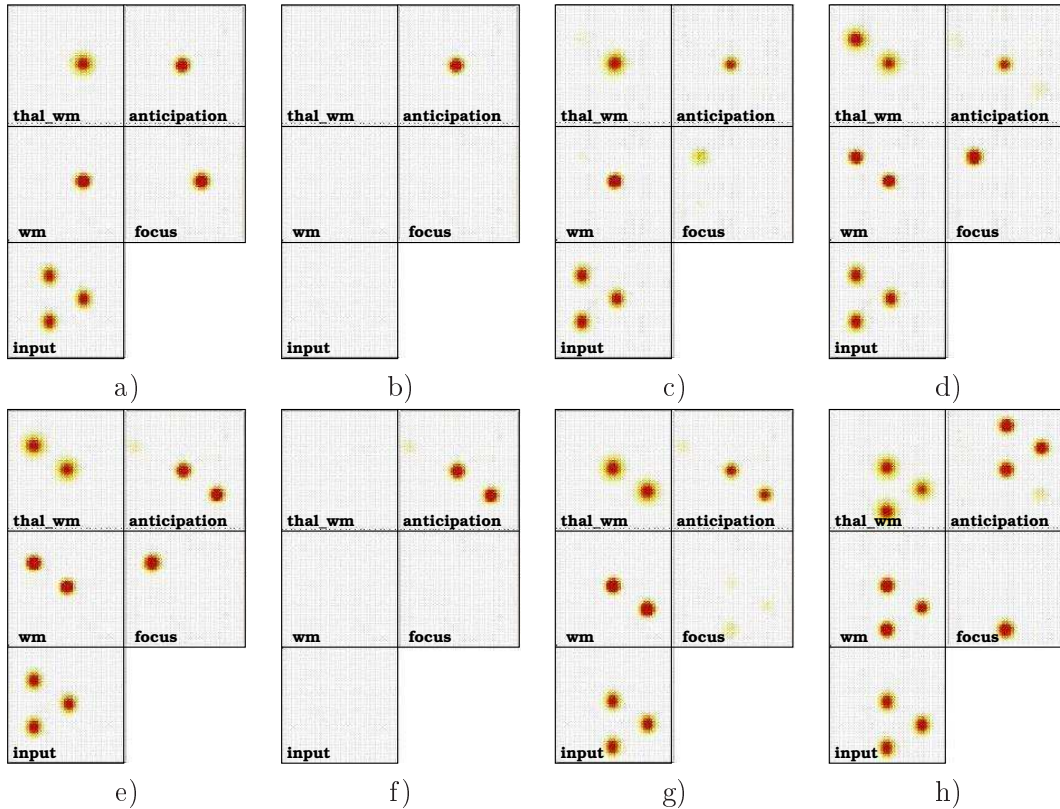


FIG. 4.9 – Une séquence particulière d'évolution du modèle durant le parcours saccadique. a) L'une des trois cibles gagne la compétition dans la carte FOCUS. La carte d'ANTICIPATION prédit l'état futur de la mémoire de travail spatiale, formée par le circuit WM-THAL_WM. b) Les paramètres des neurones sont ajustés de telle sorte que les neurones de la carte ANTICIPATION ont une latence plus longue, ce qui a pour conséquence que les activités de toutes les autres cartes décroissent rapidement mais que l'anticipation demeure plus longtemps pendant l'exécution d'une saccade. c) La mémoire de travail est mise à jour en combinant la nouvelle perception visuelle avec la prédiction et une nouvelle compétition s'engage dans la carte FOCUS. d) La cible gagnante émerge dans la mémoire de travail. e) L'état futur de la mémoire de travail est prédit. f) La saccade est exécutée et seule l'activité dans la carte ANTICIPATION demeure. g) La mémoire de travail est mise à jour en combinant la nouvelle perception visuelle et l'anticipation. h) La mémoire de travail étant mise à jour, la dernière cible est sélectionnée.

conservées en mémoire de travail non pas parce qu'elles ont été la cible d'un mouvement oculaire mais parce qu'elles occupent la position rétinotopique d'une ancienne cible mémorisée.

Une fois la saccade exécutée, la nouvelle perception visuelle est combinée avec l'anticipation pour mettre à jour la mémoire de travail. La figure 4.9c représente l'état du mécanisme peu de temps après la fin de la saccade. La mémoire de travail est mise à jour et une nouvelle cible est en train de gagner la compétition dans FOCUS. La nouvelle cible sélectionnée, combinée avec l'entrée visuelle, émerge dans la mémoire de travail (figure 4.9d). Etant donné l'état actuel de la mémoire de travail ainsi que la cible pour la prochaine saccade, la carte ANTICIPATION prédit la position des cibles mémorisées une fois que la saccade sera exécutée (figure 4.9e). Cette prédiction est combinée avec la nouvelle perception visuelle, une fois la saccade réalisée, afin de mettre à jour la mémoire de travail (figure 4.9g). Enfin, puisque l'entrée visuelle n'est constituée que de trois cibles et que deux cibles ont déjà été fixées, la dernière cible est correctement sélectionnée (figure 4.9h).

Le principal mécanisme que l'on souhaitait mettre en avant dans ce paragraphe est le mécanisme de mise à jour de la mémoire rétinotopique, en combinant la perception visuelle post-saccadique avec la prédiction de la position que les cibles mémorisées vont occuper après la saccade. Pour souligner l'importance du signal d'anticipation pour ce mécanisme, on propose d'étudier les conséquences de son inactivation.

Que se passe-t-il si la prédiction est altérée ?

On a présenté dans le paragraphe précédent un mécanisme qui permet de mettre à jour une mémoire de travail spatiale, rétinotopique, dont le principe réside dans la prédiction des conséquences de l'action motrice qui va être engagée. On propose désormais d'étudier les conséquences de son inactivation. Pour ce faire, on considère une tâche qui consiste à réaliser deux saccades successives vers deux cibles visuelles. Pendant que la tâche est réalisée, on enregistre les cellules de quatre cartes (carte de saillance, carte focus, mémoire de travail et anticipation), dont les champs récepteurs couvrent cinq positions du champ visuel notées x_0, x_1, x_2, x_3, x_4 sur la figure 4.10

Pour forcer la séquence motrice "saccade vers la cible la plus à gauche puis saccade vers la cible la plus à droite", on ajoute un biais en amplitude en faveur de la cible qui doit être sélectionnée. Ainsi, on s'assure que le mécanisme commencera par réaliser une saccade vers la cible la plus à gauche, suivie d'une saccade vers la cible la plus à droite. Etant donné ce scénario, on connaît a priori les couples de positions du champ visuel qui seront occupées par une cible :

1. les positions x_1 et x_2 avec un biais en faveur de la cible en x_1
2. les positions x_2 et x_4 , avec un biais en faveur de la cible en x_4
3. les positions x_0 et x_2

Sur la figure 4.11 sont respectivement illustrés :

- l'état courant de l'entrée visuelle avec une flèche indiquant la prochaine cible saccadique,

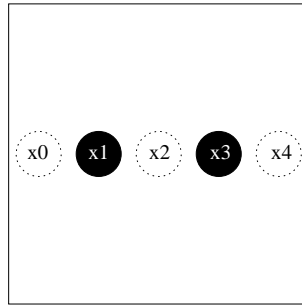


FIG. 4.10 – Pendant l'exécution de la tâche, on enregistre simultanément l'activité dans les quatre cartes INPUT, FOCUS, WM et ANTICIPATION aux cinq positions x_0, x_1, x_2, x_3, x_4 qui sont potentiellement occupées par un stimulus (visuel, moteur, mémorisé ou anticipé) durant la tâche.

- un axe temporel sur lequel figurent les moments d'exécution d'une saccade,
- l'évolution de l'activité des cellules des quatre cartes (carte de saillance, carte focus, mémoire de travail et anticipation) aux cinq positions du champ visuel

A la présentation des deux cibles, les deux positions x_1 et x_3 sont excitées dans l'entrée visuelle (courbe en tirets). Ces deux positions sont également excitées dans la carte FOCUS (courbe en trait plein). Seule la cible en x_1 gagne la compétition tandis que l'activité dans la carte FOCUS à la position x_3 s'évanouit rapidement. Puisque cette cible est sélectionnée, elle entre en mémoire de travail (courbe en tirets-pointillés). A partir de ce moment, l'anticipation se construit petit à petit et on observe que l'activité croît dans la carte ANTICIPATION (courbe en pointillés) à la position x_2 (c'est à dire au centre du champ visuel) qui sera occupée par la cible mémorisée, après la saccade. Une fois que la saccade pour fixer la cible en x_1 est exécutée, les deux cibles occupent respectivement les positions x_2 et x_4 (courbe en tiret). On observe la mise à jour de la mémoire de travail par le fait que son activité en x_2 croît, même en l'absence d'un signal de la carte FOCUS. Nous avons observé précédemment que la séquence d'activation précédant l'entrée en mémoire de travail était : INPUT-FOCUS-WM. Après la saccade, on constate que la carte FOCUS n'est pas excitée en x_2 mais c'est la coactivation de l'entrée visuelle et de l'anticipation qui permet l'entrée en mémoire de travail de la cible.

Un nouvelle compétition s'engage entre les deux cibles dans la carte FOCUS, mais puisqu'un biais est apporté en faveur de la cible présente en x_4 , elle gagne la compétition. On l'observe par l'augmentation de l'activité en x_4 dans la carte FOCUS, au détriment de l'activité en x_2 . On notera néanmoins une légère bouffée d'activité dans la carte FOCUS à la position x_2 , aux alentours de 45 pas de temps, qui s'estompe rapidement et qui correspond à la perte de la compétition de la part de la cible en x_2 .

La cible étant sélectionnée, la carte ANTICIPATION prédit que les deux cibles, pour le moment en x_2 et x_4 vont occuper les positions x_0 et x_2 . On l'observe facilement à la position x_0 où l'activité croît progressivement dans la carte ANTICIPATION. A la position x_2 , le comportement est légèrement différent qu'en x_0 . On l'observe par la petite variation d'activité aux alentours de 55 pas de temps. La diminution de l'activité, qui correspond à l'oubli de la prédiction correspondant à l'ancienne saccade, est suivie d'une augmentation d'activité qui correspond à la prédiction de

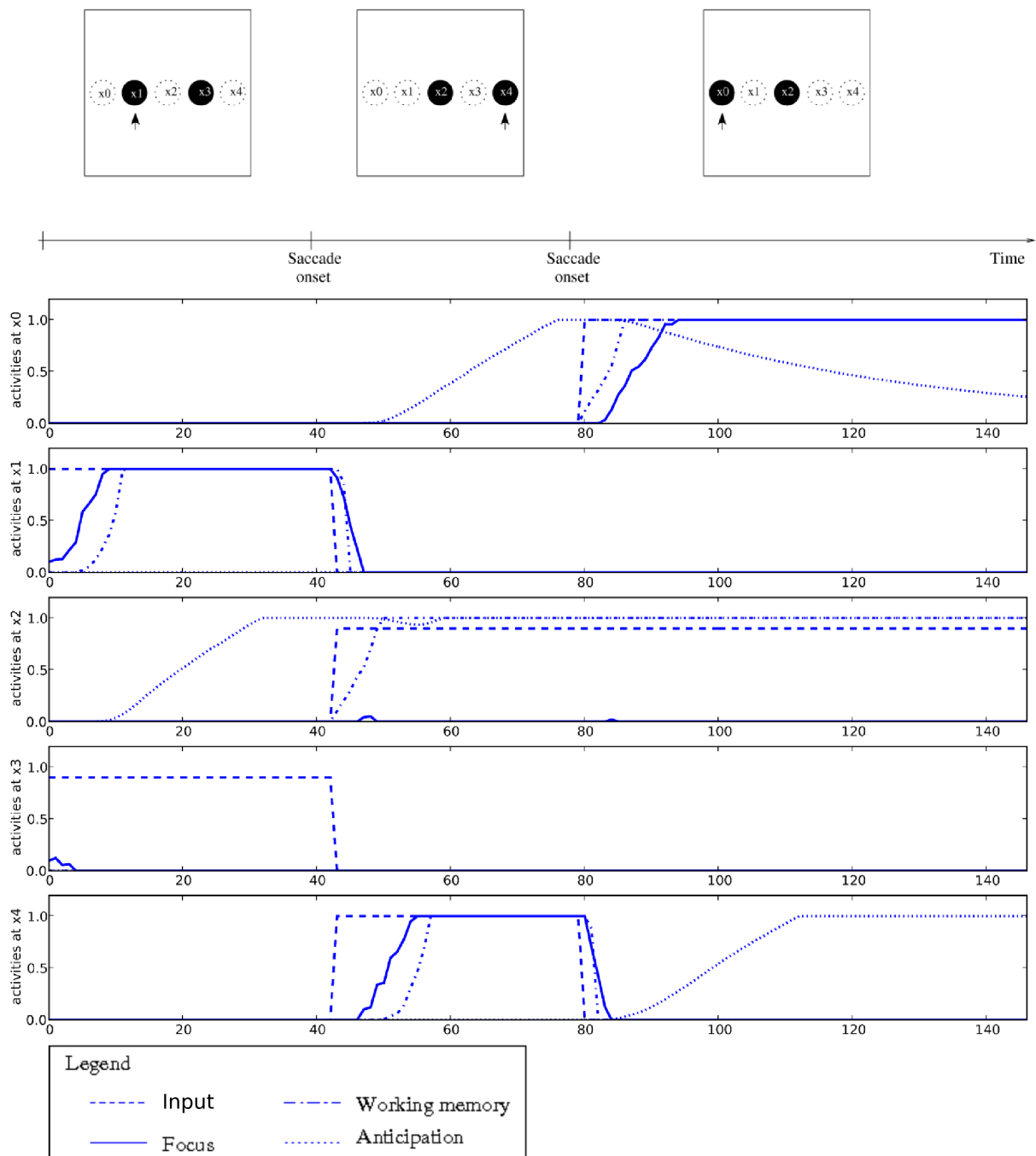


FIG. 4.11 – Haut : Pendant l'exécution de la tâche, à chaque instant, l'un des deux stimuli est physiquement plus saillant (indiqué par une flèche), pour forcer l'ordre de la séquence motrice. La première saccade, dirigée vers la cible en x_1 , amène les deux cibles en x_2 et x_4 . La seconde saccade, dirigée vers la cible en x_4 , amène les deux cibles en x_0 et x_2 . Bas : Pendant l'exécution de la tâche, on enregistre l'activité dans les quatre cartes du mécanisme aux cinq positions potentiellement occupées par une cible. Se référer au texte pour plus de détails.

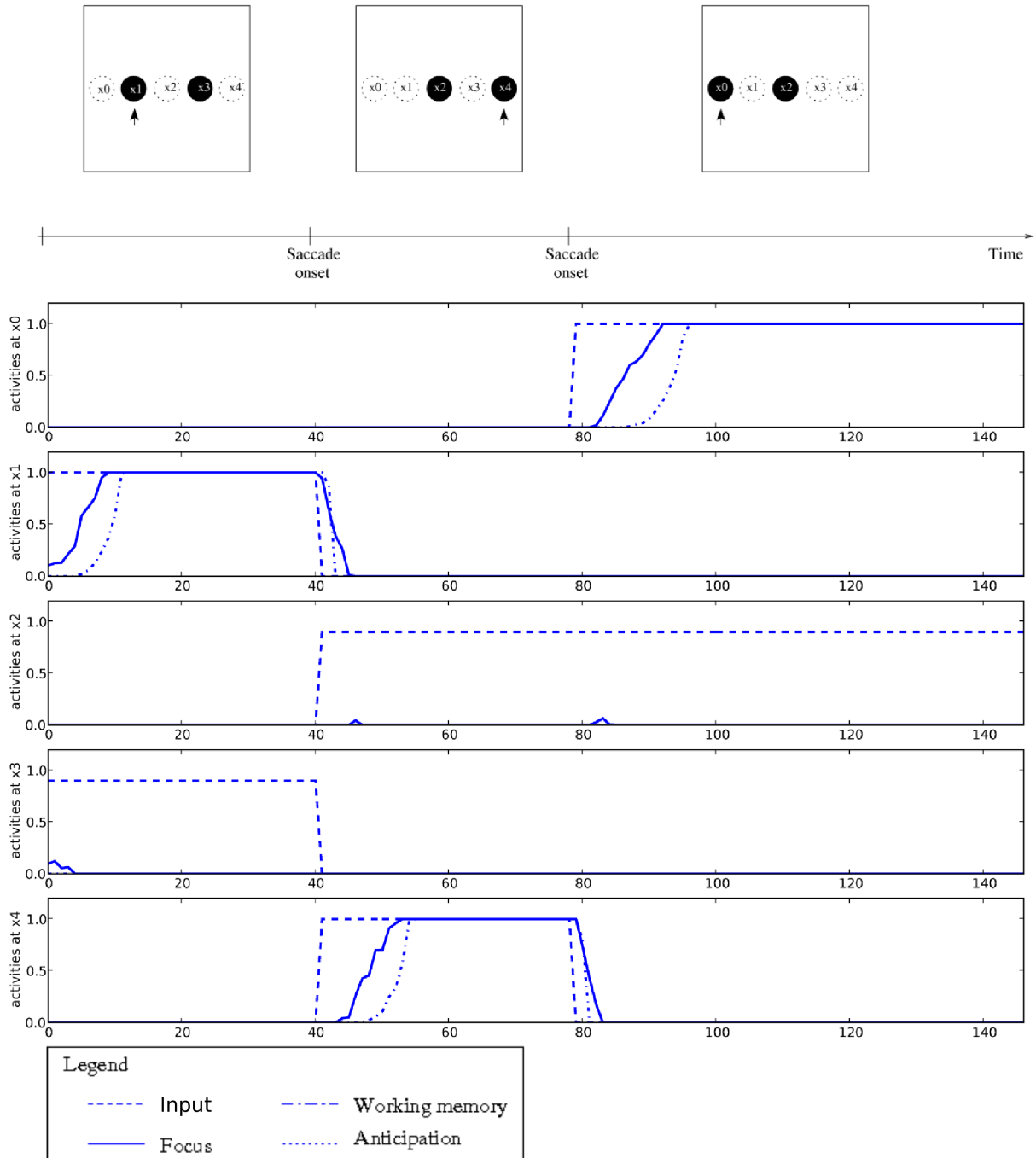


FIG. 4.12 – Le protocole expérimental est le même que pour la figure 4.11, à l'exception du signal d'anticipation qui est désactivé. Se référer au texte pour plus de détails.

la position de la cible actuellement en x_4 .

Lorsque la seconde saccade est exécutée, les deux cibles occupent les positions x_0 et x_2 . La combinaison de la nouvelle entrée visuelle et de l'anticipation conduit à la mise à jour de la mémoire.

La figure 4.12 correspond au même protocole expérimental que précédemment, dans la condition où le signal d'anticipation est inactivé. La séquence de saccades reste la même puisque la sélection de la prochaine cible est dépendante de la saillance des cibles, qui est ajustée de telle sorte que la prochaine cible saccadique soit toujours la plus saillante. Dans ce cadre expérimental, la distinction ne se fait pas au niveau du comportement obtenu mais plutôt au niveau de l'activité des cellules dans la mémoire de travail. En effet, en l'absence d'anticipation, la mémoire de travail n'est plus mise à jour. On l'observe en particulier par l'absence d'activité dans la mémoire de travail, après l'exécution de la première saccade (à 40 pas de temps), à la position x_2 . Puisque la mémoire de travail n'est pas mise à jour, le comportement du mécanisme est uniquement guidé par la saillance des stimuli visuels. Il serait en particulier incapable de réaliser une tâche de recherche visuelle dans laquelle plusieurs cibles doivent être explorées. Dans ce cas, on s'attend à ce que le mécanisme privé d'anticipation réalise plusieurs saccades vers les mêmes cibles, ce qui conduit à une exploration visuelle nettement moins efficace que pour le mécanisme qui met convenablement à jour sa mémoire de travail.

4.3.4 Conclusion

Dans cette section, nous avons présenté un mécanisme qui permet de mettre à jour une mémoire de travail spatiale rétinotopique, sans avoir à utiliser une représentation indépendante de la position de l'œil. Ces travaux ont fait l'objet d'une publication [Fix et al., 2007b]. Le principe de la mise à jour repose sur la prédiction de l'état futur de la mémoire de travail, calculée avant d'exécuter une saccade. Cette prédiction est ensuite combinée avec la nouvelle perception visuelle une fois la saccade exécutée. Malgré son architecture simpliste, le principe que l'on souhaitait mettre en avant réside dans l'utilisation d'une copie efférente d'une saccade oculaire sur le point d'être exécutée afin de prédire les conséquences de ce mouvement pour la mémoire de travail et utilise cette prédiction pour la mettre à jour.

Ce mécanisme se limite au traitement de l'information spatiale pour guider l'exécution des saccades oculaires. En particulier, en l'état, les caractéristiques visuelles des stimuli sont négligées. Dans la prochaine section, nous allons voir comment intégrer simplement les saccades oculaires dans le mécanisme de déploiement de l'attention covert que nous avons exposé au paragraphe 2.4.3.

4.4 Mécanisme de déploiement overt de l'attention visuelle spatiale

4.4.1 Introduction

Dans le paragraphe précédent, nous avons présenté un mécanisme permettant d'assurer la cohérence d'une mémoire de travail spatiale, centrée-œil, malgré des saccades oculaires. Néanmoins ce mécanisme se limite à la représentation et au traitement de l'information visuelle spatiale. Dans ce paragraphe, nous proposons de l'adjoindre au mécanisme présenté dans le paragraphe 2.4.3, qui traitait du déploiement de l'attention visuelle *spatial-based* et *feature-based*, sans mouvement oculaire. Pour illustrer le fonctionnement du modèle, on considère une tâche de recherche visuelle qui consiste à réaliser une séquence de saccades dirigées vers des cibles définies par leurs attributs visuels. Pour des raisons de simplicité, l'environnement est constitué de barres orientées et colorées ayant un ensemble limité d'attributs possibles :

- deux couleurs : bleu et vert,
- deux orientations : 45° et 135°.

Comme nous le verrons par la suite, ce mécanisme est ainsi capable de déployer l'attention sur les attributs visuels et l'attention spatiale *covert* et *overt*. La tâche a été choisie pour être relativement simple. En particulier, la règle qui dicte quelle saccade exécuter est simpliste. La raison de ce choix est que nous souhaitions avant tout nous focaliser sur un mécanisme permettant de mettre à jour une mémoire de travail spatiale, par anticipation, et montrer comment l'ensemble des mécanismes présentés jusqu'à maintenant peuvent s'intégrer.

4.4.2 Mécanisme proposé

Le mécanisme proposé est illustré sur la figure 4.13. La majeure partie du mécanisme a déjà été présentée dans le paragraphe 2.4.3. La seule différence réside dans l'ajout du mécanisme d'anticipation, présenté dans le paragraphe précédent. Les unités de la carte ANTICIPATION ont deux afférences : l'une de la carte FOCUS et l'autre de la mémoire de travail. Grâce à ces deux afférences, elles prédisent l'état futur de la mémoire de travail. Elles se projettent sur la mémoire de travail, où la prédiction est combinée avec la carte de saillance post-saccadique pour mettre à jour la mémoire.

Pour souligner le rôle de chacune des cartes ainsi que leurs interactions, on peut imaginer un scénario type de propagation des signaux dans le mécanisme. La description de ce scénario est succincte mais une description plus complète se trouve dans les paragraphes 2.4.3 et 4.3 dans lesquels chacune des parties du mécanisme est décrite plus en détails. L'entrée visuelle est filtrée selon quatre dimensions et représentée dans les cartes rétinotopiques 2D FEATURE MAPS. Ces cartes se projettent le long de deux voies. La première voie, qui regroupe deux ensembles de quatre unités sous le nom de FEATURE PROCESSING, extrait les caractéristiques visuelles les plus saillantes depuis les cartes FEATURE MAPS. Les quatre unités PERCEIVED FEATURES calculent chacune le maximum de l'activité des cartes FEATURE MAPS de la même caractéristique visuelle. Comme nous allons le décrire un peu plus tard, l'activité dans les cartes FEATURE MAPS est biaisée en faveur de la région sur laquelle l'attention spatiale est déployée, de telle sorte qu'il

est possible de décoder les attributs de la cible sur laquelle est portée l'attention dans les cartes PERCEIVED FEATURES. Ces caractéristiques sont combinées avec les attributs de la cible recherchée, codés dans quatre unités regroupées sous le nom TARGET, pour déterminer quelle action engager. On considère deux actions, chacune représentée par une cellule dans les cartes MOTOR COMMANDS. Si les caractéristiques visuelles extraites correspondent à la cible recherchée, la cellule MOVE ORDER décharge fortement, comparé à la cellule SWITCH qui décharge lorsque la cible sélectionnée spatialement ne correspond pas à la cible recherchée.

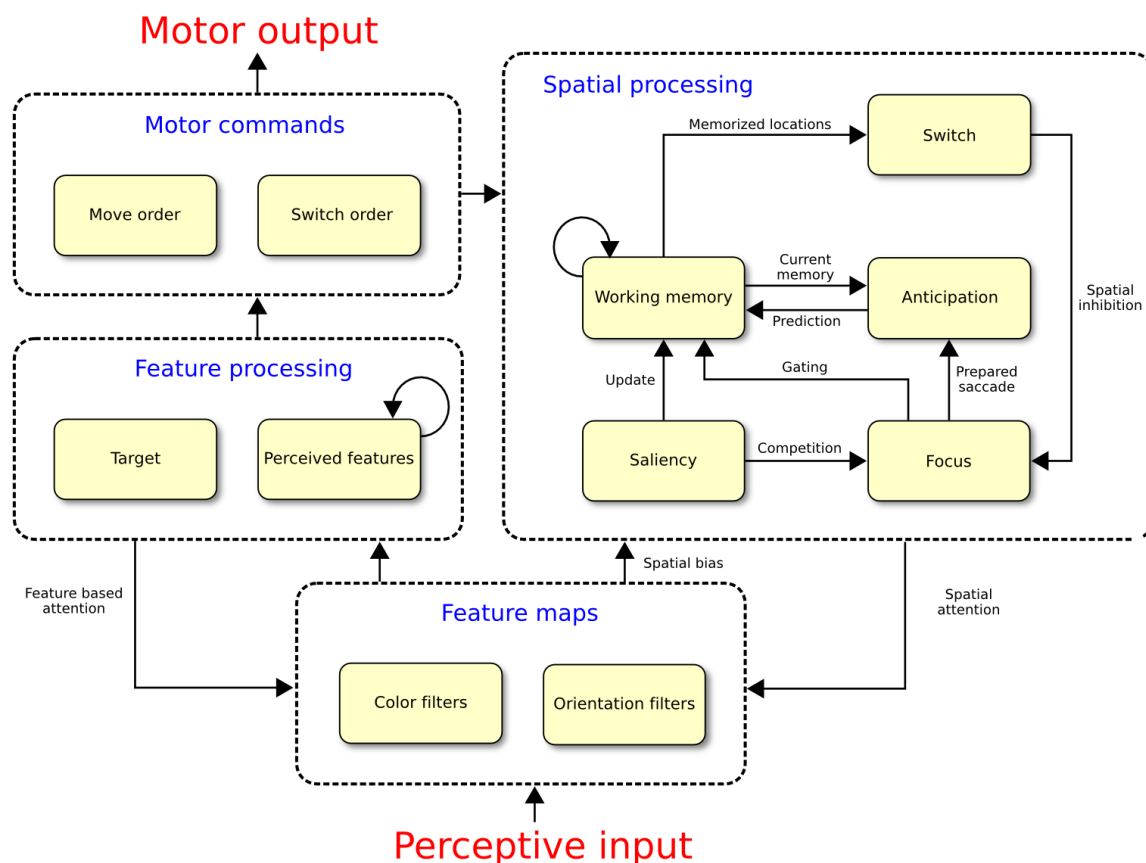


FIG. 4.13 – Le mécanisme est constitué du mécanisme de déploiement de l'attention présenté au paragraphe 2.4.3 auquel est ajouté le mécanisme d'anticipation. On trouvera plus de détails dans le texte. Les équations d'évolution des unités ainsi que les paramètres de la simulation sont disponibles dans les annexes A.1 et A.5.

La voie SPATIAL PROCESSING extrait uniquement la saillance des cartes FEATURE MAPS qui est représentée dans la carte de saillance. La carte de saillance excite la carte FOCUS dans laquelle s'engage une compétition entre les unités par l'intermédiaire de leurs connexions latérales en chapeau mexicain. L'activation conjointe de la carte de saillance et de la carte FOCUS permet l'entrée en mémoire de travail de la cible sélectionnée. La carte FOCUS projette également la cible sélectionnée vers la carte ANTICIPATION qui prédit l'état futur de la mémoire de travail à partir de son état actuel et de la cible d'une saccade. Une fois la saccade exécutée, la carte de saillance excitée par la nouvelle perception visuelle est combinée à la prédiction pour permettre la mise à jour de la mémoire de travail.

L'interaction entre les deux voies se fait par l'intermédiaire de la cellule SWITCH ORDER qui, en excitant la carte SWITCH, indique lorsque l'attention spatiale est déployée sur un distracteur.

4.4.3 Simulations et résultats

Afin d'illustrer le fonctionnement du mécanisme, nous proposons de réaliser une tâche de recherche visuelle qui nécessite d'exécuter un parcours saccadique vers les cibles définies par un attribut visuel (barres bleues) parmi un ensemble de cibles ayant des caractéristiques visuelles dans un ensemble limité d'attributs (barres bleues ou vertes, orientées à 45° ou 135°).

Nous faisons également l'hypothèse que le champ visuel est limité et l'environnement est défini de telle sorte qu'à chaque fixation, au moins trois cibles potentielles sont visibles :

- la cible fixée par la saccade précédente,
- la cible fixée,
- une cible non encore fixée.

A cela s'ajoutent des distracteurs qui ne doivent pas être la cible d'une saccade oculaire. Dans le protocole expérimental choisi, toutes les cibles vertes sont des distracteurs. La figure 4.14a représente le parcours oculaire réalisé dans ce contexte. Les cercles rouges en trait plein représentent les saccades oculaires. Le parcours oculaire est ainsi constitué de l'ensemble des cercles rouges en trait plein reliés par des flèches orientées pour indiquer le sens du parcours. Le point de départ du parcours est représenté par la flèche sans cercle à l'origine.

Les cercles en pointillés représentent le déploiement couvert de l'attention. On observe ainsi qu'à trois reprises, un distracteur a été sélectionné par l'attention spatiale, sans qu'il ne soit la cible d'un mouvement oculaire. La raison de cette sélection par l'attention est principalement la saillance de ces distracteurs. En effet, si la saillance des stimuli cibles (comprenant la contribution *bottom-up* de la cible dans son environnement et la contribution apportée par la connaissance a priori des attributs de la cible recherchée) est inférieure à la saillance des distracteurs, ce seront les distracteurs qui seront sélectionnés puisque la compétition engagée dans la carte FOCUS favorise les positions spatiales les plus saillantes. Néanmoins, une fois un distracteur sélectionné par l'attention, les attributs visuels extraits permettent de reconnaître le stimulus comme un distracteur et de désengager l'attention.

Les figures 4.14b-k illustrent ce que perçoit le mécanisme après chaque saccade, à l'exception de la figure 4.14b qui représente la perception visuelle initiale. Ces illustrations permettent de mieux se rendre compte des limites du champ visuel ainsi que de la distorsion causée par l'angle d'observation de la caméra. En particulier, sur la figure 4.14b, la caméra est dirigée face à la scène visuelle, à comparer par exemple aux figures 4.14i-k pour lesquelles l'excentricité est plus importante. Enfin, les instants d'exécution des différentes saccades sont représentés sur l'axe temporel en bas de la figure. La variabilité des temps de fixation (définie par le temps qui sépare deux

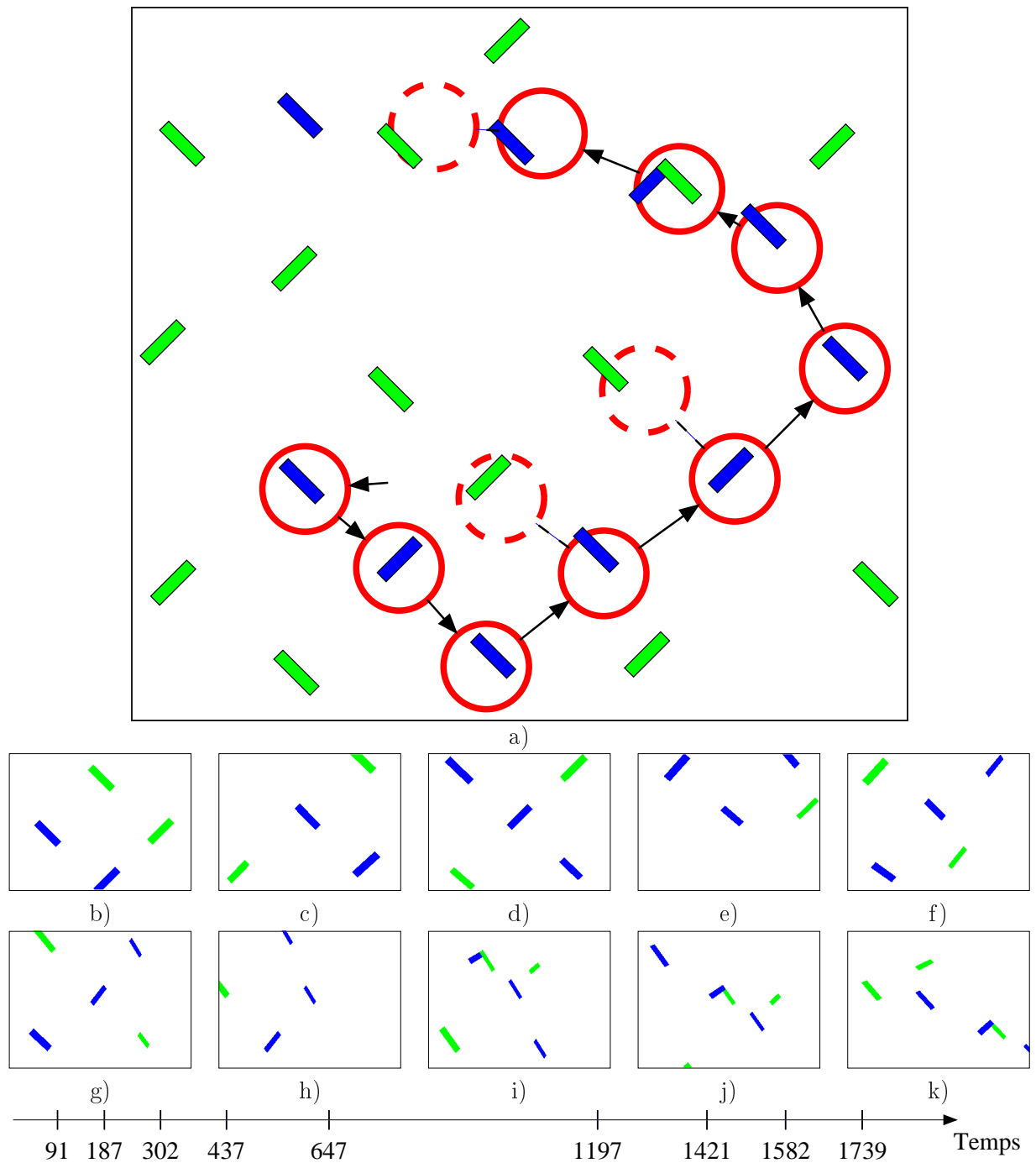


FIG. 4.14 – a) Parcours de l'attention spatiale covert et ouvert pendant l'exécution de la tâche de recherche visuelle des cibles bleues. Les cercles rouges en trait plein représentent les points de fixation successifs (le sens du parcours est indiqué par des flèches). Les cercles rouges en pointillés représentent le déploiement covert de l'attention visuelle spatiale. b-k) Ces figures représentent ce que le mécanisme perçoit après chaque saccade, à l'exception de la figure b qui représente la perception visuelle à la première fixation. Les moments d'exécution de chaque saccade sont indiqués sur l'axe temporel en bas de la figure.

saccades) peut s'expliquer par au moins deux facteurs. D'une part, lorsque l'attention *covert* est déployée sur un distracteur, et puisque la sélection par l'attention est séquentielle, l'attention doit d'abord se désengager du distracteur avant de se déployer sur une nouvelle cible. D'autre part, avec l'excentricité, les cibles deviennent moins saillantes. Puisqu'elles sont moins saillantes, les cellules de la carte de saillance dont le champ récepteur couvre leurs positions excitent moins la carte FOCUS, ce qui a pour conséquence d'allonger le temps mis pour sélectionner une cible dans cette carte.

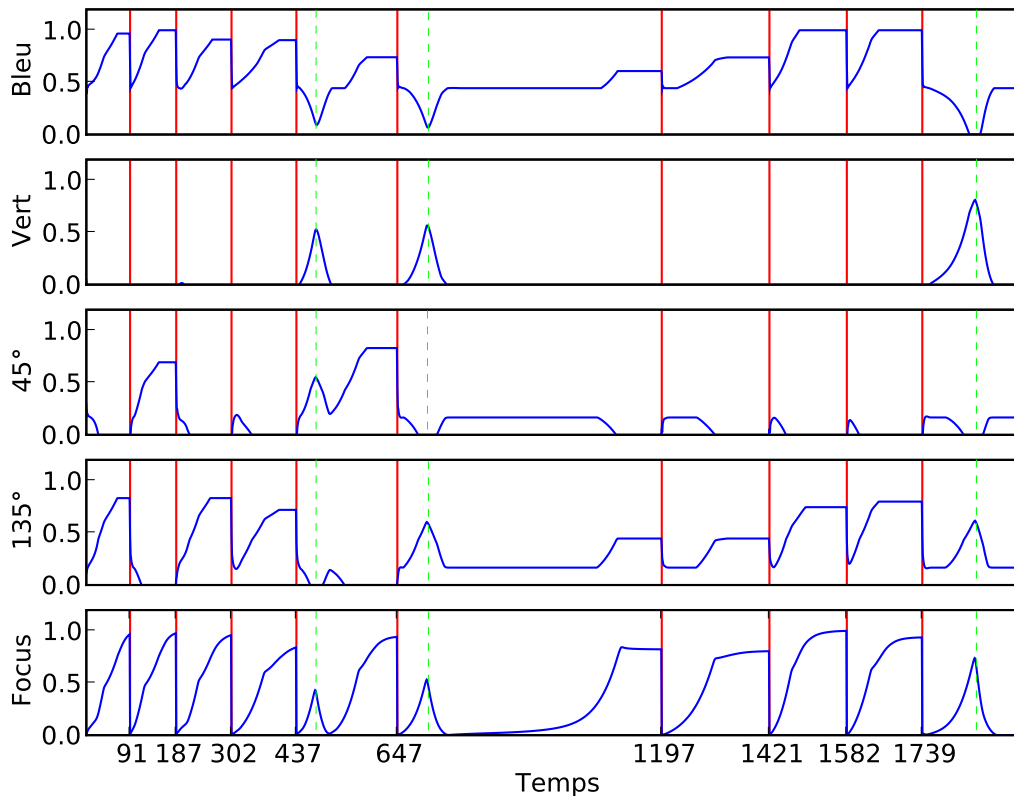


FIG. 4.15 – Les quatre premiers graphiques représentent l'activité des quatre cellules des cartes FEATURE MAPS respectivement sélectives aux attributs *bleu*, *vert*, 45° et 135° , au cours du temps. Les barres verticales rouges représentent les instants d'exécution des saccades et les barres vertes en pointillés représentent les instants de déploiement *covert* de l'attention. Ces quatre graphiques accompagnés du dernier graphique représentant l'évolution temporelle de l'activité moyenne normalisée dans la carte FOCUS fournissent de bons indicateurs pour estimer le comportement du mécanisme, à la fois sur le fait d'avoir sélectionné une cible et sur la nature de la cible sélectionnée.

Une autre façon d'illustrer le fonctionnement du mécanisme consiste à enregistrer l'activité de cellules pendant la réalisation de la tâche. En particulier, puisque la "décision" du comportement à adopter est prise par le modèle par l'intermédiaire des caractéristiques visuelles qu'il extrait, les activités dans la voie ventrale fournissent un bon indicateur du comportement choisi. L'activité dans la carte FOCUS est quant à elle un bon indicateur du déploiement de l'attention. La figure

4.15 représente l'activité des quatre cellules de la carte PERCEIVED FEATURES qui représentent respectivement l'attribut visuel bleu, vert, 45° et 135°, ainsi que l'activité normalisée dans la carte FOCUS. Les barres verticales rouges en trait plein indiquent les moments d'exécution d'une saccade tandis que les barres verticales vertes en pointillés indiquent un redéploiement couvert de l'attention spatiale.

On observe plus facilement sur ces enregistrements cellulaires la sélection par l'attention des trois distracteurs aux trois instants 477, 709 et 1849 pas de temps. Il est également plus facile de comprendre la variabilité des durées de fixation. En particulier, on constate qu'entre les pas de temps 647 et 1197, l'attention commence par se déployer sur un distracteur puis met un temps relativement long à se déployer sur une cible. Comme nous l'avons expliqué précédemment, le temps relativement long mis par l'attention pour se déployer sur cette région est une conséquence de la faible saillance de la cible sur laquelle l'attention se déploie.

On observe également un phénomène très intéressant entre les pas de temps 1421 et 1582. D'après les signaux décodés dans les populations PERCEIVED FEATURES, la cible sélectionnée par l'attention est une barre bleue orientée à 135°. Or, pendant cette période, le regard est au septième point de fixation (qui correspond à l'illustration 4.14i) et l'attention se déploie sur la région que deux cibles se partagent : une cible bleue orientée à 45° et une cible verte à 135°. C'est à dire que la région sur laquelle l'attention est déployée ne contient pas de cible bleue orientée à 135°. Les caractéristiques extraites par la population PERCEIVED FEATURES est un mélange des attributs des deux cibles qui occupent la région sur laquelle l'attention est déployée. Cette erreur s'apparente au *binding problem*, dans le sens où le système mélange les attributs visuels de deux cibles. La raison de cette erreur vient du fait que les deux cibles occupent le même champ récepteur de cellules des cartes FEATURE MAPS. Le biais apporté par la connaissance de la couleur de la cible recherchée supprime l'ambiguïté quant à la couleur de la cible de la région sur laquelle l'attention est déployée mais ne supprime pas l'ambiguïté quant à son orientation. Dans ce cas, l'orientation est fortement dépendante de l'information *bottom-up* fournie par l'environnement. Pour résoudre ce problème, on pourrait imaginer ajouter des niveaux intermédiaires entre l'entrée visuelle et les cartes FEATURE MAPS, avec des cellules dont la taille des champs récepteurs croît. En utilisant des connexions montantes et descendantes, le biais apporté aux FEATURE MAPS se propagerait sur les niveaux inférieurs pour lever l'ambiguïté sur l'orientation de la cible.

4.4.4 Conclusion

En conclusion, nous proposons de revenir sur quelques aspects négligés jusqu'à maintenant concernant le mécanisme présenté dans cette section et qui permettent d'assurer son bon fonctionnement.

Pendant le déroulement de la tâche, l'activité des cartes FEATURE MAPS, SALIENCY, FOCUS et WORKING MEMORY sont réinitialisées après chaque saccade, ce qui explique en particulier pourquoi les activités représentées sur la figure 4.15 chutent brutalement à l'exécution d'une saccade. C'est un artifice utilisé dans le mécanisme pour supprimer les interférences entre les perceptions pré et post-saccadiques. Pour illustrer sa nécessité, imaginons qu'une cible, à la posi-

tion A, soit mémorisée en mémoire de travail et qu'une saccade soit exécutée. Après la saccade, la cible mémorisée occupera une autre position, disons B, et, en combinant l'information de la carte de saillance en B avec la prédiction de la carte ANTICIPATION, la cible émergera effectivement en B dans la mémoire de travail. Mais qu'en est-il de la position en A? Si la carte de saillance ne fournit pas d'excitation en A, l'activité dans la mémoire de travail à cette position s'éteindra progressivement. Néanmoins, si une cible se trouve à cette position après la saccade, l'excitation dans la mémoire de travail demeurera à la position A, ce qui conduit à un dysfonctionnement de la mémoire de travail. La réinitialisation de l'activité dans la mémoire de travail permet de s'assurer que les anciennes cibles mémorisées ne seront plus en mémoire de travail lorsqu'elle sera mise à jour en combinant l'anticipation et la nouvelle carte de saillance. On peut l'assimiler à un "oubli" des anciennes positions mémorisées, qui ne sont plus cohérentes avec la nouvelle perception visuelle.

Il y a également un autre phénomène inhérent à l'anticipation et à la façon dont elle est utilisée dans le mécanisme. Le signal d'anticipation est nécessairement généré avant d'exécuter la saccade. Puisque l'anticipation est générée avant la saccade, elle excite la mémoire de travail avant d'exécuter le mouvement. Si les positions anticipées correspondent à des positions excitées dans la carte de saillance, ces positions émergent en mémoire de travail, justement parce que c'est la manière dont le système fonctionne. Par ailleurs, puisque la carte FOCUS représente à la fois le déploiement covert et overt de l'attention, l'anticipation peut faire émerger des cibles en mémoire de travail même lorsque l'attention est déployée sans mouvement oculaire. Pour limiter la portée de ce problème, on pourrait imaginer inhiber toniquement la carte ANTICIPATION tant qu'un mouvement oculaire n'est pas sur le point d'être exécuté. Cette solution nous paraît raisonnable puisque l'anticipation, dans le mécanisme considéré, n'a de rôle que lorsqu'une saccade est exécutée.

Une des limites du mécanisme réside dans la portée de la mémoire de travail utilisée. En effet, la mémoire de travail spatiale, en coordonnées centrées-œil, est limitée au champ visuel. Dès qu'une cible quitte le champ visuel, elle est oubliée puisqu'elle n'est plus mise à jour dans la mémoire. Cette limite est due au fait que la mémoire de travail est une représentation explicite des positions spatiales des cibles. Puisque la mémoire de travail est ici utilisée pour biaiser la sélection d'une cible pour un mouvement oculaire, on pourrait imaginer que les cibles qu'elle mémorise sont l'ensemble des cibles à mémoriser qui peuvent être atteintes par un mouvement de l'œil. Pour dépasser la limite de la portée spatiale de la mémoire, on pourrait avoir recours à des représentations spatiales de plus haut niveau, comme des représentations égocentrées centrées-tête, centrées-corps, ou une représentation allocentrée. Par exemple, on pourrait utiliser une mémoire centrée-tête qui représente l'ensemble des cibles à mémoriser qui peuvent être amenées dans le champ visuel par un mouvement de la tête. Puisque l'espace couvert par un mouvement de la tête est plus grand que l'espace couvert par un mouvement de l'œil, la mémoire centrée-tête pourrait aider à mémoriser des cibles qui ne sont plus directement accessibles par un mouvement oculaire.

Enfin, la décision d'exécuter une saccade est fortement dépendante de la tâche que nous avons réalisée avec le mécanisme. Ainsi, les cellules qui "décident" quel comportement adopter sont quelque peu ad-hoc. Puisque l'objectif du mécanisme était de s'intéresser à un mécanisme

qui permet de considérer les mouvements oculaires, nous n'avons pas détaillé ce module mais c'est bien évidemment un point qu'il serait intéressant d'étudier.

Malgré les limitations que nous venons de souligner, le mécanisme proposé, qui a fait l'objet d'une publication [Fix et al., 2007a], permet d'explorer un moyen par lequel une mémoire de travail spatiale peut être mise à jour à l'aide d'une copie éfrente d'un mouvement oculaire sur le point d'être exécuté. Il s'intègre au mécanisme de déploiement de l'attention que nous avons présenté au paragraphe 2.4.3 de telle sorte qu'un seul et même mécanisme permet à la fois de déployer l'attention sur les attributs visuels et l'attention spatiale covert et overt en anticipant les conséquences d'un mouvement oculaire imminent.

Il est également important de souligner que le mécanisme proposé doit être considéré comme une modélisation fonctionnelle dans le sens où il nous semble que cette architecture pourrait constituer une architecture minimale pour réaliser les comportements considérés, mais que la correspondance entre les cartes du modèle et les aires corticales n'est pas directe. Pour expliciter ce dernier point, nous proposons de terminer la discussion de cette section par des spéculations quant au réseau d'aires corticales qui seraient impliquées dans les mécanismes :

- de sélection de la prochaine saccadique
- de mémorisation de cibles spatiales
- de mise à jour de cette mémoire

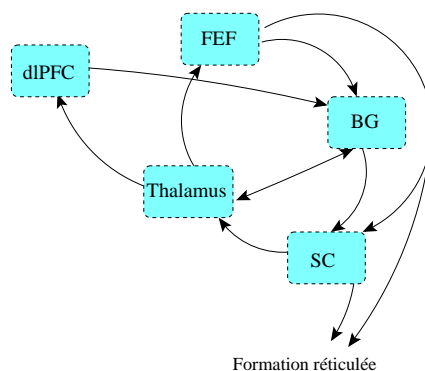


FIG. 4.16 – Quelques unes des aires corticales et sous-corticales impliquées dans le contrôle oculomoteur. Les champs oculomoteurs frontaux (FEF) et le colliculus supérieur (SC) ont des activités liées à la représentation spatiale de la prochaine cible saccadique. Le cortex préfrontal dorsolatéral (dIPFC) a des activités liées à la mémorisation à court terme d'une information spatiale. Ces trois aires sont impliquées dans des boucles avec les ganglions de la base (BG), qui a un rapport privilégié avec SC puisque l'inhibition tonique qu'ils exercent sur ce dernier doit être levée pour permettre l'exécution d'une saccade. Les circuits impliquant FEF, BG et SC pourraient être prépondérants dans la sélection de la prochaine cible saccadique, tandis que la boucle impliquant dIPFC et BG pourrait permettre de mémoriser une information spatiale à court terme, éventuellement mise à jour grâce à une copie éfrente transmise par SC et passant par le thalamus.

Nous mettons l'accent en particulier sur cinq structures corticales ou sous-corticales, représentées sur la figure 4.16, qui sont : les champs oculomoteurs frontaux (FEF), le cortex préfrontal

dorsolatéral (dlPFC), les ganglions de la base (BG), le thalamus et le colliculus supérieur (SC). Comme nous l'avons vu dans le chapitre précédent, FEF et SC ont des activités liées à la représentation de la cible d'une saccade et se projettent directement sur la formation réticulée qui est la dernière structure contrôlant les motoneurons des muscles extraoculaires. Ces aires peuvent soit être efférentes au circuit de sélection de la prochaine cible saccadique, soit faire partie de ce circuit. Les ganglions de la base ont un rôle privilégié dans l'exécution de la prochaine saccade puisque la *substantia nigra pars reticulata* (SNr), un noyau des BG, inhibe toniquement SC ; cette inhibition doit être levée pour permettre l'exécution d'une saccade. Comme nous l'avons mentionné au chapitre précédent, les ganglions de la base forment des boucles [Alexander et al., 1986]. On va distinguer deux types de boucles : les boucles cortico-basales dlPFC-BG-thalamus et FEF-BG-thalamus et sous-cortico-basales SC-thalamus-BG. La boucle cortico-basale impliquant FEF et la boucle sous-cortico-basale impliquant SC pourraient jouer un rôle prépondérant dans la sélection de la prochaine cible saccadique. La boucle dlPFC-BG-thalamus pourrait être, quant à elle, très impliquée dans la mémorisation spatiale à court terme. En effet, nous avons vu dans le chapitre précédent que dlPFC a des activités liées à la mémorisation d'une cible spatiale. Par ailleurs, d'après [Alexander et al., 1986], la boucle dlPFC-BG-thalamus implique le noyau médiodorsal du thalamus, qui, nous l'avons également vu au chapitre précédent, est une structure entre SC et FEF qui transporterait une décharge corollaire d'un mouvement saccadique. Il est ainsi possible que la boucle dlPFC-BG-thalamus soit impliquée dans la mémorisation à court terme d'une information spatiale, et qu'elle aurait accès, via la décharge corollaire provenant de SC, aux paramètres d'une saccade sur le point d'être exécutée pour mettre à jour cette mémoire.

Conclusion

L'objet de ce chapitre était l'étude de différents mécanismes qui permettent de prendre en compte la modification brutale de l'information visuelle disponible, qui résulte des saccades oculaires. Pour ce faire nous avons étudié deux principes fonctionnels distincts. Le premier, que nous avons exposé dans la section 4.2, repose sur l'utilisation d'une mémoire spatiale centrée tête. Néanmoins, puisque les structures qui déterminent *in fine* la cible d'une saccade oculaire la représente dans un référentiel centré-œil, il est nécessaire de transformer les informations spatiales mémorisées dans un référentiel centré-tête, dans le référentiel de sélection centré-œil. Cette transformation est ici réalisée grâce à un mécanisme d'intégration non-linéaire des entrées. Nous avons présenté ce mécanisme d'intégration sigma-pi dans la section 4.1. Il constitue une approche économe au calcul de transformations sensorimotrices, en ce sens que, comparé aux champs de gain, il ne nécessite pas d'utiliser une représentation intermédiaire dont les dimensions peuvent être importantes. Néanmoins, le revers de la médaille est que ce mécanisme est moins générique que les champs de gain puisque la transformation sensorimotrice est directement représenté par la connectivité du réseau.

Le deuxième principe fonctionnel, que nous avons exposé dans la section 4.3, repose sur l'utilisation de représentations centrées-œil uniquement, sans faire référence, à aucun moment, à des représentations plus évoluées (par plus évolué, nous entendons des représentations spatiales indépendantes de la position des yeux). Si la mémoire spatiale est définie par rapport à la position courante des yeux, elle doit nécessairement être mise à jour avec chaque saccade exécutée. La mise à jour de cette mémoire repose sur l'idée que, puisque les saccades sont des mouvements

volontaires, il est possible d'anticiper certaines de leurs conséquences. En particulier, la commande motrice peut être utilisée pour prédire quelles positions spatiales vont occuper les cibles mémorisées après la saccade.

Ce second mécanisme a enfin été utilisé, dans la section 4.4, pour étendre le mécanisme de déploiement de l'attention visuelle que nous présenté au chapitre 2. Le mécanisme ainsi obtenu permet d'explorer une scène visuelle, avec ou sans saccade oculaire. Il est important de remarquer que ce comportement d'exploration visuelle n'est que le résultat de l'interaction de cellules élémentaires, qui effectuent des calculs simples.

Conclusion et perspectives

L'étude que nous avons menée ces trois dernières années cherchait à mieux comprendre les phénomènes émergents de l'interaction de cellules élémentaires, qui réalisent un calcul simple. C'est à la fois un paradigme de calcul intéressant pour l'informatique, mais également un certain niveau d'abstraction du fonctionnement du système nerveux. Pour l'informatique, on part d'un constat relativement simple : certains des comportements observés dans le vivant sont souhaitables pour des systèmes informatiques mais pas encore réalisés de manière satisfaisante. On peut, à titre d'exemple, mentionner la navigation pour la robotique autonome, la planification, la résolution de problèmes, le langage ou bien encore la reconnaissance d'objets visuels. Pour les neurosciences, les modèles construits dans le cadre de ce paradigme ne peuvent évidemment pas expliquer toutes les données expérimentales recueillies sur le cerveau puisqu'ils reposent sur un ensemble nécessairement sélectif d'hypothèses sur son fonctionnement. L'hypothèse faite dans cette étude est de se limiter à l'interaction de cellules simples et de comprendre à quel point une capacité cognitive complexe peut émerger de ce substrat. Pour illustrer cela, nous nous sommes penchés sur l'attention visuelle et le contrôle oculomoteur, qui impliquent un vaste réseau d'aires corticales et sous-corticales.

Dans le chapitre 1, nous avons présenté les observations obtenues au travers d'expériences de psychologie expérimentale ou d'électrophysiologie qui révèlent quelques propriétés du phénomène d'attention visuelle *covert*. Nous en avons retenu quelques principes fonctionnels, le plus important étant la sélectivité de l'attention qui favorise le traitement d'une partie de l'information visuelle. Nous avons vu que l'attention peut être guidée par des propriétés visuelles propres à l'environnement (apparition soudaine d'un stimulus, stimulus très contrasté) ou par des buts internes (les buts d'une recherche visuelle par exemple). Le chapitre s'est conclu sur la présentation de trois modèles d'attention visuelle. Les modèles à base de carte de saillance apportent une solution intéressante à la manière dont l'attention peut être guidée par des propriétés intrinsèques de la scène visuelle. Le *Selective Tuning Model* met, quant à lui, l'accent sur la compétition pour la représentation. Enfin le modèle de F. Hamker repose sur des hypothèses biologiques fortes comme la ségrégation entre une voie spécialisée dans le traitement de l'information spatiale et une voie spécialisée dans l'extraction des attributs visuels. Il apporte également une contribution intéressante sous la forme de la *reentry hypothesis* selon laquelle le signal de préparation d'une saccade en direction d'une position spatiale peut servir à moduler des représentations visuelles de bas niveau. C'est une manière d'introduire un biais *top down*, un but, pour guider le déploiement de l'attention.

Dans le chapitre 2, nous avons commencé par exposer le cadre de modélisation au travers de quatre principes fondamentaux que sont le calcul local, numérique, distribué et adaptatif.

La présentation que nous avons faite de ces quatre principes nous a permis d'expliciter ce que nous entendons par l'émergence d'un comportement complexe d'un substrat de calculs simples. Nous avons poursuivi l'exposé en présentant les champs neuronaux qui s'avèrent être un modèle relativement simple de populations de neurones et qui contient trois de ces propriétés que sont le calcul local, numérique et distribué. Notre choix s'est porté sur ce modèle puisqu'il nous paraît être le plus simple pour observer des propriétés dynamiques émergentes intéressantes. Sa simplicité de mise en œuvre peut, par exemple, être comparée à la difficulté de mettre en œuvre des réseaux de neurones à spikes qui requièrent d'utiliser des méthodes d'évaluation événementielles [Rochel, 2004]. La dernière partie du chapitre 2 a été consacrée à la présentation d'un modèle d'attention visuelle dans le cadre des champs neuronaux. Pour des raisons didactiques, nous avons isolé ce qui nous paraît être des fonctions élémentaires comme la sélection d'une cible ou la mémorisation d'une information spatiale en montrant comment ces fonctions pouvaient émerger de calculs locaux, numériques et distribués. Enfin, nous avons montré comment ce système pouvait s'adjoindre au système d'attention visuelle développé par F. Hamker pour ajouter le guidage de l'attention visuelle par des attributs non spatiaux.

Le système ainsi obtenu peut être utilisé, par exemple, pour réaliser l'exploration d'une scène visuelle. Même si l'illustration que nous avons considérée se limite à rechercher une barre orientée colorée dans un environnement simpliste, les travaux sur l'attention visuelle *bottom up* nous laissent imaginer que ce système pourrait s'appliquer sur des images naturelles. Néanmoins, ce système reste limité à l'analyse de scènes visuelles statiques. Cette limite n'est pas désirable pour un système de vision par ordinateur et n'est pas réaliste d'un point de vue biologique si on se souvient que les primates exécutent de l'ordre de trois saccades oculaires par seconde. Par ailleurs, la théorie prémotrice de l'attention propose que ce sont les mêmes aires corticales qui sont impliquées dans le déploiement de l'attention visuelle spatiale et dans le contrôle oculomoteur, ce qui nous a poussé à étudier comment il était possible d'étendre le modèle d'attention visuelle, présenté au chapitre 2, pour prendre en compte une des conséquences des saccades oculaires qu'est le brutal changement de l'information visuelle disponible, qui semble aller à l'encontre de l'apparente stabilité de l'expérience visuelle.

Pour mieux apprécier ce que sont les saccades oculaires ainsi que les aires corticales et sous-corticales impliquées dans leur contrôle, nous avons proposé dans le chapitre 3 une revue des connaissances actuelles dans ce domaine. Nous avons essentiellement retenu de cette présentation les référentiels dans lesquels l'information spatiale est représentée et nous avons identifié des fonctions cognitives (sélection d'une cible saccadique, mémorisation d'une information spatiale, mémorisation d'une séquence de saccades) corrélées à l'activité observée dans différentes aires corticales et structures sous-corticales (essentiellement des aires du pariétal, du lobe frontal ainsi que les ganglions de la base).

Dans le chapitre 4, nous avons proposé deux mécanismes pour mieux comprendre l'apparente stabilité de l'expérience visuelle ainsi qu'un mécanisme qui modélise à la fois le déploiement *covert* et *overt* de l'attention visuelle. Les deux premiers mécanismes reposent sur deux principes distincts pour construire des représentations de l'information spatiale : d'une part l'utilisation de la proprioception et d'une représentation spatiale indépendante de la position de l'œil et d'autre part l'utilisation d'une copie efférente d'un mouvement oculaire pour mettre à jour une représen-

tation spatiale définie par rapport à la position courante de l'œil. Ce deuxième mécanisme part du principe que, puisque les saccades sont des mouvements oculaires volontaires, il est possible d'en prédire certaines conséquences. En particulier, comme nous l'avons montré, il est possible d'anticiper les positions spatiales que vont occuper des cibles mémorisées après ce mouvement oculaire. Nous avons alors utilisé ce mécanisme pour étendre le modèle de déploiement *covert* de l'attention visuelle, présenté dans le chapitre 2, pour obtenir un modèle de déploiement de l'attention visuelle spatiale avec ou sans mouvement oculaire. Ce dernier mécanisme a été appliqué à une tâche de recherche visuelle avec mouvements oculaires. La tâche considérée contraint fortement la décision d'exécuter une saccade, et il faudrait poursuivre ces travaux pour étudier les mécanismes qui en motivent l'exécution.

Au travers de ces différents modèles, nous avons montré comment une fonction globale, analogue à l'attention visuelle chez le primate, est le résultat émergent de l'interaction de cellules élémentaires simples. Ces travaux sont très encourageants puisqu'ils illustrent comment une fonction globale, non formalisée, émerge d'un substrat générique qui n'est pas dédié, a priori, à l'émergence de cette fonction. La brique de base de ces modèles est une cellule simple, dont le coût en ressources de calculs est faible. Cette cellule ne fait qu'intégrer ses entrées et transmettre un signal résultant de cette intégration aux unités suivantes. C'est en ce sens qu'elle est générique, qu'elle n'est pas dédiée à une fonction particulière, et qu'elle peut être utilisée dans des architectures pour étudier d'autres capacités cognitives complexes comme la navigation, la résolution de problèmes, la planification, la reconnaissance d'objets visuels, qui sont autant de capacités dont l'étude est intéressante aussi bien du point de vue des neurosciences que du point de vue de l'informatique. Pour l'informatique, ce substrat de calculs locaux, numériques et distribués est très intéressant puisqu'il est intrinsèquement parallèle et ainsi, en théorie, implémentable sur une architecture parallèle [Girau, 2007, Boniface, 2000]. Sa nature distribuée rend également ce paradigme robuste aux pannes et au bruit : chaque cellule ne traite ou ne représente qu'une partie de l'information globale en se recoupant partiellement, de telle sorte qu'une perturbation sur l'un des éléments n'entraîne pas nécessairement de conséquences fâcheuses sur le système. Les principes de ce paradigme de calcul sont à l'heure actuelle des sujets de recherche en vue, notamment, de proposer de nouveaux supports matériels de calcul [Berry, 2008].

Néanmoins, le développement d'un mécanisme dans le cadre de ce paradigme reste très complexe puisque la définition de ses paramètres est dans une certaine mesure empirique, notamment parce qu'il est difficile de formaliser la fonction globale qui en émerge. Nous disons que sa définition est empirique dans une certaine mesure puisqu'il ne faut pas oublier que l'inspiration biologique nous apporte beaucoup, en nous éclairant sur les flux d'informations à privilégier ou sur la manière dont l'information peut être traitée (le cortex visuel est spécialisé dans le traitement de l'information visuelle, certaines aires sont sélectives à des orientations, d'autres à des formes plus complexes, les champs oculomoteurs frontaux ont des réponses visuelles, motrices, visiomotrices, etc...). Les paramètres que nous proposons en annexe A pour ces modèles sont un jeu de paramètres parmi d'autres possibles. Ces paramètres sont obtenus par essais/erreurs en se reposant sur l'intuition du modélisateur sur l'influence des différents paramètres du système sur son comportement global. Même si les études mathématiques sur les champs neuronaux considèrent pour le moment des conditions très restrictives sur le système pour qu'une étude analytique soit envisageable, peut être que ce que le modélisateur considère comme intuition pourrait être formalisé pour l'aider dans la définition de ses mécanismes, ce que nous avons commencé à faire

dans [Alexandre et al., 2008].

Si on souhaite appliquer ces paradigmes de calculs, la non maîtrise du résultat émergent est très certainement une faiblesse. Les architectures que nous avons proposées dans cette thèse contiennent une partie non négligeable d'a priori : nous fixons le nombre de cellules, le nombre de couches, les poids entre ces couches, la dynamique du réseau, etc... Ces architectures sont ainsi très rigides et ne sont pas nécessairement les plus adaptées pour résoudre les tâches que nous proposons ni n'ont la possibilité de s'adapter à la dynamique de l'environnement dans lequel le système peut être amené à évoluer. Une direction à explorer pour relâcher cette contrainte est l'apprentissage. L'apprentissage peut se voir à différents niveaux : soit en partant d'une architecture contrainte mais qui peut évoluer au cours du temps, soit en adoptant une approche, que je qualifierais de développementale, dans laquelle les fonctions émergent au fur et à mesure de l'interaction du système avec l'environnement. Cette seconde approche est très certainement la plus prometteuse puisque d'une part elle met l'accent sur la modularité de l'architecture et d'autre part elle nous amène à considérer sous quelle forme cet apprentissage doit avoir lieu, ce qui n'est pas un problème facile. La présentation que nous avons faite du mécanisme d'attention visuelle *covert* et *overt* est un exemple illustratif de ce que nous entendons par modularité. Il est possible d'interpréter les activités des neurones d'une carte comme une fonction élémentaire du comportement global (sélection d'une cible, mémoire spatiale, anticipation). Nous avons montré dans le chapitre 4 que la destruction de la carte d'anticipation conduit à la perte intégrale de la faculté d'anticiper les conséquences d'une saccade oculaire sur la position des cibles mémorisées. Si on considère que le système s'est construit par un apprentissage itératif, dans lequel les fonctions émergent au fur à mesure de l'interaction du système avec son environnement, à la suite de la destruction d'une des parties du système, une population de cellules pourrait être recrutée pour pallier cette "panne" du système. Cette forme de plasticité, dans laquelle une partie du cortex est impliquée dans une fonction autre que la fonction qui lui est habituellement attribuée, est observée expérimentalement chez des sujets humains. Les travaux de [Chollet et al., 1991, Ward et Frackowiak, 2006] montrent en particulier comment le cortex moteur se réorganise pour compenser la perte d'une fonction motrice à la suite d'un accident vasculaire cérébral (AVC). Cet exemple met en avant le fait que le cortex n'est pas une structure rigide mais qu'il peut se réorganiser pour compenser la perte de tout ou partie d'une capacité cognitive.

Les travaux présentés dans cette thèse ne constituent qu'une première étape dans la compréhension des phénomènes émergents et, ne serait-ce que pour un système de vision biologique ou artificielle, plusieurs extensions sont envisageables. En particulier, si on se limite aux modèles présentés dans ce manuscrit, les tâches utilisées contraignent très fortement le déploiement de l'attention visuelle, avec ou sans mouvement oculaire. Une manière de relâcher cette contrainte consisterait à lier les informations spatiales et non spatiales, qui sont pour le moment traitées dans deux voies distinctes et qui ne "communiquent" que par l'intermédiaire des couches visuelles de plus bas niveau, pour que le système construise ses propres représentations, acquiert des connaissances sur l'environnement, pour les utiliser afin de produire ses propres buts internes. Ces perspectives pointent en direction d'un système de perception active [Noe, 2005]. Nous l'avons illustré au travers de la vision, mais on peut également le concevoir pour d'autres modalités sensorielles et motrices. Imaginons par exemple que nous soyons en train de palper avec la main une bouteille, les yeux fermés, de telle sorte que seule la sensation tactile nous permet d'identifier l'objet. La perception d'une bouteille ne se fait qu'au travers de l'exploration tactile de l'objet et

il est surprenant que, malgré le fait que les sensations tactiles soient discontinues et qu'à aucun moment l'intégralité de l'objet n'est sentie, nous avons malgré tout l'impression de percevoir une bouteille.

La notion de perception active met ainsi l'accent sur l'implication du sujet dans la construction d'une perception ; ce n'est plus seulement un sujet passif qui reçoit des informations de l'environnement mais c'est un sujet qui est activement impliqué dans la perception. Cela implique que le sujet soit capable de guider ses actions ; l'anticipation pourrait jouer ici un rôle important. Robert Rosen définit un système anticipateur ainsi : *“A system containing a predictive model of itself and/or its environment, which allows it to change state at an instant in accord with the model's predictions pertaining to a latter instant”*. Dans nos travaux, nous avons surtout montré en quoi la prédiction des conséquences d'une action pouvait être intéressante pour maintenir une représentation spatiale cohérente. Un système qui disposerait en plus d'un modèle prédictif de l'environnement serait capable de générer des hypothèses pour guider efficacement son comportement.

Ces dernières extensions que nous mentionnons restent cantonnées à la vision (attention visuelle, perception visuelle active) et différents travaux, ces dernières années, soulignent l'intérêt du connexionnisme dans l'étude d'autres fonctions cognitives complexes. On peut mentionner, à titre d'exemples, des travaux de modélisation dans ce cadre sur le rôle de l'hippocampe dans la mémoire [Rougier, 2000], sur le rôle du cortex préfrontal dans la navigation [Frezza-Buet, 1999], sur le système moteur [Vitay, 2006] ou sur le langage [Alexandre, 2008]. Le connexionnisme est une science jeune, mais les travaux exposés dans cette thèse ainsi que les différents travaux mentionnés précédemment nous laissent imaginer un avenir prometteur et des résultats intéressants que ce soit pour la maîtrise du calcul distribué, ses applications à la robotique autonome, la compréhension des phénomènes émergents et la mesure dans laquelle les capacités cognitives complexes peuvent s'expliquer dans ce contexte.

Annexe A

Paramètres des modèles

Les modèles décrits dans cette annexe utilisent tous des cartes neuronales qui sont des versions discrètes des champs neuronaux présentés dans la section 2.3. Ces cartes comportent $n \times n$ neurones auxquels est attribuée une position spatiale x_{ij} en fonction de leur position sur la carte $(i, j) \in [0, n - 1]^2$:

$$x_{ij} = \left(\frac{i}{n} - 0.5, \frac{j}{n} - 0.5 \right)$$

La mise à jour de l'activité $u(x_{ij}, t)$ d'un neurone est réalisée en évaluant de manière asynchrone les équations suivantes :

$$\begin{aligned} u(x_{ij}, t + \Delta t) &= f(u(x_{ij}, t) + \Delta u(x_{ij}, t)) \\ \tau \Delta u(x_{ij}, t) &= -u(x_{ij}, t) + \frac{1}{\alpha} \left(\sum_{k=0}^n \sum_{l=0}^n w(\|x_{ij} - x_{kl}\|) u(x_{kl}, t) + s(x_{ij}, t) \right) + h \\ f(x) &= \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x > 0 \end{cases} \\ w(x) &= A_+ e^{-\frac{x^2}{\sigma_+^2}} - A_- e^{-\frac{x^2}{\sigma_-^2}} \end{aligned}$$

Les projections entre deux cartes prennent deux formes que l'on appelle respectivement neurones sigma et neurones sigma-pi. Chaque neurone sigma de la carte *destination* de position x_{ij} intègre ses afférences par une somme pondérée des activités de la carte *source* selon l'équation A.1.

$$s(x_{ij}, t) = C \times \sum_{k,l} \exp\left(-\frac{(i-k)^2 + (j-l)^2}{\sigma_c^2}\right) I_{kl}(t) \quad (\text{A.1})$$

où $I_{kl}(t)$ est l'activité du neurone de la carte *source* de position (k, l) . Pour des raisons de simplicité, nous noterons abusivement par la suite $\sigma_c = 0$ pour des projections point-à-point. Bien que cela puisse paraître incohérent, pour des raisons de concision, nous noterons $\sigma_c = \max$ lorsqu'un neurone intègre ses entrées en calculant le maximum de l'activité des neurones afférents :

$$s(x_{ij}, t) = C \times \max_{kl} I_{kl}(t)$$

Chaque neurone sigma-pi de la carte *destination* de position x_{ij} intègre ses afférences par une somme pondérée du produit de l'activité de la carte *source* et *source₂* selon l'équation A.2.

$$s(x_{ij}, t) = C \times \sum_{k,l} I_{kl}(t) \times J_{\alpha \binom{i}{j} + \beta \binom{k}{l} + \delta}(t) \quad (\text{A.2})$$

où $I_{kl}(t)$ est l'activité du neurone de la carte *source* de position (k, l) et $J_{mn}(t)$ est l'activité du neurone de la carte *source₂* de position (m, n) . Les paramètres α, β, δ permettent de définir la transformation calculée.

A.1 Equations du mécanisme des paragraphes 2.4.3 et 4.4

Dans ce paragraphe, nous détaillons explicitement toutes les équations d'évolution des unités du modèle présenté aux paragraphes 2.4.3 et 4.4.

L'entrée visuelle est filtrée selon quatre dimensions $d \in \{bleu, vert, 45^\circ, 135^\circ\}$. Le résultat de ces filtres est représenté dans les cartes $I_{bleu}, I_{vert}, I_{45^\circ}, I_{135^\circ}$. Pour chaque dimension $d \in \{bleu, vert, 45^\circ, 135^\circ\}$, on définit \bar{d} la dimension antagoniste telle que :

$$\begin{aligned}\overline{bleu} &= vert \\ \overline{vert} &= bleu \\ \overline{45^\circ} &= 135^\circ \\ \overline{135^\circ} &= 45^\circ\end{aligned}$$

Les équations proposées contiennent des variables symboliques dont la valeur est donnée dans les annexes A.2 et A.5.

A.1.1 V4

Le niveau V4 est constitué de quatre cartes de $N \times N$ cellules, avec une carte pour chaque dimension $d \in \{bleu, vert, 45^\circ, 135^\circ\}$. L'équation d'évolution des cellules des quatre cartes sont :

$$\begin{aligned}\forall x \in [1..N]^2, \forall d \in \{bleu, vert, 45^\circ, 135^\circ\} \\ u_{V4,d}(x, t+1) &= f(u_{V4,d}(x, t) + \Delta u_{V4,d}(x, t))\end{aligned}$$

$$\tau \Delta u_{V4,d}(x, t) = -u_{V4,d}(x, t) + \frac{1}{\alpha} u_{input,d}(x, t) (0.25 + 0.5 u_{focus}(x, t) + 0.15 u_{PF,d}(t))$$

A.1.2 Percept

$$u_{percept}(x, t+1) = f(u_{percept}(x, t) + \Delta u_{percept}(x, t))$$

$$\begin{aligned}\tau \Delta u_{percept}(x, t) &= -u_{percept}(x, t) + \frac{1}{\alpha} \left(\sum_y w_{percept \rightarrow percept}(x, y) u_{percept}(y, t) \right) \\ &+ \frac{1}{\alpha} \max_d (u_{V4,d}(x, t)) + h_{percept}\end{aligned}$$

A.1.3 Focus

$$u_{focus}(x, t+1) = f(u_{focus}(x, t) + \Delta u_{focus}(x, t))$$

$$\begin{aligned}
\tau \Delta u_{focus}(x, t) = & -u_{focus}(x, t) + \frac{1}{\alpha} \left(\sum_y w_{focus \rightarrow focus}(x, y) u_{focus}(y, t) \right) \\
& + \frac{1}{\alpha} \left(\sum_y w_{percept \rightarrow focus}(x, y) u_{percept}(y, t) \right) \\
& + \frac{1}{\alpha} (-4.0 u_{wm}(x, t) u_{sw}(t))
\end{aligned}$$

A.1.4 Wm

$$\begin{aligned}
u_{wm}(x, t + 1) &= f(u_{wm}(x, t) + \Delta u_{wm}(x, t)) \\
\tau \Delta u_{wm}(x, t) = & -u_{wm}(x, t) + \frac{1}{\alpha} \left(\sum_y w_{wm \rightarrow wm}(x, y) u_{wm}(y, t) \right) \\
& + \frac{1}{\alpha} \left(\sum_y w_{percept \rightarrow wm}(x, y) u_{percept}(y, t) \right) \\
& + \frac{1}{\alpha} \left(\sum_y w_{focus \rightarrow wm}(x, y) u_{focus}(y, t) \right) \\
& + \frac{1}{\alpha} \left(\sum_y w_{anticipation \rightarrow wm}(x, y) u_{anticipation}(y, t) \right)
\end{aligned}$$

A.1.5 Anticipation

$$\begin{aligned}
u_{anticipation}(x, t + 1) &= f(u_{anticipation}(x, t) + \Delta u_{anticipation}(x, t)) \\
\tau \Delta u_{anticipation}(x, t) = & -u_{anticipation}(x, t) + \frac{1}{\alpha} \left(\sum_y w_{anticipation \rightarrow anticipation}(x, y) u_{anticipation}(y, t) \right) \\
& + \frac{1}{\alpha} \left(0.05 \sum_y u_{wm}(y, t) u_{focus}(\alpha x + \beta y + \delta, t) \right)
\end{aligned}$$

A.1.6 IT

$$\begin{aligned}
\forall d \in \{bleu, vert, 45^\circ, 135^\circ\} \\
u_{IT,d}(t + 1) &= f(u_{IT,d}(t) + \Delta u_{IT,d}(t)) \\
\tau \Delta u_{IT,d}(t) = & -u_{IT,d}(t) + \frac{1}{\alpha} (0.6 u_{IT,d}(t) - 0.6 u_{IT,\bar{d}}(t)) \\
& + \frac{1}{\alpha} max_x(u_{V4,d}(x, t))
\end{aligned}$$

A.1.7 PF

L'activité des unités PF sont définies par l'utilisateur puisqu'elles spécifient quelle est la cible.

A.1.8 Move

$$u_{mv}(x, t + 1) = f(u_{mv}(x, t) + \Delta u_{mv}(x, t))$$

$$\begin{aligned} \tau \Delta u_{mv}(x, t) = -u_{mv}(x, t) &+ \frac{1}{\alpha} (0.5 \sum_d u_{PF,d}(t) u_{IT,d}(t)) \\ &- \frac{1}{\alpha} (0.5 \sum_d u_{PF,d}(t) u_{IT,\bar{d}}(t)) \end{aligned}$$

A.1.9 Switch

$$u_{sw}(x, t + 1) = f(u_{sw}(x, t) + \Delta u_{sw}(x, t))$$

$$\tau \Delta u_{sw}(x, t) = -u_{sw}(x, t) + \frac{1}{\alpha} (\sum_d u_{PF,d}(t) u_{IT,\bar{d}}(t))$$

A.2 Paramètres du mécanisme du paragraphe 2.4.3

Les paramètres des neurones de toutes les cartes de ce modèle (I_* , $V4_*$, IT_* , PF_* , PERCEPT, FOCUS, WM, SWITCH) sont donnés dans la table A.1. Les paramètres des fonctions de voisinage dans chaque carte sont donnés dans la table A.2. Les paramètres des connexions entre les cartes sont définies dans la table A.3. Il y a une particularité pour les cartes IT . Puisque le champ récepteur de ces cellules est très large (il couvre tout le champ visuel), on ajoute des connexions latérales pour amplifier le contraste entre les cellules représentant des attributs antagonistes dans le monde qu'on considère. C'est à dire que les cellules IT s'autoexcitent avec une amplitude de 0.6 et inhibent l'unité IT qui représente l'attribut antagoniste avec une amplitude de -0.6 . On considère que les attributs bleu et vert sont antagonistes ainsi que les attributs 45° et 135° . Si l'espace des attributs était discrétisé moins brutalement, on pourrait imaginer des projections latérales localement excitatrices et inhibitrices sur de plus grandes distances.

<i>Carte</i>	<i>Taille</i>	<i>h</i>	α	τ
I_BLEU, I_VERT, I_PI_4, I_3PI_4	40x40		-	-
V4_BLEU, V4_VERT, V4_PI_4, V4_3PI_4	40x40	0.0	1.0	0.75
IT_BLEU, IT_VERT, IT_PI_4, IT_3PI_4	40x40	0.0	1.5	0.75
PF_BLEU, PF_VERT, PF_PI_4, PF_3PI_4	40x40	-	-	-
PERCEPT	40x40	-0.1	0.5	2.0
FOCUS	40x40	0.0	6.0	0.75
WM	40x40	-0.2	13.0	0.6
MV	1x1	0.0	1.0	0.75
SW	1x1	0.0	0.6	0.75

TAB. A.1 – Paramètres des neurones de chaque carte du modèle.

<i>Carte</i>	<i>A</i>	<i>B</i>	σ_a	σ_b
WM	3.0	0.5	2.0	4.0
FOCUS	1.0	0.65	5.0	$N\sqrt{2}$

TAB. A.2 – Paramètres des fonctions de voisinage des cartes du modèle.

<i>Source</i>	<i>Destination</i>	<i>C</i>	σ_c
I_*	V4_*	0.25	<i>point à point</i>
I_*, PF_*	V4_*	0.15	<i>point à point</i>
I_*, FOCUS	V4_*	0.5	<i>point à point</i>
V4_*	IT_*	1.0	<i>max</i>
V4_*	percept	1.0	<i>max</i>
PERCEPT	FOCUS	0.4	2.0
PERCEPT	WM	0.3	2.0
FOCUS	WM	0.2	2.0
SW, WM	FOCUS	-4.0	<i>point à point</i>
PF, IT	MV	-	-
PF, IT	SW	-	-

TAB. A.3 – Paramètres des connexions entre les cartes du modèle.

A.3 Paramètres du mécanisme du paragraphe 4.2

Les paramètres des neurones de toutes les cartes de ce modèle (LIP, FEFV, FEFVM, FEFM, SEF, DLPFC et SC) sont donnés dans la table A.4. Les paramètres des fonctions de voisinage dans chaque carte sont donnés dans la table A.5. Les paramètres des connexions entre les cartes sont définies dans les tables A.6 et A.7.

<i>Carte</i>	<i>Taille</i>	h	α	τ
LIP	30*30	0.0	13.0	0.75
FEFV	30*30	0.0	13.0	0.75
FEFVM	30*30	0.0	5.0	0.75
FEFM	30*30	0.0	5.0	0.75
SEF	30*30	0.0	13.0	0.75
DLPFC	30*30	0.0	10.0	0.75
SC	30*30	0.0	13.0	0.75

TAB. A.4 – Paramètres des neurones de chaque carte du modèle.

<i>Carte</i>	A	B	σ_a	σ_b
LIP	1.2	1.0	3.0	4.0
FEFV	1.2	1.0	3.0	4.0
FEFVM	1.2	1.0	3.0	4.0
FEFM	0.95	0.65	5.0	30.0
SEF	2.5	1.0	2.0	4.0

TAB. A.5 – Paramètres des fonctions de voisinage des cartes du modèle.

<i>Source</i>	<i>Destination</i>	C	σ_c
INPUT	LIP	0.85	2.0
LIP	FEFV	2.0	2.0
FEFV	FEFVM	0.25	2.0
FEFVM	FEFM	0.4	2.0
FEFM	LIP	0.25	2.0
SEF	DLPFC	2.35	1.2
DLPFC	SEF	2.4	1.2
FEFM	SC	2.0	2.0
FIXATION	SC	-1.0	<i>point à point</i>

TAB. A.6 – Paramètres des connexions entre les cartes du modèle.

<i>Sources</i>	<i>Destination</i>	α	β	δ	C
LIP, EYE POSITION	SEF	1.0	-1.0	$\begin{pmatrix} w/2.0 \\ h/2.0 \end{pmatrix}$	0.2
SEF, EYE POSITION	FEFVM	-1.0	1.0	$\begin{pmatrix} w/2.0 \\ h/2.0 \end{pmatrix}$	1.0

TAB. A.7 – Paramètres des connexions entre les cartes du modèle pour les unités sigma-pi

A.4 Paramètres du mécanisme du paragraphe 4.3

Les paramètres des neurones de toutes les cartes de ce modèle (INPUT, FOCUS, WM, THAL_WM, ANTICIPATION, SWITCH) sont donnés dans la table A.8. Les paramètres des fonctions de voisinage dans chaque carte sont donnés dans la table A.9. Les paramètres des connexions entre les cartes sont définies dans la table A.10.

<i>Carte</i>	<i>Taille</i>	h	α	τ
INPUT	40*40	0.0	13.0	0.75
FOCUS	40*40	-0.05	14.0	0.75
WM	40*40	-0.2	13.0	0.6
THAL_WM	40*40	0.0	13.0	0.6
ANTICIPATION	40*40	0.0	5.0	4.0
SWITCH	1*1	0.0	1.0	10.0

TAB. A.8 – Paramètres des neurones de chaque carte du modèle.

<i>Carte</i>	A	B	σ_a	σ_b
FOCUS	1.7	0.65	4.0	25.0
WM	2.5	1.4	2.0	3.5
ANTICIPATION	1.6	1.0	3.0	4.0

TAB. A.9 – Paramètres des fonctions de voisinage des cartes du modèle.

<i>Source</i>	<i>Destination</i>	<i>C</i>	σ_c
INPUT	FOCUS	0.25	2.0
INPUT	WM	0.3	2.0
FOCUS	WM	0.20	2.0
WM	THAL_WM	2.35	1.5
THAL_WM	WM	2.40	1.5
ANTICIPATION	WM	0.2	2.0
WM, SWITCH	FOCUS	-5.0	<i>point à point</i>

TAB. A.10 – Paramètres des connexions entre les cartes du modèle.

<i>Sources</i>	<i>Destination</i>	α	β	δ	<i>C</i>
WM, FOCUS	ANTICIPATION	-1.0	1.0	$\begin{pmatrix} w/2.0 \\ h/2.0 \end{pmatrix}$	0.05

TAB. A.11 – Paramètres des connexions entre les cartes du modèle pour les unités sigma-pi

A.5 Paramètres du mécanisme du paragraphe 4.4

Les paramètres des neurones de la carte ANTICIPATION sont donnés dans la table A.12. Les paramètres des fonctions de voisinage de cette carte sont donnés dans la table A.13. Les paramètres des connexions entre les cartes sont définies dans la table A.14. Les paramètres des cartes et des projections du reste du modèle sont les mêmes que ceux de l'annexe A.2.

<i>Carte</i>	<i>Taille</i>	<i>h</i>	α	τ
ANTICIPATION	40*40	0.0	5.0	4.0

TAB. A.12 – Paramètres des neurones de chaque carte du modèle.

<i>Carte</i>	<i>A</i>	<i>B</i>	σ_a	σ_b
ANTICIPATION	1.6	1.0	3.0	4.0

TAB. A.13 – Paramètres des fonctions de voisinage des cartes du modèle.

<i>Source</i>	<i>Destination</i>	<i>C</i>	σ_c
ANTICIPATION	WM	0.30	2.0

TAB. A.14 – Paramètres des connexions entre les cartes du modèle.

<i>Sources</i>	<i>Destination</i>	α	β	δ	<i>C</i>
WM, FOCUS	ANTICIPATION	-1.0	1.0	$\begin{pmatrix} w/2.0 \\ h/2.0 \end{pmatrix}$	0.05

TAB. A.15 – Paramètres des connexions entre les cartes du modèle pour les unités sigma-pi

Annexe B

Décomposition en valeurs singulières

B.1 Introduction

La Décomposition en Valeurs Singulières (SVD, *Singular Value Decomposition*) est une méthode de décomposition d'une matrice, réelle ou complexe, de dimensions quelconques que l'on peut appliquer pour optimiser le calcul d'un produit de convolution. Dans le cas de matrices réelles, carrées, ce théorème peut s'énoncer de la manière suivante :

Théorème 1 *Soit M une matrice de $\mathbb{R}_{m,n}$, il existe une matrice unitaire $U \in \mathbb{R}_{m,m}$, une matrice $\Sigma \in \mathbb{R}_{m,n}$ diagonale dont les coefficients diagonaux sont positifs ou nuls, et une matrice unitaire $V \in \mathbb{R}_{n,n}$ telles que :*

$$M = U \cdot \Sigma \cdot V$$

B.2 Application au calcul d'un produit de convolution

B.2.1 Position du problème

Soit M une matrice réelle de taille $n \times n$ et K une matrice réelle de taille $2m+1 \times 2m+1$. On utilise un noyau de taille impaire uniquement pour des raisons de simplicité des notations mais le résultat reste valable pour des matrices de taille paire. On peut écrire le produit convolution de la forme suivante :

$$\forall (i, j) \in [1..n]^2 : I_{i,j} = \sum_{k,l} M_{i+k-m, j+l-m} K_{k,l} \quad (\text{B.1})$$

Cette notation n'est pas la forme standard d'un produit de convolution mais elle est équivalente et nous utiliserons cette notation par la suite pour l'optimisation d'un produit de matrice ; Dans l'équation B.1, nous n'avons pas précisé les bornes des sommes. Elles ne sont évidemment pas

infinies mais bornées par les dimensions des matrices :

$$\begin{aligned} \max(i - n + 1, -m) &\leq k \leq \min(i, m) \\ \max(j - n + 1, -m) &\leq l \leq \min(j, m) \end{aligned}$$

Pour des raisons de simplicité des notations, nous ne précisons pas par la suite les bornes des sommes.

D'après le théorème de Décomposition en Valeurs Singulières, il existe deux matrices unitaires U et V et une matrice diagonale S telles que :

$$K = USV \quad (\text{B.2})$$

La matrice S est diagonale à coefficients positifs ou nuls. Ses valeurs diagonales sont appelées valeurs singulières de la matrice K et sont les valeurs propres de la matrice KK^* où K^* est la matrice transposée de K . Notons $VS(K)$ l'ensemble des valeurs singulières de K . Par définition :

$$\forall i, j \in [1..2m+1]^2 : S_{i,j} = \begin{cases} 0 & \text{si } i \neq j \\ \sigma_i & \text{si } i = j \text{ avec } \sigma_i \in VS(K) \end{cases}$$

Ainsi, on peut réécrire B.2 par :

$$\begin{aligned} \forall (k, l) \in [1..2m+1]^2 : K_{k,l} &= (USV)_{k,l} \\ &= \sum_o \left(\sum_p U_{k,p} S_{p,o} \right) V_{o,l} \\ &= \sum_o U_{k,o} S_{o,o} V_{o,l} \\ &= \sum_o U_{k,o} \sigma_o V_{o,l} \end{aligned}$$

Revenons-en à notre calcul initial. L'équation B.1 peut alors s'écrire :

$$\begin{aligned} \forall (i, j) \in [1..n]^2 : I_{i,j} &= \sum_{k,l} M_{i+k-m, j+l-m} K_{k,l} \\ &= \sum_{k,l} M_{i+k-m, j+l-m} \sum_o U_{k,o} \sigma_o V_{o,l} \\ &= \sum_o \sum_k \sum_l M_{i+k-m, j+l-m} \sum_o U_{k,o} \sigma_o V_{o,l} \\ &= \sum_o \sum_l \sigma_o V_{o,l} \left[\sum_k M_{i+k-m, j+l-m} U_{k,o} \right] \end{aligned}$$

L'intérêt de cette dernière écriture est de pouvoir décomposer le calcul B.1 en deux étapes, comme suit :

$$\forall i, j \in [1..n]^2, \sigma_o \in VS(K) : temp_{i,j,o} = \sum_k M_{i+k-m, j+l-m} U_{k,o} \quad (\text{B.3})$$

$$I_{i,j} = \sum_o \sum_l \sigma_o V_{o,l} temp_{i,j,o} \quad (\text{B.4})$$

Notons $m' = 2m + 1$ l'une des dimensions de K . Le calcul naïf de B.1 coûte de l'ordre de $n^2 \times m'^2$ opérations. Notons $r = \text{card}(VS(K))$ le nombre de valeurs singulières non nulles de K qui est également son range. Les deux étapes B.4 sont de l'ordre de $n^2 \times r \times m'$ opérations. Puisque le rang d'une matrice est toujours plus petit que la plus petite des dimensions alors $r \leq m'$.

Pour comparer la complexité des deux méthodes (naïve et avec SVD), il faut inclure le temps de calcul de la décomposition SVD qui peut être calculée en $O(m'^3)$ [Golub et VanLoan, 1996, Press et al., 2007]. Néanmoins, pour l'application que nous allons considérer au paragraphe suivant, on peut négliger ce temps de calcul. En effet, nous allons appliquer la décomposition SVD pour évaluer les afférences d'un neurone. La matrice K correspond alors à la matrice des poids des liens afférents de la cellule. Sous la condition que les poids ne varient pas pendant la simulation, la matrice K est constante, sa décomposition SVD l'est donc également et il n'est nécessaire de la calculer qu'à la première exécution de la simulation.

B.3 Application aux réseaux de neurones

B.3.1 Principe

Il est relativement courant dans nos modèles de calculer la contribution de neurones afférents ou latéraux en convoluant un ensemble d'unités par un noyau représentant la matrice de projection. Afin d'optimiser ces calculs, et si on ne considère pas d'apprentissage (ce qui constitue une restriction importante), il suffit de précalculer la décomposition en valeurs singulières des matrices de projection. Ce calcul étant réalisé au lancement de la simulation, il n'entraîne pas de surcoût pendant la simulation.

B.3.2 Cas des projections afférentes

Considérons deux couches de neurones, pour le moment distinctes. Chaque neurone x_{ij} de la couche L_2 calcule la contribution des neurones afférents, de la couche L_1 par :

$$\forall x_{ij} \in L_2 : I_{i,j} = \sum_{x_{kl} \in L_1} w(i-k, k-l) \cdot u(x_{kl}) \quad (\text{B.5})$$

L'équation B.5 est l'écriture classique de la somme pondérée des afférences d'une cellule. En jouant simplement sur les indices et en considérant une matrice de poids W de taille $2m + 1 \times 2m + 1$ centrée en m, m , on peut réécrire cette équation sous la forme :

$$\forall x_{ij} \in L_2 : I_{i,j} = \sum_{k,l} W_{k,l} u_{i+k-m, j+l-m}$$

Cette équation est très ressemblante à l'équation B.1 du paragraphe précédent. On peut ainsi utiliser le même raisonnement et utiliser la décomposition SVD de la matrice de poids $W = USV$. L'évaluation des afférences de la cellule se calcule à deux étapes.

Etape 1 : Notons $VS(W)$ les valeurs singulières de la matrice de poids W . On construit une matrice temporaire $temp$ comme suit :

$$\forall i, j, \sigma_o \in VS(W) : temp_{i,j,o} = \sum_k u_{i+k-m, j+l-m} U_{k,o} \quad (B.6)$$

Etape 2 : Etant donnée la matrice temporaire $temp$ calculée, les afférences de la cellule i, j sont calculées comme suit :

$$\forall i, j : I_{i,j} = \sum_o \sum_l \sigma_o V_{o,l} temp_{i,j,o} \quad (B.7)$$

Ainsi, pour calculer les afférences des unités de la couche L_2 , on commence par calculer la matrice $temp$ par l'équation B.6, commune à toutes les unités de la couche L_2 puis on évalue indépendamment les afférences de chaque unité grâce à l'équation B.7. En supposant les poids partagés (c'est à dire que la matrice de poids entre la couche L_1 et la couche L_2 est la même quel que soit le neurone de la couche L_2) et n'évoluant pas pendant la simulation (c'est à dire qu'on ne réalise pas d'apprentissage), alors la décomposition SVD de la matrice de poids peut être calculée à l'exécution de la simulation. Ainsi, le calcul des afférences des cellules ne se fait qu'en évaluant les deux équations B.6 et B.7.

B.3.3 Cas des projections latérales, évaluation asynchrone par carte

Si le réseau est simulé en utilisant un asynchronisme total, dès qu'une unité est évaluée, elle diffuse sa nouvelle activité. Ainsi, dans le cas d'un asynchronisme total, il faut adjoindre aux deux étapes précédentes une étape supplémentaire qui consiste à modifier les matrices temporaires des cellules **efférentes**. Pour les simulations présentées dans cette thèse, on utilise un asynchronisme partiel qu'on va appeler asynchronisme par carte. En effet, l'évaluation des cartes est séquentielle et l'évaluation des unités au sein d'une carte est asynchrone totale. En particulier, les cartes sont évaluées les unes après les autres. Ainsi, la mise à jour des matrices temporaires, dans cette forme d'asynchronisme, n'est nécessaire que si les unités ont des projections latérales. Dans ce cas, après chaque mise à jour de l'activité d'une cellule, il est nécessaire de mettre à jour la matrice temporaire de la carte à laquelle appartient la cellule selon l'équation suivante :

$$\forall k, \sigma_o \in VS(W) : temp_{x-k,y,o} = \sigma_o U_{m+k,o} [u_t(x, y) - u_{t-1}(x, y)] \quad (B.8)$$

Dans le cas d'un asynchronisme total, il aurait fallu mettre à jour les matrices temporaires de toutes les cartes contenant des **efférences** de l'unité, c'est à dire des unités sur lesquelles cette unité se projette.

B.3.4 Evaluation des performances

Dans ce paragraphe, on présente brièvement les résultats d'occupation mémoire et de temps de calcul pour la simulation de deux architectures avec ou sans SVD. Les tests ont été réalisés sur un Intel Core 2 Duo avec 2Go de mémoire.

Architecture à deux cartes On considère dans ce paragraphe une architecture constituée de deux cartes qu'on appelle INPUT et FOCUS. La carte INPUT se projette sur la carte FOCUS avec des projections gaussiennes et les unités de la carte FOCUS ont des projections latérales en différence de gaussiennes. La figure B.1 présente le temps mis pour calculer 100 pas ainsi que l'occupation mémoire pour la simulation, dans laquelle les projections sont définies avec ou sans SVD. Sur la figure B.1a sont représentés les temps mis pour calculer 100 pas avec ou sans SVD. Lorsque la dimension des cartes augmentent, la simulation utilisant la décomposition SVD va presque 4 fois plus vite. Sur la figure B.1b est représentée l'occupation mémoire, en pourcentage de la mémoire totale, pour les deux types de projections (avec ou sans SVD), en fonction de la taille des cartes. On constate qu'au delà de cartes de 55×55 cellules, la mémoire sature avec les projections normales, tandis qu'en utilisant les projections SVD, puisque les poids sont partagés, la mémoire sature moins vite.

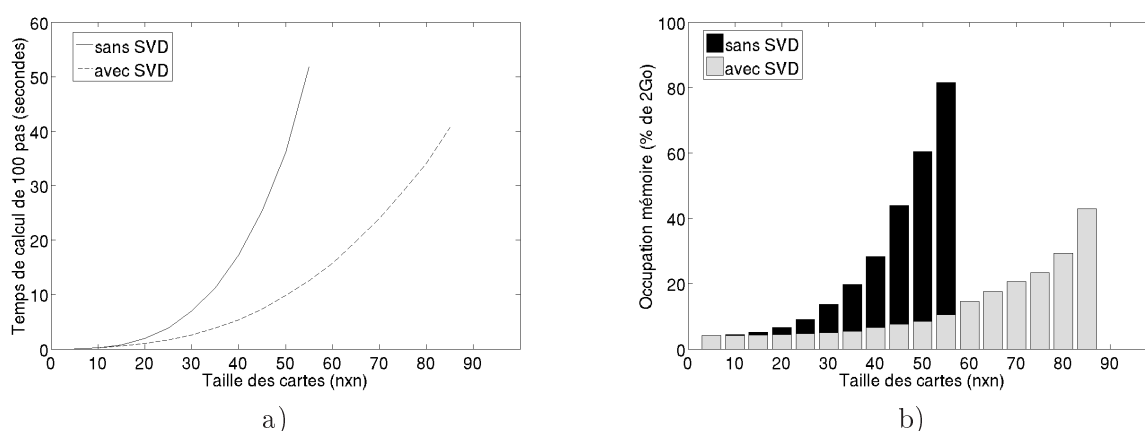


FIG. B.1 – L'architecture simulée contient deux cartes INPUT et FOCUS. La carte INPUT a des projections gaussiennes sur FOCUS. Les unités de la carte FOCUS ont des projections latérales en différence de gaussiennes. a) Temps mis pour simuler 100 pas avec ou sans SVD. La courbe du temps de calcul sans SVD est limitée à des cartes de 55×55 cellules puisque la mémoire sature au delà. b) Occupation mémoire pour la simulation avec ou sans SVD, en fonction de la taille des cartes. On constate par exemple que pour une même occupation mémoire, on peut simuler un réseau avec des cartes de 45×45 cellules sans SVD et un réseau avec des cartes de 85×85 cellules avec SVD.

Architecture à quatre cartes On considère ici une architecture constituée de quatre cartes similaires à celle présentée dans la section 2.4.2. Sans décrire en détail l'architecture, on utilise cinq projections gaussiennes ainsi que trois projections latérales en différence de gaussiennes. La figure B.2a illustre le temps pour calculer 100 pas avec ou sans SVD. La courbe du temps de calcul sans SVD est tronquée à 35×35 cellules puisque la mémoire sature pour des dimensions plus grandes. A titre d'exemple, à temps de simulation constant, on peut simuler un réseau avec des cartes de 35×35 cellules sans SVD et un réseau avec des cartes de 60×60 cellules avec SVD. La figure B.2b illustre l'occupation mémoire pour les deux simulations, en fonction de la dimension des cartes. Le gain en occupation mémoire est plus important avec cette architecture qu'avec l'architecture du paragraphe précédent. Par exemple, à occupation mémoire constante, on peut simuler un réseau avec des cartes de 30×30 cellules sans SVD et un réseau avec des

cartes de 80×80 cellules avec SVD.

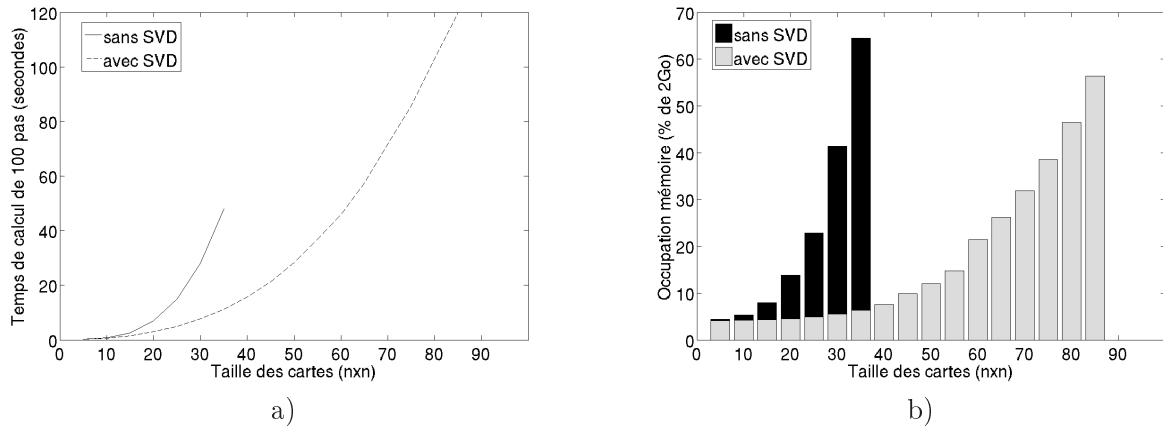


FIG. B.2 – On considère une architecture constituée de 4 cartes. Sans détailler l'architecture, on distingue cinq projections gaussiennes et trois projections latérales en différence de gaussiennes. a) Temps mis pour simuler 100 pas avec ou sans SVD. La courbe du temps de calcul sans SVD est limitée à des cartes de 30×30 cellules puisque la mémoire sature au delà. b) Occupation mémoire pour la simulation avec ou sans SVD, en fonction de la taille des cartes. On constate par exemple que pour une même occupation mémoire, on peut simuler un réseau avec des cartes de 30×30 cellules sans SVD et un réseau avec des cartes de 80×80 cellules avec SVD.

Annexe C

Distributed Asynchronous Numerical and Adaptative computations

C.1 Introduction

D.A.N.A. est l'acronyme de "Distributed Asynchronous Numerical and Adaptative", une plateforme développée par N. Rougier que j'ai utilisé tout au long de ma thèse pour réaliser mes simulations. Cette plateforme logiciel est constituée d'un coeur écrit en C++ et interfacé au langage Python. L'interfaçage en Python permet une interaction très facile et très riche avec la simulation tout en profitant de l'ensemble des outils développés par la communauté (matplotlib pour les calculs scientifiques et le tracé de graphiques, pyopengl pour de la visualisation 3D, ...). La librairie est encore en cours de développement donc nous ne présentons qu'un aperçu de la librairie dans l'état de développement dans lequel elle est au moment de réaliser nos simulations.

DANA est divisé en différents packages compilés indépendamment. Plutôt que de faire une liste exhaustive des différentes classes et méthodes disponibles dans chaque package, on va voir comment les différents termes distribué, asynchrone, numérique et adaptatif sont instanciés. Je présenterais également succinctement cinq packages que j'ai développé, à savoir :

- `dana.learn` pour réaliser de l'apprentissage ;
- `dana.sigmapi` pour coder des unités sigma-pi ;
- `dana.svd` qui permet d'optimiser les calculs en reposant sur la décomposition en valeurs singulières présentés dans l'annexe B ;
- `dana.image` pour traiter des images et alimenter un réseau DANA avec les résultats ;
- `gpython.world` qui propose un environnement virtuel constitué d'objets basiques et dans lequel un robot virtuel peut naviguer.

C.2 Présentation des différents packages

C.2.1 `dana.core` : la librairie de base, calcul local et asynchrone

Description

La librairie `dana.core` est la librairie de base qui définit l'ensemble des classes permettant de construire un réseau, constitué de cartes, elles-mêmes constituées d'unités. Ce package fournit ainsi le cadre permettant de définir les unités élémentaires de calculs. Elle définit également une classe `LINK` qui représente un lien entre deux unités, et qui contient trois attributs :

- une unité source,
- une unité destinatrice,
- un poids.

On dispose de deux façons pour connecter des unités entre elle. D'une part il est possible de les connecter une par une, depuis le script de simulation en spécifiant une unité source, destinatrice et un poids. Il est également possible d'utiliser une définition plus élaborée des poids. Pour ce faire, la classe `PROJECTION` utilise quatre attributs, chaque attribut pouvant prendre un ensemble limité de valeurs :

- une forme : `Point`, `Box`, `Disc` ;
- une distance : `Euclidian`, `Manhattan`, `Max` ;
- une densité : `Full`, `Sparse`, `Sparsen` qui définissent chacune une probabilité de créer une connexion ;
- un profil : `Constant`, `Linear`, `Uniform`, `Gaussian`, `DoG` (*Difference of Gaussians*).

Ce package définit également le schéma d'évaluation asynchrone des unités. Les simulations étant exécutées sur une machine séquentielle, l'asynchronisme ne peut être que simulé. Il existe différentes façons de simuler l'asynchronisme sur une machine séquentielle :

- l'évaluation de chaque unité du réseau est conditionnée par une probabilité. Ainsi, à chaque pas d'évaluation, une unité peut ou non être évaluée,
- les unités sont regroupées dans un vecteur, mélangé aléatoirement à chaque pas d'évaluation. Dans ce cas, à chaque pas de temps, toutes les unités sont évaluées une fois, en ce dans un ordre aléatoire,
- plusieurs séquences d'évaluation des unités sont générées à l'exécution de la simulation et à chaque pas d'évaluation, l'une de ces séquences d'évaluation est choisie aléatoirement.

Dans `DANA`, le schéma d'évaluation n'est pas rigide et il est possible de le modifier assez facilement. Les simulations présentées dans cette thèse ont néanmoins été effectuées avec le troisième schéma d'évaluation dans lequel plusieurs séquences d'évaluation sont générées à l'avance et, à chaque pas de temps, l'une de ces séquences est choisie aléatoirement.

Exemples

La structure générale des scripts de simulation est :

- inclusion des bibliothèques,
- construction du réseau, des cartes, des unités,
- définition des liens entre les unités,
- définition de méthodes ad-hoc pour la simulation,
- création de la fenêtre d'affichage.

Le script C.1 définit un réseau très simple constitué d'une seule carte, contenant 30×30 unités. Les connexions latérales sont une différence de gaussiennes. La figure C.1 illustre les projections ainsi construites (le poids des connexions est représenté par un gradient de couleur de bleu vers rouge correspondant à des projections inhibitrices puis excitatrices).

```

import dana.core as core
import dana.projection as proj

model = core.Model()
net = core.Network ()
model.append(net)
width = 30
height = width

Map = core.Map ( (width,height), (0,0) )
Map.append(core.Layer())
net.append(Map)

p = proj.Projection()
p.self_connect = True
p.distance = proj.distance.Euclidean(True)
p.density = proj.density.Full()
p.profile = proj.profile.DoG(2.20,3.0/width,0.55,1.0)
p.shape = proj.shape.Disc(2.0)
p.src = Map[0]
p.dst = Map[0]
p.connect()

```

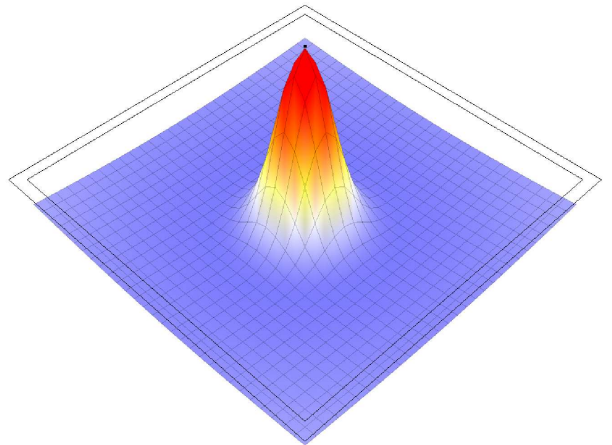


FIG. C.1 – Le script permet de créer une carte de 30×30 cellules. Les connexions latérales dans la carte sont définies par une différence gaussienne, illustrée à droite.

C.2.2 dana.cnft : définition de l'équation d'évolution des cellules, calcul numérique

Description

Le package `dana.cnft` introduit les unités dont l'activité est régie par l'équation d'évolution du premier ordre discrétisée suivante :

$$u(t+1) = f\left(u(t) + \frac{1}{\tau}(-u(t) + h + \frac{lateral + input}{\alpha})\right) \quad (C.1)$$

$$f(x) = \begin{cases} max_{act} & \text{si } x \geq max_{act} \\ x & \text{si } min_{act} \leq x < max_{act} \\ min_{act} & \text{si } x < min_{act} \end{cases} \quad (C.2)$$

$$(C.3)$$

où α, τ sont des paramètres permettant d'ajuster la dynamique de l'unité, h son taux de décharge spontanée et *lateral* et *input* respectivement la contribution des unités d'une même carte et la contribution des unités des autres cartes. La distinction entre les unités latérales (pour le terme *lateral*) et afférentes (pour le terme *input*) est uniquement conceptuelle.

Exemples

L'utilisation d'unités du package `dana.cnft` est relativement simple, il suffit de le spécifier lorsque les cartes sont construites. Il est par ailleurs nécessaire de définir les paramètres de l'équation, soit $\alpha, \tau, min_{act}, max_{act}, h$. L'illustration C.2 est un exemple de script permettant d'utiliser des unités de la librairie `dana.cnft`.

```
import dana.core as core
import dana.projection as proj
import dana.cnft as cnft

model = core.Model()
net = core.Network ()
model.append(net)
width = 30
height = width

Focus = core.Map ( (width,height), (1,0) )
Focus.append(core.Layer())
Focus[0].fill(cnft.Unit)
Focus.name = 'Focus'

Focus.spec = cnft.Spec()
Focus.spec.tau = 0.75
Focus.spec.baseline = 0.0
Focus.spec.alpha = 13.5
Focus.spec.min_act = 0.0
Focus.spec.max_act = 1.0

net.append(Focus)
```

FIG. C.2 – Pour utiliser les unités définies par `dana.cnft`, il suffit de le mentionner à la construction de la carte. Il est également nécessaire de spécifier les paramètres qui régissent la dynamique des unités. Ces paramètres sont communs à toutes les unités d'une même carte.

C.2.3 glpython : une librairie de visualisation

Description

La librairie `glpython` fournit l'ensemble des objets et des méthodes qui permettent de visualiser un réseau. Il est nécessaire de construire deux objets, un objet *figure* et un objet *window*. L'objet *window* est le conteneur principal dans lequel plusieurs figures peuvent être insérées. En particulier, il faut définir un objet *figure* qui contient le réseau. On peut également définir des zones et un certain nombre d'autres objets.

Exemples

Les illustrations C.3 représentent la fenêtre de visualisation du réseau. Aucun réseau n'apparaît puisqu'aucun réseau n'a été défini dans le script. En bas de la fenêtre apparaît un terminal python grâce auquel il est possible de facilement interagir avec la simulation (lancer la simulation, l'arrêter, modifier les paramètres des poids, des unités, visualiser les activités des cellules, etc...).

```

from glpython import window
from dana.visualization.glpython import Figure

fig = Figure()
win,fig = window (figure=fig, has_terminal=True)
fig.network (net, style='flat', show_colorbar=False)
fig.text (size=1, position = (.5, -.05), text="Un exemple de texte")
win.show()

```



FIG. C.3 – Le script illustre quels packages inclure et quelles méthodes appeler pour construire la fenêtre graphique. Il est complémentaire aux scripts présentés précédemment et qui permettent de construire un réseau. La figure de droite illustre la fenêtre de visualisation.

C.2.4 dana.learn : une librairie d'apprentissage, calcul adaptatif

Description

Le package `dana.learn` permet d'introduire de l'apprentissage dans les réseaux de neurones dans le cadre définis par DANA. Le principe de la définition d'une règle d'apprentissage repose sur la proposition de W. Gerstner [Gerstner et Kistler, 2002] d'une définition assez large d'une

règle d'apprentissage :

$$\frac{dw_{ij}}{dt} = F(w_{ij}, \nu_i, \nu_j) \quad (\text{C.4})$$

où w_{ij} est le poids entre l'unité pré-synaptique j et l'unité post-synaptique i , ν_j et ν_i sont respectivement l'activité du neurone pré-synaptique et du neurone post-synaptique. C'est une définition assez générale qui repose sur le principe de localité du postulat de Hebb.

Le deuxième principe du postulat de Hebb est le principe de coopération qu'on retrouve si, en considérant la fonction F suffisamment dérivable, on calcule un développement de Taylor de l'expression précédente autour de $\nu_i = \nu_j = 0$:

$$\frac{dw_{ij}}{dt} = c_0(w_{ij}) + c_1^{post}(w_{ij})\nu_i + c_1^{pre}(w_{ij})\nu_j \quad (\text{C.5})$$

$$+ c_2^{post}(w_{ij})\nu_i^2 + c_2^{pre}(w_{ij})\nu_j^2 + c_2^{corr}(w_{ij})\nu_i\nu_j + \mathcal{O}(\nu^3) \quad (\text{C.6})$$

Le principe de coopération, qui stipule que l'efficacité synaptique dépend de la coactivation des unités pré et post-synaptiques est représenté par le terme c_2^{corr} . En le prenant égal à une constante et en considérant les autres coefficients du développement de Taylor nuls, on retrouve la règle de Hebb. W. Gerstner souligne qu'on peut considérer un développement de Taylor à un ordre plus élevé et la définition des coefficients (qui peuvent être des fonctions) conduit à différentes règles d'apprentissage :

- La règle de Hebb $\frac{dw_{ij}}{dt} = \gamma\nu_i\nu_j$ est obtenue par $c_2^{corr}(w_{ij}) = \gamma$, tous les autres coefficients étant nuls,
- La règle de Oja [Oja, 1982] $\frac{dw_{ij}}{dt} = \gamma(\nu_i\nu_j - w_{ij}\nu_i^2)$ est obtenue par $c_2^{post} = -\gamma w_{ij}$, $c_2^{corr} = \gamma$.

Dans `dana.learn`, une règle d'apprentissage est définie en spécifiant les paramètres c_k du développement de Taylor. Comme nous l'avons vu, ces paramètres peuvent être des fonctions mais pour des raisons de simplicité d'implémentation, le package se limite à des fonctions polynomiales en w_{ij} . Ainsi, chaque paramètre c_k est défini en spécifiant les coefficients d'un polynôme en w_{ij} :

$$c_k = \sum_l b_l(w_{ij})^l \quad (\text{C.7})$$

Les méthodes de la classe `Learner` permettent de :

- définir la carte source des liens à apprendre ;
- définir la carte destinatrice des liens à apprendre ;
- ajouter un terme à la règle d'apprentissage ;
- apprendre les liens en donnant en paramètre le taux d'apprentissage.

Un terme de la règle d'apprentissage est défini par un tableau à trois composantes :

- la puissance à laquelle est élevée l'activité pré-synaptique ;
- la puissance à laquelle est élevée l'activité post-synaptique ;
- un tableau qui contient les coefficients du polynôme qui définit le coefficient du développement de Taylor de la règle d'apprentissage.

Exemples

Les deux figures C.4 illustrent la définition de la règle de Hebb et de la règle de Oja. Après avoir construit un objet de la classe `Learner`, il suffit de définir chacun des termes de la règle d'apprentissage selon le formalisme présenté au paragraphe précédent. Lorsque la règle est définie, un pas d'apprentissage est effectué en appelant la méthode `learn` et en lui spécifiant le taux d'apprentissage.

<pre> import dana.learn as learn learner = learn.Learner() learner.set_source(source_map) learner.set_destination(destination_map) learner.add_one([1,1,[1,0]]) learner.connect() # Apprentissage learner.learn(lrate) </pre>	<pre> import dana.learn as learn learner = learn.Learner() learner.set_source(source_map) learner.set_destination(destination_map) learner.add_one([1,1,[1]]) learner.add_one([2,0,[0,-1]]) learner.connect() # Apprentissage learner.learn(lrate) </pre>
---	---

FIG. C.4 – Les deux scripts illustrent comment définir une règle de Hebb ou une règle de Oja. La règle est définie en spécifiant chacun de ses termes grâce aux appels de la méthode `add_one`.

C.2.5 `dana.sigmapi` : Unités Sigma-Pi

Le package `dana.sigmapi` définit la classe et les méthodes permettant de créer et de connecter des unités sigma-pi. La seule différence entre une unité sigma-pi et une unité sommative est la manière dont l'unité intègre ses entrées. La librairie définit ainsi un type `LINK` ainsi qu'un type `PROJECTION` permettant de définir ces liens. Un objet de `LINK` contient le tableau des unités dont le produit des activités doit être calculé, ainsi qu'un poids. Cette implémentation permet ainsi de définir l'intégration d'entrées de la forme :

$$input = \sum_i w_i link_i \quad (C.8)$$

$$= \sum_i w_i \prod_{u \in E_i} u \quad (C.9)$$

où E_i représente l'ensemble des unités connectées par le lien $link_i$.

Bien qu'il soit possible de définir les liens à la main depuis le script de simulation, la classe `PROJECTION` introduit un certains nombres de méthodes permettant de créer des liens d'une "forme" prédéfinie, utilisés pour les simulations présentées dans ce manuscrit, et qui manquent certainement de généralité, à comparer aux méthodes de définition des liens proposées dans la librairie `dana.core`.

C.2.6 `dana.svd` : Optimisation des calculs

Description

Dans le cas où les matrices de poids ne sont pas apprises et restent donc constantes tout au long de la simulation, il est possible d'optimiser l'intégration des entrées d'une unité en utilisant la décomposition en valeurs singulières (SVD) comme nous l'avons mentionné dans l'annexe B. Le package `dana.svd` est la librairie permettant d'utiliser la décomposition SVD. On ne va pas, dans ce paragraphe, entrer dans une description trop détaillée de la librairie mais simplement évoquer les points essentiels.

La librairie propose trois types d'évaluation :

- standard, comme proposé dans la librairie de base `dana.core`,
- partagée : les poids sont partagés par les unités d'une même carte et sont stockés dans cette carte,
- `svd` : les poids sont partagés et évalués avec la méthode de décomposition en valeurs singulières.

Pour définir quel type d'évaluation on souhaite utiliser, il suffit, à la construction des liens dans le script de simulation, de spécifier l'attribut *separable* respectivement à 0, 1 ou 2. On va maintenant détailler un tout petit peu plus la manière dont l'évaluation est implémentée. Le principe général consiste à calculer, à la première exécution du script de simulation, les décompositions SVD des matrices de poids. Ce calcul n'est à faire qu'à la première exécution puisque les poids sont supposés ne pas évoluer pendant la simulation ; c'est un pré-requis de l'implémentation proposée qui permet d'optimiser le temps de calcul en ne calculant la décomposition SVD qu'une seule fois.

Comme nous l'avons vu dans l'annexe B, l'application de cette méthode se passe en deux étapes. Une première étape est commune à toutes les unités et une seconde est exécutée pour chaque unité. A l'exécution du script de simulation, lorsqu'un poids est défini à l'aide de l'objet de type `PROJECTION`, la décomposition SVD des poids est calculée et les vecteurs U, S, V obtenus sont stockés dans la carte qui contient les unités post-synaptiques. A l'évaluation d'un pas de temps, les cartes commencent par calculer une matrice temporaire, commune à toutes les unités et qui implique les activités des unités pré-synaptiques ainsi que les matrices S et V . Puis les cellules évaluent leur activité. Pour ce faire, chaque unité récupère le vecteur U de la décomposition SVD dans la carte qui la contient. Puisque les unités d'une carte peuvent avoir des afférences de plusieurs autres cartes, un indice est affecté à chaque décomposition SVD et transmis aux unités pour qu'elles puissent récupérer le bon vecteur de la transformation. L'unité utilise ainsi le vecteur V ainsi que la matrice temporaire calculée par la carte pour mettre à jour son activité. Lorsque l'activité d'une unité est mise à jour, il est nécessaire de répercuter cette mise à jour dans les matrices temporaires calculées au niveau des cartes. Dans le schéma d'évaluation utilisé pour les simulations qu'on présente dans ce manuscrit, les cartes sont évaluées séquentiellement et les unités sont évaluées dans un ordre aléatoire. Ainsi, lorsqu'une unité est mise à jour, il suffit de mettre à jour la matrice temporaire de la carte qui contient la cellule si et seulement si les unités de cette carte ont des projections latérales. Pour utiliser un schéma d'évaluation dans lequel les unités de tous le réseau seraient évaluées dans un ordre aléatoire (c'est à dire en se

passant de la séquentialité d'évaluation des cartes), il serait nécessaire de modifier la librairie.

Exemples

La figure C.5 illustre un script permettant d'utiliser la décomposition SVD pour évaluer les poids entre les cartes intitulées *input* et *focus*.

C.2.7 dana.image : Librairie de traitement d'image

Description

Le package `dana.image` propose un ensemble de méthodes pour réaliser des opérations sur des images et injecter le résultat dans les cartes DANA. Cette librairie repose sur les outils fournis avec la librairie `Mirage C++`¹⁸ qui permet de faire du traitement d'image. Elle réalise également les différentes opérations relatives à l'extraction des différents canaux permettant de construire la carte de saillance proposée par [Itti, 2000, Frintrop, 2005b]. Enfin elle fait le lien avec les réseaux de neurones DANA en projetant les résultats des opérations appliquées sur les images, sur les cartes de cellules DANA.

Nous ne sommes pas, à proprement parler, intéressés par le calcul de la carte de saillance mais plutôt par l'ensemble des calculs qui amènent jusqu'à la construction des cartes intermédiaires qui sont, dans l'implémentation standard de L. Itti, combinées pour former la carte de saillance. Le principe du calcul de ces cartes de caractéristiques (*feature maps*) reposent sur la détection, pour des attributs visuels donnés, des régions les plus contrastées.

Exemples

Les illustrations C.6 représentent l'image source ainsi que le résultat des différentes opérations permettant de construire :

- le canal d'opposition rouge-vert ;
- le canal d'opposition vert-rouge ;
- le canal d'opposition bleu-jaune ;
- le canal d'opposition jaune-bleu ;
- le canal d'intensité ;
- les canaux de Sobel pour les orientations 0°, 45°, 90°, 135°.

Le bas de la figure illustre l'état du réseau lorsque le résultat des filtres est utilisé pour définir l'activité des cellules du réseau.

¹⁸La librairie `Mirage C++` est en libre accès à l'adresse <http://www.metz.supelec.fr/metz/recherche/ersidp/>

```

import dana.core as core
import dana.projection as proj
import dana.svd as svd

model = core.Model()
net = core.Network ()
model.append(net)
width = 50
height = width

# Create the input map
Input = core.Map ( (width,height), (0,0) )
Input.append(core.Layer())
Input[0].fill(core.Unit)
Input.name = 'Input'
net.append(Input)

# Create the focus map
Focus = core.Map ( (width,height), (1,0) )
Focus.append(svd.Layer())
Focus[0].fill(svd.Unit)
Focus.name = 'Focus'

Focus.spec = cnft.Spec()
Focus.spec.tau = 10.0
Focus.spec.baseline = 0.0
Focus.spec.alpha = 8.0
Focus.spec.min_act = 0.0
Focus.spec.max_act = 1.0
Focus.spec.wp = 1
Focus.spec.wm = 1

net.append(Focus)

# Create input to focus connections
p1 = svd.projection()

p1.self_connect = True
p1.separable = 2
p1.distance = proj.distance.Euclidean (True)
p1.profile = proj.profile.Gaussian(0.5,0.05)
p1.density = proj.density.Full(1)
p1.shape = proj.shape.Disc(1)
p1.src = Input[0]
p1.dst = Focus[0]
p1.connect()

```

FIG. C.5 – Pour utiliser la décomposition SVD dans l'utilisation des poids, il suffit de définir les poids avec la classe PROJECTION en spécifiant l'attribut *separable* et d'utiliser les unités proposées par le package. Ces unités héritent des unités de la librairie `dana.cnft`

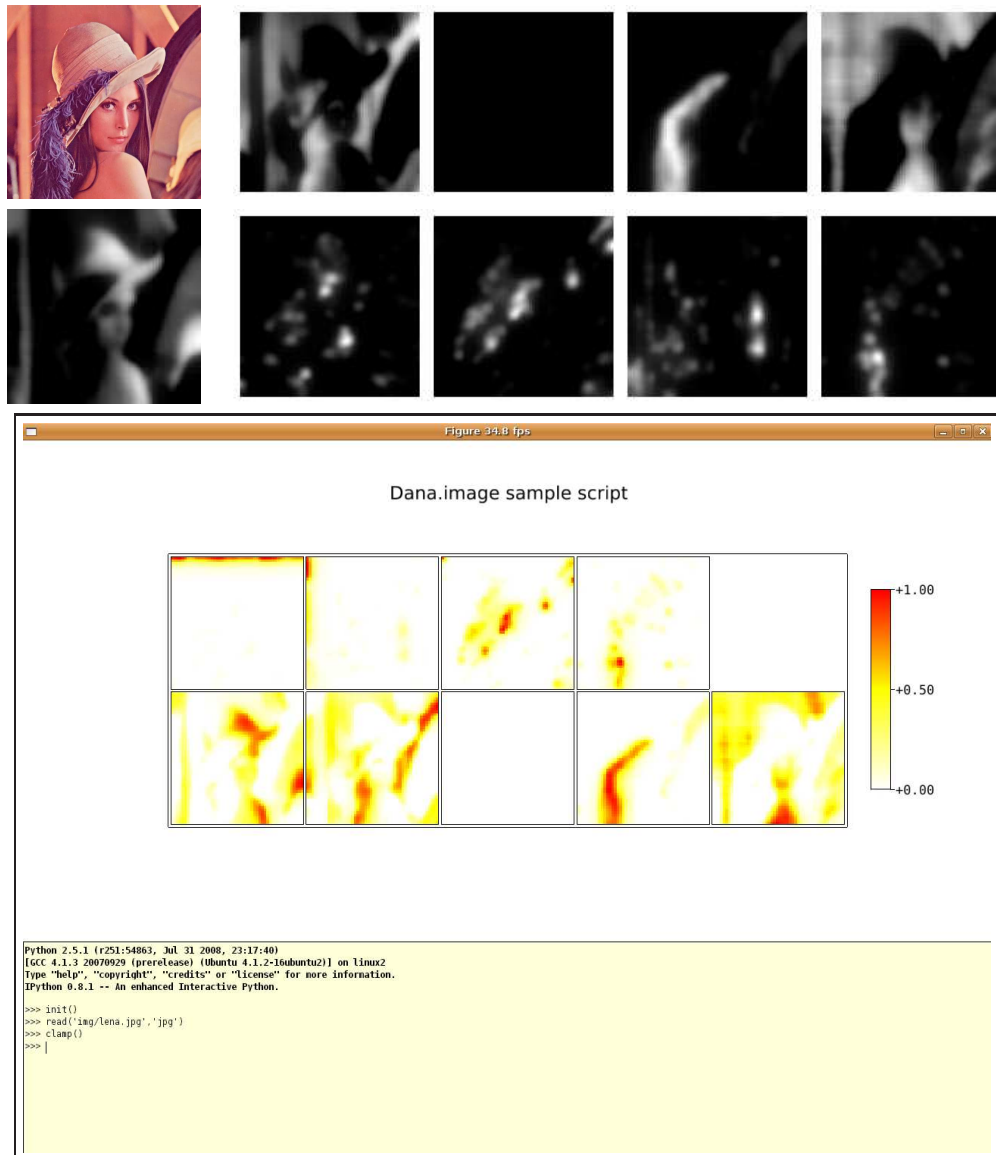


FIG. C.6 – Illustration du package `dana.image`. Les illustrations du haut représentent l'image source ainsi que les résultats des différents filtres : opposition rouge-vert, vert-rouge, bleu-jaune, jaune-bleu, intensité, filtre de sobel orienté à 0° , 45° , 90° , 135° . L'illustration du bas représente la fenêtre de simulation dans laquelle le résultat des filtres est utilisé pour définir l'activité des unités du réseau.

C.2.8 glpython.world : Un environnement virtuel

Description

La librairie `glpython.world` définit un environnement virtuel dans lequel un robot virtuel peut naviguer. On distingue principalement deux types d'entités dans l'environnement :

- les objets qui peuvent être des barres orientées colorées ou des disques colorés ou d'autres objets qu'on peut facilement ajouter à la librairie en surchargeant la classe `OBJECT` fournis par `glpython` ;
- le robot virtuel qui peut naviguer dans l'environnement et le percevoir au travers d'une caméra virtuelle.

La librairie propose ainsi différents objets de base qu'on peut introduire dans l'environnement et dont on peut modifier les propriétés (position ou attributs visuels comme la couleur, l'orientation ou la forme d'une barre). Nous ne détaillerons pas plus les objets, il est simplement intéressant de noter qu'il est très simple d'en introduire de nouveaux. La librairie définit également une classe `ROGER` qui représente un robot virtuel constitué d'un corps et d'une caméra. Le corps et la caméra peuvent être déplacés indépendamment en rotation mais sont liés en translation. Par ailleurs, il est possible de capturer l'image de la caméra, soit pour la faire apparaître dans l'environnement de simulation, soit pour l'utiliser comme entrée à la librairie `dana.saliency` qui alimentera ainsi le réseau avec une entrée visuelle.

La principale difficulté relative au déplacement de la caméra concerne la transformation entre les coordonnées de la caméra et l'espace neuronal utilisé dans le mécanisme pour représenter la commande motrice. La correspondance entre l'espace visuel de la caméra et l'espace neuronal utilisé pour représenter la commande motrice fait intervenir trois paramètres :

- la résolution de l'image capturée, notée *image_horiz_res*, *image_vert_res* ;
- l'ouverture (*aperture*) de la caméra, notée *aperture* qui définit l'étendue du champ visuel vertical ;
- la taille de la carte neuronale utilisée pour représenter l'information motrice, notée *width*, *height*.

Ces trois paramètres sont combinés pour obtenir des coefficients permettant de transformer linéairement une position dans la carte neuronale en une commande motrice pour la caméra dans l'environnement.

Pour ce faire, on commence par évaluer l'étendue du champ visuel de la caméra :

$$\begin{aligned} \text{vert_visual_field} &= \text{aperture} \\ \text{horiz_visual_field} &= \text{aperture} \frac{\text{image_horiz_res}}{\text{image_vert_res}} \end{aligned}$$

Les coefficients qu'on cherche à déterminer sont simplement le rapport entre les dimensions de la carte neuronale et l'étendue du champ visuel :

$$\begin{aligned} \text{horiz_visiomotor} &= \frac{\text{width}}{\text{horiz_visual_field}} \\ \text{vert_visiomotor} &= \frac{\text{height}}{\text{vert_visual_field}} \end{aligned}$$

Ainsi, lorsqu'une position $pan2do, tilt2do$ du centre de masse est décodée d'une carte neuronale représentant la commande motrice à exécuter, il suffit d'envoyer la commande de rotation de la caméra au robot avec les paramètres $\frac{pan2do}{horiz_visiomotor}, \frac{tilt2do}{vert_visiomotor}$

Exemples

L'illustration C.7 représente la fenêtre de simulation gpython dans laquelle apparaît une vision à la troisième personne de l'environnement et du robot, ainsi que ce que perçoit le robot virtuel.

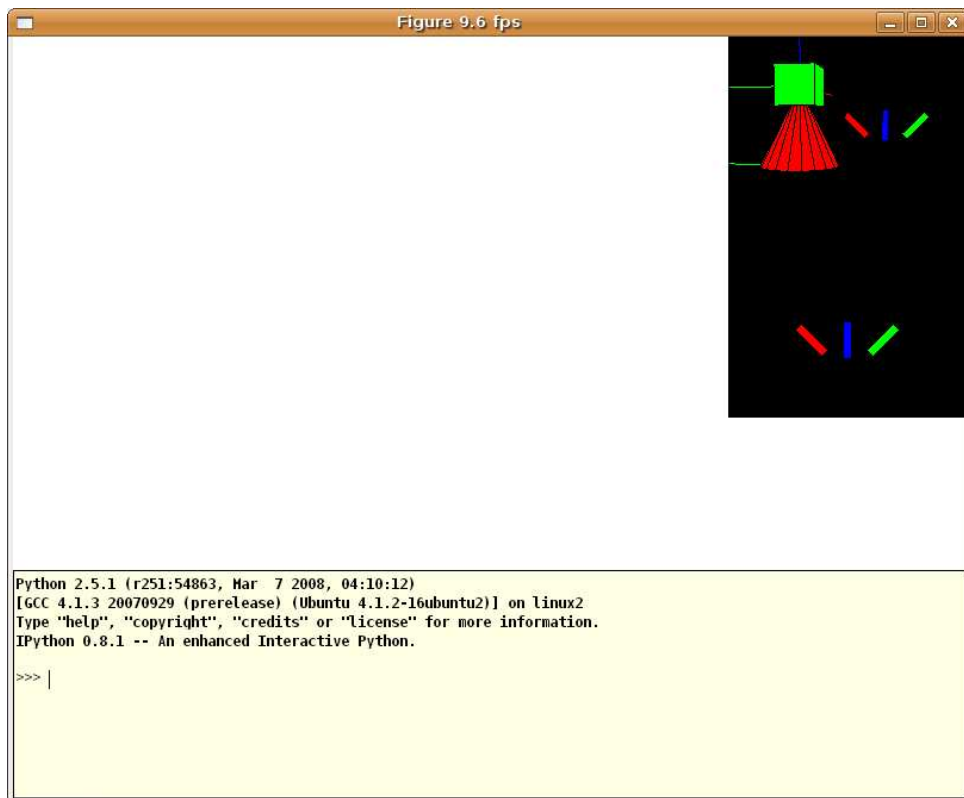


FIG. C.7 – Illustration de la librairie gpython_world

Annexe D

Codage de la proprioception : monotone ou explicite

D.1 Introduction

Une des informations utilisées dans le modèle que nous avons présenté dans la section 4.2 est la position de l’oeil. Chez les primates, la position de l’oeil est fournie par la proprioception qui indique l’état de contraction des muscles extraoculaires. Comme nous l’avons souligné, les mesures expérimentales indiquent que la proprioception est codée par des populations monotones (les cellules qui codent la position horizontale ou verticale de l’oeil ont une fréquence de décharge qui dépend de façon monotone de l’excentricité horizontale ou verticale de l’oeil). Or, dans le mécanisme proposé, nous utilisons un codage explicite. C’est à dire que la position de l’oeil est représentée par une gaussienne dans une carte 2D. Dans cette annexe, nous montrons qu’il est possible de se passer d’une représentation explicite de la position de l’oeil puisque cette position explicite peut être décodée linéairement d’une représentation monotone.

D.2 Étude empirique

D.2.1 Définition du problème

La question à laquelle nous cherchons à répondre est la suivante : est-il possible de décoder linéairement une représentation explicite 2D de la position de l’oeil, à partir de quatre représentations monotones 1D codant respectivement la position horizontale et la position verticale de l’oeil (avec deux populations pour chaque composante : une population monotone croissante et une population monotone décroissante).

Pour répondre à cette question, on mène une étude empirique d’apprentissage supervisé entre les couches d’entrées qui codent de façon monotone la composante horizontale ou verticale de la position de l’oeil et une couche de sortie qui code explicitement la position de l’oeil par une

gaussienne dans une carte 2D. L'architecture considérée est représentée sur la figure D.1

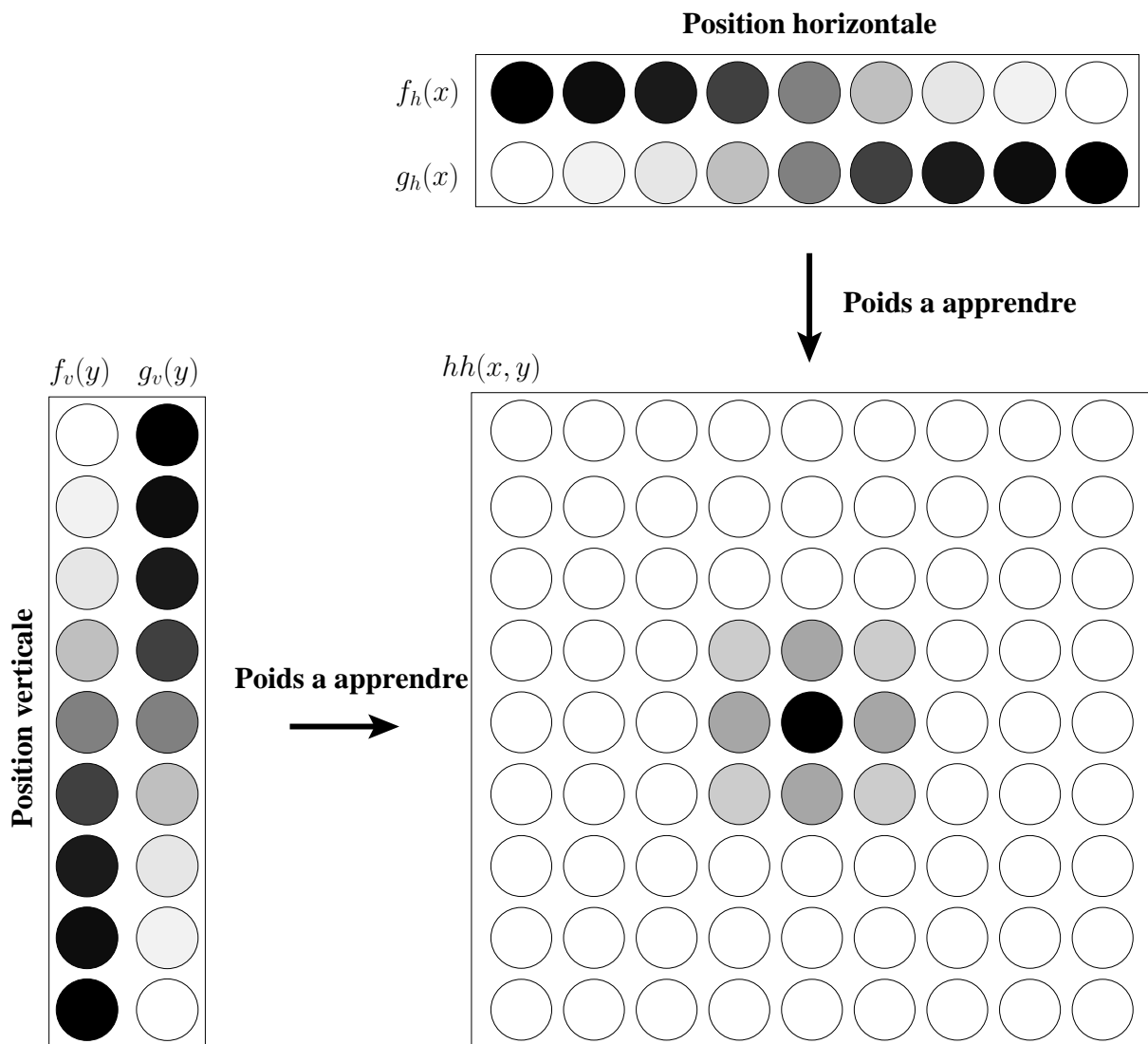


FIG. D.1 – Architecture du modèle appris par apprentissage supervisé

L'apprentissage supervisé est réalisé en définissant les entrées $f_h(x)$, $g_h(x)$, $f_v(y)$, $g_v(y)$ ainsi que la sortie désirée $hh_d(x, y)$ pour des positions x_0, y_0 choisies aléatoirement. Les activités de

ces différentes couches sont définies par :

$$\begin{aligned} f_h(x) &= \frac{1}{1 + e^{-\frac{x-x_0}{\sigma_0}}} \\ g_h(x) &= \frac{1}{1 + e^{-\frac{x_0-x}{\sigma_0}}} \\ f_v(y) &= \frac{1}{1 + e^{-\frac{y-y_0}{\sigma_0}}} \\ g_v(y) &= \frac{1}{1 + e^{-\frac{y_0-y}{\sigma_0}}} \\ hh_d(x, y) &= e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{\sigma_0^2}} \end{aligned}$$

Pour réaliser l'apprentissage, on connecte intégralement les couches d'entrées et la couche de sortie, les neurones de la couche de sortie ayant pour fonction de transfert une sigmoïde. Les activités des neurones de la couche de sortie sont ainsi définies par :

$$\begin{aligned} hh(x, y) &= f\left(\sum_{x'} w_{0,h}(x, x') f_h(x') + w_{1,h}(x, x') g_h(x') + \sum_{y'} w_{0,v}(y, y') f_v(y') + w_{1,v}(y, y') g_v(y')\right) \\ f(x) &= \frac{1}{1 + \exp(-x)} \end{aligned} \quad (\text{D.1})$$

Les poids des projections sont appris par apprentissage supervisé en utilisant la règle de rétropropagation du gradient :

$$\Delta w = \eta \times pre \times \delta \quad (\text{D.2})$$

$$\delta = hh(x, y)(1 - hh(x, y))(hh_d(x, y) - hh(x, y)) \quad (\text{D.3})$$

Le terme $hh(x, y)(1 - hh(x, y))$ étant obtenu en calculant la dérivée de la fonction de transfert qui est une sigmoïde et qui a une dérivée facile à exprimer en fonction de l'activité du neurone post-synaptique.

D.2.2 Simulation et résultats

Les poids sont initialisés par des valeurs aléatoires de petite amplitude. Le réseau, avec une couche de sortie de 30×30 cellules, est entraîné sur 400 000 présentations de positions horizontales et verticales choisies aléatoirement.

Une fois l'apprentissage terminé, on constate (ça n'est pas illustré) que les poids d'une colonne de la couche de sortie avec les populations f_h, g_h sont constants et les poids entre une ligne de la couche de sortie et les populations f_v, g_v sont constants. On propose alors de moyenner sur chaque colonne les poids entre f_h, g_h et la couche de sortie ainsi que de moyenner sur chaque ligne les poids entre f_v, g_v et la couche de sortie. La figure D.2 illustrent les quatre jeux de poids ainsi obtenus.

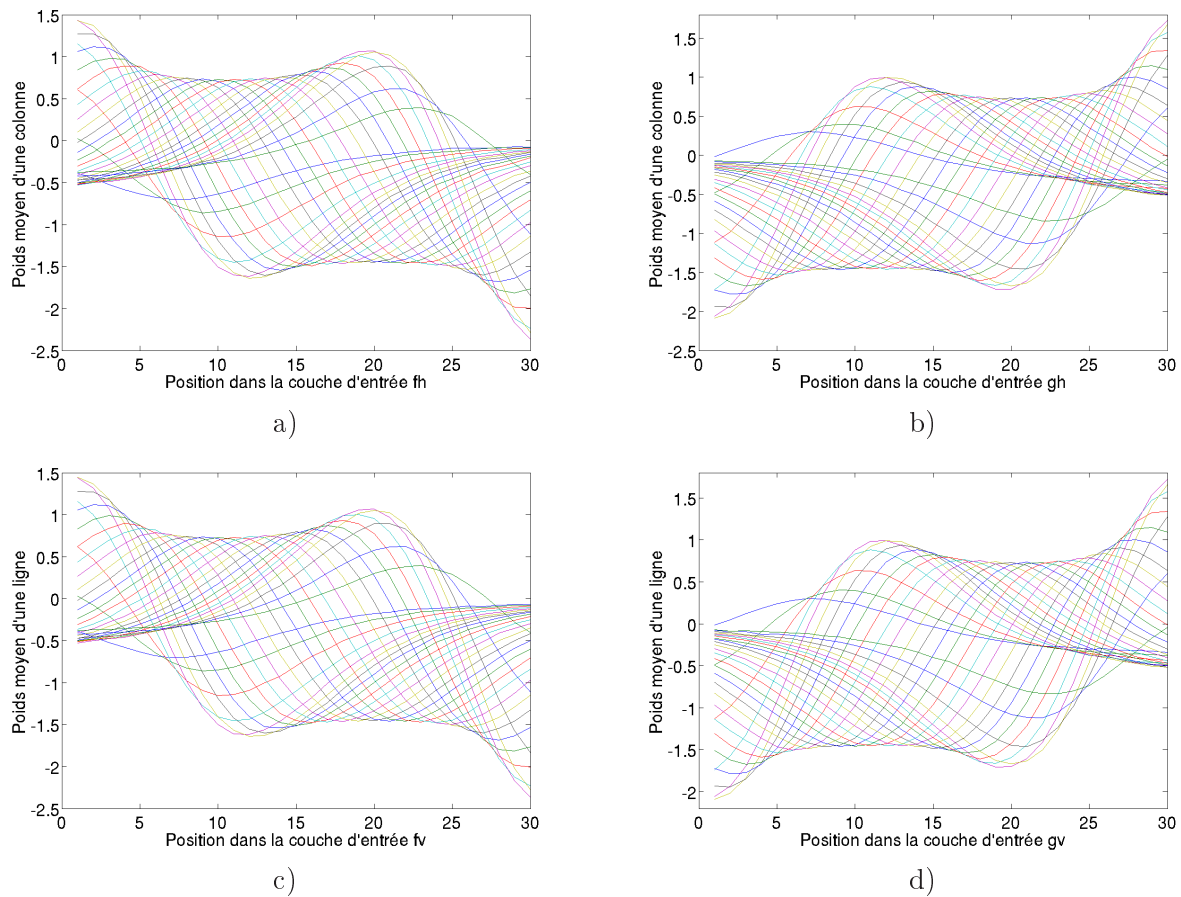


FIG. D.2 – Chaque courbe représente le poids entre un neurone de la couche cachée et un neurone d'une des couches d'entrée. Les projections entre les populations f_h et g_h et la couche cachée sont les mêmes sur une colonne de la couche cachée. Les projections entre les populations f_v et g_v et la couche cachée sont les mêmes sur une ligne de la couche cachée. a) Projections entre les neurones d'une ligne de la couche cachée et la population f_h . b) Projections entre les neurones d'une ligne de la couche cachée et la population g_h c) Projections entre les neurones d'une colonne de la couche cachée et la population f_v d) Projections entre les neurones d'une colonne de la couche cachée et la population g_v

A partir des figures D.2 on observe une certaine régularité des poids. On observe en effet qu'ils ont globalement la forme de la dérivée d'une gaussienne dont le signe dépend de la population présynaptique considérée. Il y a cependant deux paramètres qui distinguent les poids obtenus de la dérivée d'une gaussienne : un offset et un facteur d'amplitude, qui sont deux facteurs d'autant plus importants qu'on s'approche des bords des populations. On pourrait en effet continuer l'analyse des poids et constater que les poids peuvent être approximés par la dérivée d'une gaussienne à laquelle s'ajoute un offset qui dépend de la distance par rapport au centre de la population. Il y a également une modulation en amplitude qui dépend de la distance au centre de la population. Néanmoins, nous arrêtons l'analyse des poids en proposant qu'ils ont la forme de la dérivée d'une gaussienne sous condition qu'on soit suffisamment loin des bords. L'apprentissage a l'avantage de gérer les effets de bords qui correspondent au fait que lorsqu'on se trouve sur les extrêmes des populations, l'information est plus ambiguë. Dans le prochain paragraphe, on propose une étude analytique confirmant la forme globale des poids en supposant les populations infinies et continues.

D.3 Résultat analytique

Nous proposons dans ce paragraphe un résultat analytique en 1D, dans le cas continu, obtenu en collaboration avec Thierry Vieville. Ces restrictions permettent de proposer un résultat analytique qui est certes réducteur (puisque dans le paragraphe précédent, on travaille dans un espace 2D, discret et borné) mais qui confirme en partie les intuitions sur la forme des poids que nous avons formulées à la fin du paragraphe précédent. L'extension au cas 2D, continu est brièvement abordé à la fin du paragraphe.

Étant données les deux fonctions f_h, g_h qui codent monotonement la position horizontale de l'oeil :

$$f_h(x) = \frac{1}{1 + \exp\left(-\frac{x-x_0}{\sigma_0}\right)} \quad (\text{D.4})$$

$$g_h(x) = \frac{1}{1 + \exp\left(-\frac{x_0-x}{\sigma_0}\right)} \quad (\text{D.5})$$

et le noyau $K(u, c, \sigma)$:

$$K(u, c_0, \sigma_0) = \frac{c_0 u}{\sigma_0} e^{-\frac{u^2}{\sigma_0^2}} \quad (\text{D.6})$$

Alors la fonction $hh(x)$ définie par :

$$hh(x) = \int_{-\infty}^{\infty} (K(u, -c_0, \sigma_0) f_h(x-u) + K(u, c_0, \sigma_0) g_h(x-u)) du \quad (\text{D.7})$$

$$= \frac{c_0}{\sigma_0} \int_{-\infty}^{\infty} \frac{(e^{\frac{u+x_0}{\sigma_0}} - e^{\frac{x}{\sigma_0}}) u}{(e^{\frac{u+x_0}{\sigma_0}} + e^{\frac{x}{\sigma_0}}) e^{\frac{u^2}{\sigma_0^2}}} \quad (\text{D.8})$$

est *très proche* d'une gaussienne définie par :

$$ce^{-\frac{(x-x_0)^2}{\sigma^2}} \quad (\text{D.9})$$

avec $c = 0.4c_0\sigma_0$ et $\sigma = 2.32\sigma_0$

On peut facilement étendre le résultat précédent dans le cas 2D, continu, en considérant la fonction suivante :

$$hhh(x, y) = f(hh(x) + hh(y) - hh(0)) \quad (\text{D.10})$$

$$f(x) = \frac{x + |x|}{2} \quad (\text{D.11})$$

La fonction f est une rectification. Le résultat indique que dans le cas 2D, on peut également décoder “linéairement” (dans le sens où les neurones intègrent linéairement leurs entrées) une représentation explicite 2D à partir de quatre représentations monotones (respectivement deux représentations monotones pour la position horizontale et deux représentations monotones pour la position verticale), sous condition d’utiliser une non-linéarité dans la fonction de transfert qui est une simple rectification.

Annexe E

Publications

Ces travaux de thèse ont donné lieu aux publications suivantes :

Revue internationale

- J. Fix, N. Rougier et F. Alexandre (2007). "From physiological principles to computational models of the cortex", *Journal of Physiology - Paris*, 101(1-3), pp. 32-39,
- N. Rougier, J. Fix (2008). "DANA, Distributed Asynchronous Numerical Adaptive computing framework", *Frontiers in Neuroinformatics : Python in Neuroscience*, (en cours de soumission).

Chapitre de livre

- J. Fix, J. Vitay et N. Rougier (2007). "A Distributed Computational Model of Spatial Memory Anticipation During a Visual Search Task", in *Anticipatory Behavior in Adaptive Learning Systems*, M.V. Butz, O. Sigaud, G. Pezzulo, G. Baldassarre (Editeurs), pages 170-188, Springer-Verlag.

Conférences internationales avec comité de lecture, communication orale

- J. Fix, J. Vitay et N. Rougier (2006). "A Computational Model of Spatial Memory Anticipation during Visual Search", dans *Anticipatory Behavior in Adaptive Learning Systems*, ABIALS 2006, Rome (Italie),
- J. Fix, N. Rougier et F. Alexandre (2007). "A top-down attentional system scanning multiple targets with saccades", dans *From Computational Cognitive Neuroscience to Computer Vision*, CCNCV 2007, Bielefeld (Allemagne),
- J. Fix (2007). "A computational approach to the control of voluntary saccadic eye movements" dans *1st International Conference on Cognitive Neurodynamics*, ICCN 2007, Shanghai (Chine),
- J. Fix, N. Rougier, F. Alexandre (2008). "A computational approach to the covert and overt deployment of spatial visual attention", dans *2ième Conférence Francophone de Neurosciences Computationnelles*, Neurocomp 2008, Marseille (France).

Conférences internationales avec comité de lecture, poster

- J. Fix (2006). "Principes de transformations sensorimotrices à l'aide de neurones Sigma-Pi", dans *1ère Conférence Francophone de Neurosciences Computationnelles*, Neurocomp 2006, Pont-à-Mousson (France),
- J. Fix, N. Rougier, F. Alexandre (2008). "A computational approach to the covert and overt deployment of spatial visual attention", dans *Munich Symposium of Visual Search and Selective Attention*, MVSSA 2008, Fribourg (Suisse),
- F. Alexandre, J. Fix, A. Hutt, N. Rougier, T. Viéville (2008). "On practical neural field parameters adjustment", dans *2ième Conférence Francophone de Neurosciences Computationnelles*, Neurocomp 2008, Marseille (France).

Le texte des deux publications les plus importantes suit.

E.1 A Distributed Computational Model of Spatial Memory Anticipation During a Visual Search Task

Jeremy Fix, Julien Vitay et Nicolas Rougier

Anticipatory Behavior in Adaptive Learning Systems, p. 170–188, 2007

Abstract

Some visual search tasks require to memorize the location of stimuli that have been previously focused. Considerations about the eye movements raise the question of how we are able to maintain a coherent memory, despite the frequent drastic changes in the perception. In this article, we present a computational model that is able to anticipate the consequences of eye movements on visual perception in order to update a spatial working memory.

E.1.1 Introduction

In the most general framework of behavior, the notion of anticipation is intimately linked with the possibility to predict the consequences and the outcomes of a given action. If we consider that any action is goal-motivated, then an action is carried out in the first place because it is anticipated that this action could lead to a situation where it is more straightforward to reach the goal. In this framework, anticipation can be viewed as a prediction of the future and is tightly linked to the notion of goal-directed behavior. However, there also exists more structural reasons why anticipation is necessary.

For example, when dealing with both accurate and very fast movements like catching a ball or scanning a visual scene, brain representations should be updated very quickly (even in advance in some cases) in accordance with the task that is carried out. The problem in this context is that the time scale required for carrying out such tasks may be dramatically smaller than the time scale of a single neuron. Moreover, those neurons are also in interaction with other neurons in the network and the resulting dynamic may be even slower. One solution to cope with this problem is to use a forward predictive model that is able to anticipate the consequences and outcomes of a motor action. The resulting dynamic at the level of the model is then faster than the dynamic of its components.

Let us consider the ability to anticipate changes in the visual information resulting from an eye saccade. This anticipation is known to be largely based on unconscious mechanisms that provide us with a feeling of stability while the whole retina is submerged by different information at each saccade : producing a saccade results in a complete change in the visual perception of the outer world. If a system is unable to anticipate its own saccadic movements, it cannot pretend to obtain a coherent view of the world : each image would be totally uncorrelated from the others. One stimulus being at one location before a saccade could not be easily identified as being the same stimulus at another location after the saccade. Consequently, the saccadic eye movements should be anticipated in order to keep the coherence of the scene and to be able to track down

interesting targets. A number of works already addressed the specific problem of visual search of a target among a set of distractors. However, most of the resulting models do not deal with the problem of saccadic eye movements that produce drastic changes in the available visual information.

Using neural fields introduced by [Amari, 1977] for the one dimensional case and later extended to higher dimensions by [Taylor, 1999], we would like to address in this paper the specific problem of anticipation during visual search using a purely distributed and numerical neural substrate. After briefly reviewing literature related to visual search in the first section, we introduce a very simple visual experiment that helps illustrating the underlying mechanisms of the model that is detailed in the same section.

E.1.2 Visual search

Visual search is a cognitive task that most generally involves an active scan of a visual scene for finding one or several given targets among distractors. It is deeply anchored in most animal behaviors, from a predator looking for a prey in the environment, to the prey looking for a safe place to avoid being seen by the predator. Psychological experiments may be less ecological and may propose for example to find a given letter among an array of other letters, measuring the efficiency of the visual search in terms of reaction time (the average time to find the target given the experimental paradigm). In the early eighties, [Treisman et Gelade, 1980] suggested that the brain actually extracts some basic features from the visual field in order to perform the search. Among these basic features that have been recently reviewed by [Wolfe, 1998], one can find features such as color, shape, motion or curvature. Finding a target is then equivalent to finding the conjunction of features (that may be unique) that best describes the target. In this sense, [Treisman et Gelade, 1980] distinguished two main paradigms (a more tempered point of view can be found in [Duncan et Humphreys, 1989]).

Feature search refers to a search where the target differs from distractors against exactly one feature.

Conjunction search refers to a search where the target differs from distractors against two or more features.

What characterizes best the feature search is a constant search time that does not depend on the number of distractors. The target is sufficiently different from the distractors to pop out. However, in the case of conjunction search, the mean time needed to find the target is roughly proportional to the number of distractors that share at least one feature with the target (cf. Fig. E.1). These observations lead to the question of how a visual stimulus could be represented in the brain. The explanation given by [Treisman et Gelade, 1980], the *Feature-Integration Theory*, proposed that elementary features are processed in separated feature maps. Competition inside one map would lead to Feature search, based on the idea that the item differing the most from its background would win the competition and be represented. For targets differing from distractors by more than two features, there can not be any global competition. This would mean that finding the target requires to successively scan every potential candidate until the correct target

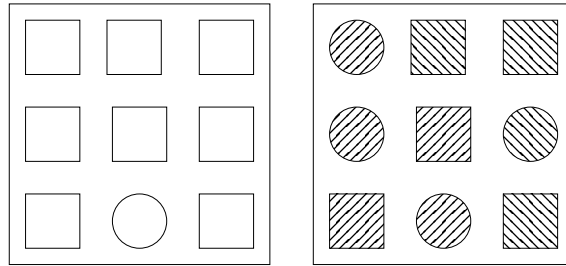


FIG. E.1 – Feature search can be performed very quickly as illustrated on the left part of the figure ; the disc shape literally pops out from the scene. However, as illustrated on the right part of the figure, if the stimuli share at least two features, the pop out effect is suppressed. Hence, finding the disc shape with the stripes going from up-left to down-right requires an active scan of the visual scene.

is found. This explains the dependence of the search time on the number of similar distractors in conjunction search tasks.

The main prediction of this theory is that processing visual inputs is not a global feed-forward processing, but more an iterative and sequential process on sensory representations. We describe below the strategies used by the brain to achieve this sequential search, by putting emphasis on saccadic eye movements and visual attention. The scope of this article is therefore to model the cognitive structures involved in this sequential processing of visual objects, and not the visual processing of features by itself.

Saccadic eye movements

The eye movements may have different behavioral goals, leading to five different categories of movements : saccades, vestibulo-ocular reflex, optokinetic reflex, smooth-pursuit and vergence. However, in this article we will only focus on saccades (for a detailed study of eye movements, see [Leigh et Zee, 1999], [Carpenter, 1988]).

Saccades are fast and frequent eye movements that move quickly the eye from the current point of gaze to a new location in order to center a visual stimulus on the fovea, a small area on the retina where the resolution is at its highest. The velocity of the eyes depends on the amplitude of the movement and can reach up to 700 degrees per second at a frequency of 3 Hz. The question we would like to address is how the brain may give the illusion of a stable visual space while the visual perception is drastically modified every 300 ms.

While the debate to decide whether or not the brain is blind during a saccade has not been settled (see [Li et Matin, 1997, Burr, 2004, Kleiser et al., 2004, Ross et al., 2001] for the notion of saccadic suppression and [O'Regan et Noë, 2001] for a discussion about the necessity of a saccadic suppression mechanism), the coherence between the perception before and after a saccade cannot be established accurately solely based on perception. One solution is to consider that the brain may use an efferent copy of the voluntary eye movement to remap the representation it has

built of the visual world. Several studies shed light on pre-saccadic activities in areas such as V4 and LIP where the locations of relevant stimuli are supposed to be represented. In [Moore et al., 1998], the authors suggest that “the presaccadic enhancement exhibited by V4 neurons [...] provides a mechanism by which a clear perception of the saccade goal can be maintained during the execution of the saccade, perhaps for the purpose of establishing continuity across eye movements”. In [Merriam et Colby, 2005], the authors review evidences that LIP neurons, whose receptive field will land on a previously stimulated screen location after a saccade, are excited even if the stimulus disappears during the saccade. In a recent study, [Sommer et Wurtz, 2006] showed neurons in FEF that receive projections from the superior colliculus that could explain the origin of a corollary discharge signal, responsible for the presaccadic activity exhibited by these neurons.

Visual attention

Focusing on a given stimulus of the visual scene is a particular aspect of the more general concept of attention that has been defined as the capacity to concentrate cognitive resources on a restricted subset of sensory information ([James, 1890]). In this context of visual attention, only a small subset of the retinal information is available at any given time to elaborate motor plans or cognitive reasoning (cf. *change blindness* experiments presented in [O’Regan et Noë, 2001], [Simons, 2000]). A visual scene is not processed as a whole but rather processed by successively focusing on interesting parts of it, possibly involving eye movements, but this is not necessary. The selection of a target for an eye movement is then closely related to the notion of spatial attention ([Moore et Fallah, 2001]) that is classically divided into two types : **overt attention** which involves a saccade to center a stimulus on the fovea and **covert attention** in which no eye movement is triggered. These two types of spatial attention were first supposed to be independent ([Posner et Petersen, 1990]) but recent studies such as the premotor theory of attention proposed in [Rizzolatti et al., 1987] (see also [Chelazzi et al., 1993], [Kowler et al., 1995], [Craighero et al., 1999]) consider that covert and overt attention rely on the same neural structures but the movement is inhibited in covert attention. A more general discussion about the covert and overt stages of action can be found in [Jeannerod, 2001].

The deployment of attention on a specific part of the visual information can be the consequence of two phenomenas. Firstly it can rely on the saliency of a stimulus, compared to its surrounding (for example a sudden strong flash light) ; this is known as bottom-up attention. Secondly, it can also depend on the task in which the subject is involved, which could need to enhance some parts of the perception (for example, imagine that you have to find an orange among apples and bananas, the color information could be a good criteria to find the target rapidly).

In [Moran et Desimone, 1985], the authors shed light on the neural correlates of attention on the response of neurons in the visual and temporal cortices. If we consider a specific neuron tuned to a given orientation in his receptive field, one can distinguish several cases :

- the response of the neuron is high when an oriented bar with the preferred orientation (called good stimulus) is presented in its receptive field
- the response of the neuron is low when an oriented bar with an orientation different from

the preferred one (called bad stimulus) is presented in its receptive field

- the response is between the two precedings when both a good and bad stimulus is presented

When a monkey is involved in a task that requires to select one of the two stimuli, for example the good one, the response of the neuron is enhanced. The study of this suppressive interaction phenomena was extended by further authors ([Luck et al., 1997], [Reynolds et Desimone, 1999], [Treue et Maunsell, 1996]).

As we will see in section E.1.3.0, we do not deal with how the salience of the visual stimuli is computed, whether or not it is a bottom-up or top-down processing. The main points are that for each location in the visual space, we are able to compute its behavioral relevance, and that considering eye movements necessarily implies dealing with overt attention.

Computational models

Over the past few years, several attempts at modeling visual attention have been engaged ([Koch et Ullman, 1985], [Tsotsos et al., 1995], [Wolfe, 2000], [Itti et Koch, 2001], [Hamker, 2004]). The basic idea behind most of these models is to find a way to select interesting locations in the visual space given their behavioral relevance and whether or not they have already been focused. The two central notions in this context have been proposed by [Koch et Ullman, 1985] and [Posner et Cohen, 1984] :

- saliency map
- inhibition of return (IOR).

The saliency map is a single spatial map, in retinotopic coordinates, where all the available visual information converge in order to obtain a unified representation of stimuli, according to their behavioral relevances. A winner-take-all algorithm can be easily used to find which stimulus is the most salient within the visual scene, that is identified as the locus of attention. However, in order to be able to go to the next stimuli, it is important to bias the winner-take-all algorithm in such a way that it prevents going backward to an already focused stimulus. The goal of the inhibition of return mechanism is precisely to feed the saliency map with such a bias. The idea is to have another neural map that records focused stimuli and inhibits the corresponding locations in the saliency map. Since an already focused stimulus is actively inhibited by this map, it cannot pretend to win the winner-take-all competition, even if it is the most salient.

The existence of a single saliency map is still not proved. In [Hamker, 2004] the author proposes a more distributed representation of these relevances, making a clear anatomical distinction between the processing of the visual attributes of an object and its spatial position (according to the What and Where pathways hypothesized by [Ungerleider et Mishkin, 1982], see also [Goodale et Milner, 1992]). In this model, spatial competition occurs in a motor map instead of a perceptive one. It exhibits good performances regarding visual search task in natural scene, but is restricted to covert attention. In most of the previously proposed models, the authors do not take into account eye movements and the visual scene is supposed to remain stable : scanning is done without any saccade. During the rest of this article, we will keep the saliency map hypothesis, even if controverted, in order to illustrate the anticipatory mechanism.

E.1.3 A model of visual search with overt attention

The goal of our model is to show the basic mechanisms necessary to achieve sequential search in a visual scene using both overt and covert attention. Using a saliency map, we need to compute the location of the most interesting stimulus that will be processed to achieve recognition. This focus of attention on a stimulus has to be displaced in two situations : in covert attention, this focus has to be dynamically inhibited to represent another stimulus. There is therefore a need for an inhibition-of-return mechanism than can inhibit the current focus of attention. Moreover, we have to memorize the locations of previously attended stimuli, by the mean of a dynamic spatial working memory.

The second situation to be considered is when eye movements can center the stimulus that is attended to. The spatial working memory has to be updated by the eye movement so that its state corresponds to the post-saccadic locations of memorized stimuli. This is where an anticipatory mechanism is mandatory.

To describe these mechanisms, we first present an experimental setup for which previous computational models would fail to achieve efficient sequential search. We then present the architecture of our model and report simulated results.

Experiment

In order to accurately evaluate the model, we setup a simple experimental framework in which some identical stimuli are drawn on a blackboard and are observed by a camera. The task is to successively focus (i.e. center) each one of the stimuli without focusing twice on any of them. We estimate the performance of the model in terms of how many times a stimulus has been focused. Hence, the point is not to analyze the strategy of deciding which stimulus has to be focused next (see [Findlay et Brown, 2006a, Findlay et Brown, 2006b] for details on this matter). In the context of the proposed model, the strategy is simply to go from the most salient stimulus to the least salient one, and to randomly pick one stimulus if the remaining ones are equally salient.

Figure E.2 illustrates an experiment composed of four identical stimuli where the visual scan path has been materialized. The effect of making a saccade from one stimulus to another is shown and underlines the difficulty (for a computational model) of identifying a stimulus before and after a saccade. Each one of the stimuli being identical to the others, it is impossible to perform an identification based solely on features. The only criteria that can be used is the spatial location of the stimuli.

Model

The model is based on three distinct mechanisms (cf. Fig. E.3 for a schematic view of the model). The first one is a competition mechanism that involves potential targets represented in a saliency map that were previously computed according to visual input. Second, to be able to focus

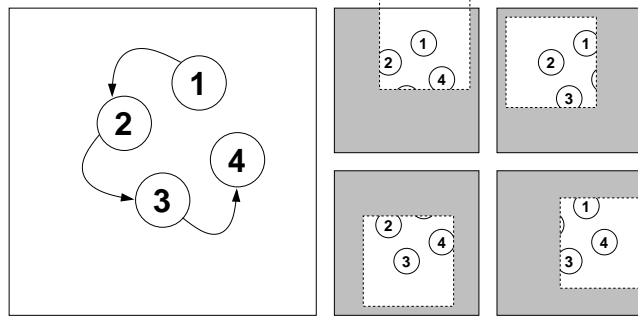


FIG. E.2 – When scanning a visual scene, going for example from stimulus 1 to stimulus 4, as illustrated on the left of the figure, the image received on the retina is radically changed when each stimulus is centered on the retina, as illustrated on the right of the figure. The difficulty in this situation is to be able to remember which stimuli have already been centered in order to center another one. The figures on the stimuli are shown only for explanation purpose and do not appear on the screen ; all the stimuli are identical.

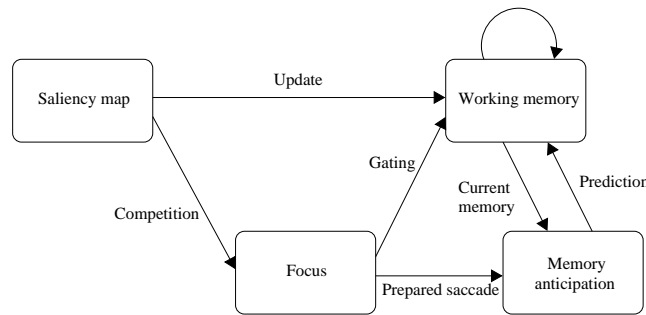


FIG. E.3 – Schematic view of the architecture of the model. The image captured by the camera is filtered and represented in the saliency map. This information feeds two pathways : one to the memory and one to the focus map. A competition in the focus map leads to the most salient location that is the target for the next saccade. The anticipation circuit predicts the future state of the memory with its current content and the programmed saccade.

only once on each stimulus, the locations of the scanned targets are stored in a memory map using retinotopic coordinates. Finally, since we are considering overt attention, the model is required to produce a camera movement, centering the target on the fovea, used to update the working memory. This third mechanism works in conjunction with two inputs : current memory and parameters of the next saccade. This allows the model to compute quite accurately a prediction of the future state of the visual space, restricted to the targets that have already been memorized.

The model is based on the computational paradigm of two dimensional discrete neural fields (the mathematical basis of this paradigm can be found in [Amari, 1977] for the one dimensional case, extended to a two dimensional study in [Taylor, 1999]). The model consists of five $n \times n$ maps of units, characterized by their position, denoted $\mathbf{x} \in [1..n]^2$ and their activity as a function of their position and time, denoted $u(\mathbf{x}, t)$. The basic dynamical equation that follows the activity of a unit at position \mathbf{x} , depends on its input $I(\mathbf{x}, t)$. Equation (E.1) is the equation proposed

in [Amari, 1977], discretized in space.

$$\tau \cdot \frac{\partial u(\mathbf{x}, t)}{\partial t} = -u(\mathbf{x}, t) + \text{baseline} + \frac{1}{\alpha} I(\mathbf{x}, t) \quad (\text{E.1})$$

We distinguish two kinds of units. The first are sigma units that compute their input as a weighted sum of the activity of afferent neurons, where afferent neurons are defined as neurons in other maps. We also consider lateral connections that involve units in the same map. If we denote w_{aff} the weighting function for the afferent connections and w_{lat} the weighting function for the lateral connections, the input $I(\mathbf{x}, t)$ of a unit \mathbf{x} at time t can be written :

$$I(\mathbf{x}, t) = \sum_{aff} w_{aff} u_{aff}(t) + \sum_{lat} w_{lat} u_{lat}(t) \quad (\text{E.2})$$

where (E.3) define the lateral and afferent weighting functions as a Gaussian and difference of Gaussians.

$$\begin{aligned} w_{aff}(\mathbf{x}, \mathbf{y}) &= A \cdot e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{a^2}} \text{ with } A, a \in \mathfrak{R}^{*+}, \mathbf{x}, \mathbf{y} \in [1..n]^2 \\ w_{lat}(\mathbf{x}, \mathbf{y}) &= B \cdot e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{b^2}} - C \cdot e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{c^2}} \text{ with } B, C, b, c \in \mathfrak{R}^{*+}, \mathbf{x}, \mathbf{y} \in [1..n]^2 \end{aligned} \quad (\text{E.3})$$

The second kind of units we consider are sigma-pi units ([Rumelhart et al., 1987]) that compute their input as a sum of product of the activity of afferent neurons. We also consider the lateral connection term so that the input of a unit \mathbf{x} at time t can be written :

$$I(\mathbf{x}, t) = \sum_{i \in I} w_{aff_i} \prod_{j \in E_i} u_{aff_j}(t) + \sum_{lat} w_{lat} u_{lat}(t). \quad (\text{E.4})$$

All the parameters of the previous equations used in the simulation are summarized in the appendix.

We now describe how the different maps interact. Since the scope of this article is the anticipation mechanism, the description of the saliency map, the focus map and the working memory will not be accurate ; a more detailed explanation, with the appropriate dynamical equations, can be found in [Vitay et Rougier, 2005].

Saliency map : The saliency map, also referred to as INPUT in the following, is computed by convolving the image captured with the camera of a robot used for the simulation with gaussian filters. The stimuli we use are easily discriminable from the background on the basis of the color information. This computation leads to a representation of the visual stimuli with gaussian patterns of activity in a single saliency map. We do not deal with how this saliency map is computed, whether or not it is due to bottom-up or top-down attention ; we only consider that we are able to compute a spatial map, in retinotopic coordinates, that represents the behavioral relevance of each location in the visual space. We point out again that this is one of our working hypothesis, detailed in section E.1.2.0.

Focus : The units in the FOCUS map have direct excitatory feedforward inputs from the saliency map. The lateral connections are locally excitatory and widely inhibitory so that a

competition between the units within the map leads to the emergence of only one stimulus in the focus map. This mechanism is not just a dynamical *winner-take-all* algorithm because the winning stimulus will still be represented in this map, even if the other stimuli in the visual scene become more salient through time than it : it has to be explicitly inhibited. This focused stimulus is considered as the next target to focus on and the movement to perform to center it on the fovea is decoded from this map. This map then codes the parameters of the next saccade to make.

Working memory : Once a stimulus has appeared within the focus map and because it is also present in the saliency map at the same location, it emerges within the working memory. Both the excitations from the focus map and the saliency map (at a same location) are necessary for the emergence of a stimulus in the working memory area. If the focused stimulus changes, it will not be present anymore in the focus map such that an additional mechanism is needed to maintain it in the memory. It is not shown on the schematic illustration (E.3) but the memory consists in two maps WM and THAL_WM that share excitatory connections in the two ways : the first map excites the second and the second excites the first, weighted so that the excitation is limited in space.

Memory anticipation : The memory anticipation mechanism aims at predicting what should be the state of the working memory, after an eye movement needed to center the stimulus in the focus map, before the movement is triggered. This map, filled with sigma-pi units, has two inputs : units of the focus map and units of the working memory. If we denote $wm(\mathbf{x},t)$ the activity of the unit \mathbf{x} of the working memory at time t , and $f(\mathbf{x},t)$ the activity of the unit \mathbf{x} of the focus map at time t , we define the input $I(\mathbf{x},t)$ of the unit \mathbf{x} in the anticipation map as :

$$I(\mathbf{x}, t) = w_{sigma-pi} \sum_{\mathbf{y} \in \mathbb{R}^2} wm(\mathbf{y}, t) f(\mathbf{y} - \mathbf{x}, t) + \sum_{aff} w_{aff} u_{aff}(t) \quad (\text{E.5})$$

The input of each unit in the anticipation map is computed as a convolution product of the working memory and the focus map, centered on its coordinates. To make (E.5) clearer, the condition of the sum is weaker than the one that should be used : since the input maps are discrete sets of units, the two vectors \mathbf{y} and $\mathbf{y}-\mathbf{x}$ mustn't exceed the size of the maps. The equation (E.5) should also take into account that the position *eye centered* is represented by a bell-shaped pattern of activity centered in the focus map, so that an offset should be included in the first sum when determining which unit of the focus map multiplies $wm(\mathbf{y}, t)$

From (E.1) and (E.5), the activity of the units in the anticipation map, without lateral connections, satisfies (E.6).

$$\tau \cdot \frac{\partial u(\mathbf{x}, t)}{\partial t} = -u(\mathbf{x}, t) + baseline + w_{sigma-pi} \sum_{\mathbf{y} \in \mathbb{R}^2} wm(\mathbf{y}, t) f(\mathbf{y} - \mathbf{x}, t) \quad (\text{E.6})$$

Then, the shape of activity in the anticipation map converges to the convolution product of the working memory and the focus map. Since the activity in the focus map has a gaussian shape and the working memory can be written as a sum of gaussian functions, the convolution product of the working memory and the focus map leads to an activity profile that is the profile in the working memory translated by the vector represented in the focus map. This profile is the prediction of the future state of the working memory and is then used to slightly excite the

working memory. After the eye movement and when the saliency map is updated, the previously scanned stimuli emerge in the working memory as a result of the conjunction of the visual stimuli in the saliency map and the prediction of the working memory : the prediction is combined with the new perception. This is exactly the same mechanism than the one used when a stimulus emerges in the working memory owing to the conjunction of the activity in the saliency map and the focus map.

Simulation and results

The visual environment consists in three identical stimuli that the robot is expected to scan successively exactly once. A stimulus is easily discriminable from the background, namely a green lemon on a white table. A complete activation sequence of the different maps is illustrated on Fig. E.4. The saliency map is filled by convolving the image captured from the camera by a green filter in HSV coordinates such that it leads to three distinct stimuli¹⁹.

At the beginning of the simulation (Fig. E.4a), only one of the three stimuli emerges in the focus map, thanks to the strong lateral competition that occurs within this map. This stimulus, present both in the focus map and in the saliency map, emerges in the working memory. The activation within the anticipation map reflects what should be the state of the saliency map, restricted to the stimuli that are in the working memory, after the movement that brings the focused one in the center of the visual field. During the eye movement (Fig. E.4b), no visual information is available and the parameter τ in (E.1) and (E.6) is adjusted so that only the units in the anticipation map remain active, whereas the activity of the others tends to zero. After the eye movement and as soon as the saliency map is fed with the new visual input, the working memory is updated thanks to the excitation from both saliency and anticipation map at a same location : the prediction of the state of the visual memory is compared with the current visual information. A new target can now be elicited in the focus map thanks to a switch mechanism similar to that described in [Vitay et Rougier, 2005], but not detailed here. This mechanism acts like the inhibition of return presented in section E.1.2.0 : the memorized locations in the working memory are inhibited in the focus map, therefore biasing the competition in it, so that only a non-already focused stimulus can be the next target to focus on.

In order to illustrate more explicitly the role of the anticipatory signal, we now consider a second experiment. In that experiment, the visual scene consists in only two identical stimuli (Fig. E.5).

The task is the same as previously, namely the robot must scan each stimulus only once, but the experimental conditions are slightly different : we will enforce the robot to scan these targets in a predefined order. To bias the spatial attention toward one of the two targets, we first increase the intensity of the leftmost target. Then, when the saccade to center that target is performed, we refresh the display and increase the intensity of the rightmost target. In that way, the scenario is the following :

1. Select the leftmost target
2. Focus on that target

¹⁹A video of the model is available at <http://www.loria.fr/~fix/publications.php>

3. After the saccade, when the display is refreshed, select the rightmost target
4. Perform the saccade to center it

The visual bias we add makes us able to get the same experimental conditions over the trials. During a trial, we record the activity of the neurons whose receptive field covers one of the five positions denoted x_0 , x_1 , x_2 , x_3 and x_4 on the figures, in the four maps : visual, focus, wm and anticipation. In a typical trial, we will have a target at x_1 and x_3 , then, after the first saccade, the targets will be at x_2 and x_4 to finally occupy, after the last saccade, the positions x_0 and x_2 (Fig. E.6 Top).

Moreover, two conditions are considered ; In the first one (Fig. E.6), the anticipation is enabled whereas in the second one (Fig. E.7), the anticipatory signal is disabled. At the beginning of the trial, the targets are at positions x_1 and x_3 so that the neurons in the visual map at these positions are excited (dashed line) whereas the neurons at the other positions remain silent. The two positions x_1 and x_3 compete for the spatial attention. Since we added a bias toward the target at position x_1 , the spatial attention is on target x_1 , rather than on target x_3 , so that the activity of the neuron at position x_1 in the focus map (solid line) grows up, whereas the activity of the neuron at position x_3 in the same map decreases to zero. The attention on target x_1 enables it to emerge in the working memory (dash-dot line). The task is now to produce an eye movement that will center that target. The anticipatory mechanism predicts that when that target will be centered, it will occupy the position x_2 : the activity of the neuron at position x_2 grows up (dotted line). As soon as the saccade is performed, we refresh the display. The two targets now occupy the positions x_2 and x_4 . The bias toward the rightmost target enforces that target to be attended. The activity of the neurons at position x_4 in the focus map and the working memory grows up. Whereas the target at position x_4 emerges in the working by the conjunction of an activity in the visual input and the focus map, the target at position x_2 emerges thanks to the visual input and the anticipatory signal. As we can see on figure E.7 in which the anticipatory signal was disabled, the position of the first attended target cannot be updated at position x_2 . Finally, a saccade to center the target at position x_4 is performed. In the case the anticipation is present, the new positions of the two targets are in the working memory at x_0 and x_2 , whereas when there is no anticipation, only the last attended target is in the working memory.

E.1.4 Discussion

In this paper, we have presented a continuous attractor network model that is able to anticipate the consequences of its own movements by actually predicting the visual scene as it is supposed to be after the execution of an action. Furthermore, the model also illustrates how this information is used in the context of a serial search of a target among a set of distractors : each already focused target is kept within a working memory area that is updated with regards to eye movements.

The model is of a completely distributed nature and does not require any central supervisor. All the units in the model satisfy a dynamical equation. When dealing with this kind of dynamic model, the integration time of the units is a critical factor as shown in [Rougier et Vitay, 2006],

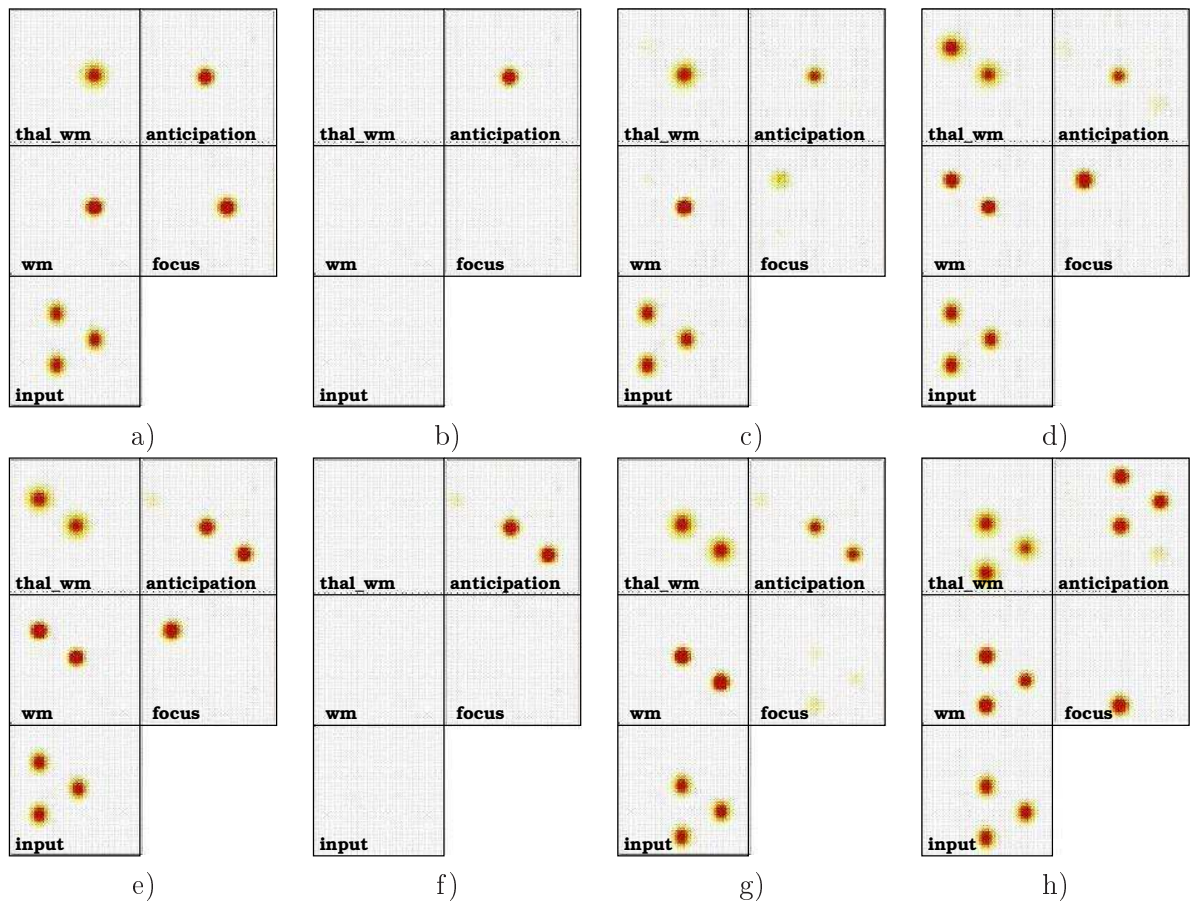


FIG. E.4 – A sequence of evolution of the model during an overt visual scan trial. a) One of the three stimuli emerges in the focus map and the anticipation’s units predict the future state of the visual memory (the maps *wm* and *thal_wm*). b) During the execution of the saccade, only the units in the anticipation map remain active. c) The focused stimulus emerge in the memory since it is both in the saliency map and the anticipation map at the same location. d) A new target to focus is elicited. e) The future state of the memory is anticipated. f) The saccade is executed and only the prediction remains. g) The two already focused stimuli emerge in the memory. h) The attentional focus lands on the last target.

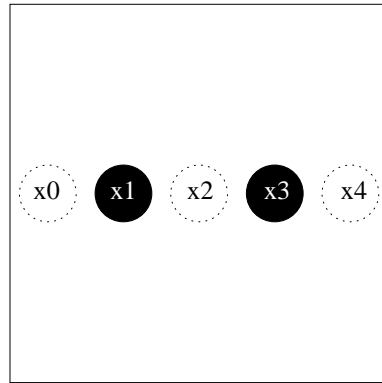


FIG. E.5 – The scene consists in two identical stimuli, the black blobs, initially symmetrically positioned around the center of gaze. The task is to successively focus on each target. During a trial, we measure the activity of neurons whose receptive field covers the five positions represented by the dashed circles and denoted as x_0 , x_1 , x_2 , x_3 and x_4 , in several maps.

which shares some ideas with the present model. It means that in our case, even if we make the hypothesis that the perception is available during the saccade (ignoring also that the perception is smeared), the working memory could be updated dynamically with the perception only if the movement's speed doesn't exceed a critical limit. In the case of saccadic eye movements, it is then necessary to have an anticipatory mechanism. We are definitely speaking about anticipation since a prediction about the futur perception is used to maintain a coherent memory which is mandatory to accomplish the task we designed. It is nonetheless not limited to that particular case since scanning several potential targets is one of the basic primitives we use when performing a visual search task.

The question of learning the underlying transformation of the anticipatory mechanism, namely the convolution product of the focus map and the working memory, remains open and is still under study. We did implement a learning mechanism, under restrictions and strong hypotheses that relies heavily on the difference between the pre-saccadic prediction and the post-saccadic actual perception. This self generated signal is able to measure to what extent the prediction is correct or not. Hence, it is quite easy to modify the weights accordingly. The main difficulty during learning remains the sampling distribution of examples within the input space which is a well known problem in information and learning theory. Without any additional motivational system that could bias the examples according to a given task, it is quite unrealistic to rely on a regular distribution of examples.

Finally, the coherence of the visual world is solely based on an anticipatory mechanism that ultimately allows to identify targets before and after a saccade despite drastic changes in the visual perception. The prediction of the future state of the visual memory enriches the perception of the visual world in order, for example, to prevent focusing twice on the same stimulus. Of course, this model does not pretend to be complete nor accurate and does not tackle a number of problems that are directly related to visual perception. However, we think that the possibility to unconsciously anticipate our own actions using a dynamic working memory could be extended to other motor tasks involving other perception as well.

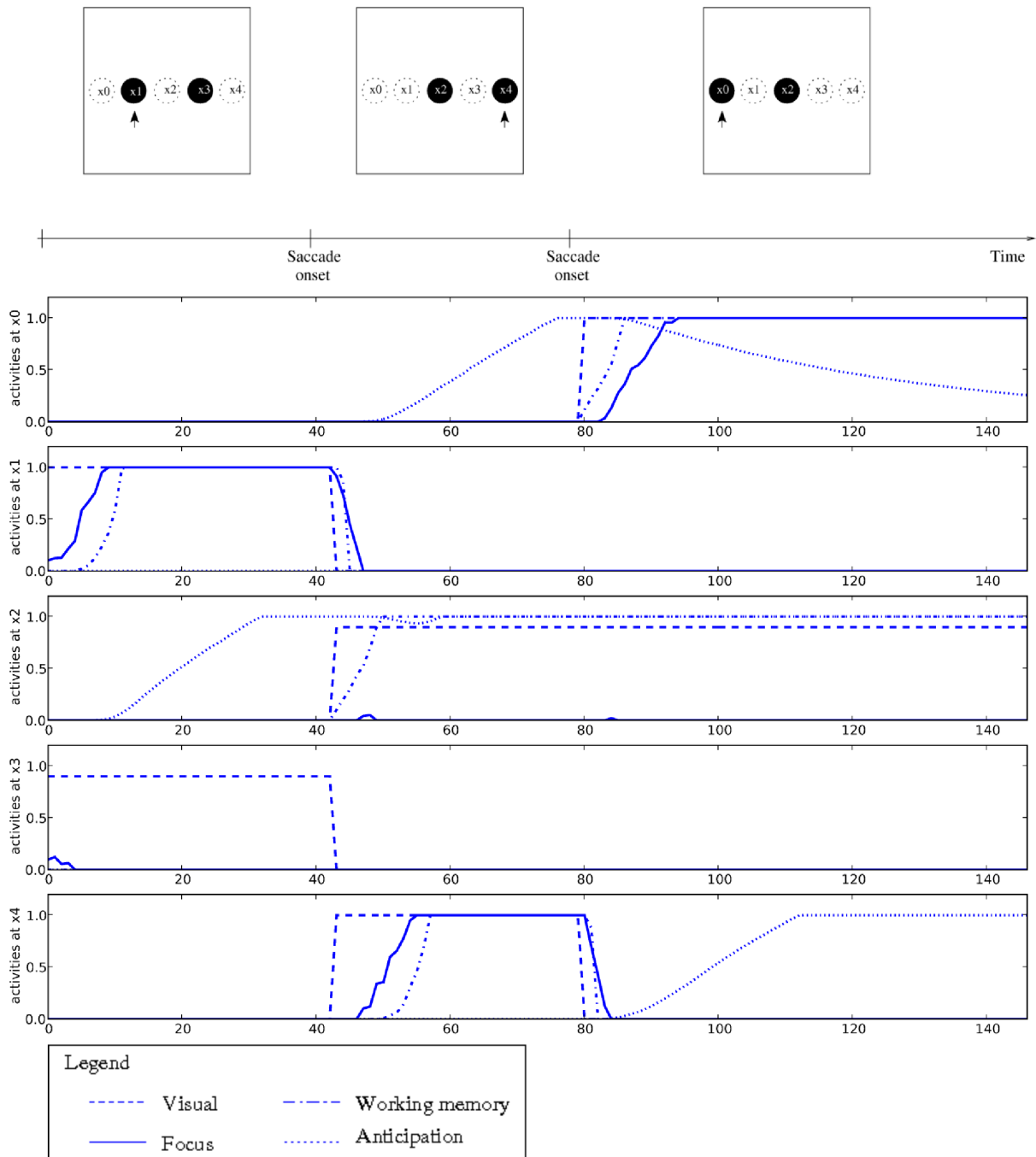


FIG. E.6 – Case with the anticipatory signal enabled. We record the activity of neurons whose receptive field covers one of the five positions x_0 , x_1 , x_2 , x_3 and x_4 , in the four maps : visual, focus, wm and anticipation. During the trial, we add a bias toward one of the targets so that the attention directs to the biased target (that target is shown by the arrow). Each subplot represents the activity of the neurons in each map at a given position. The dashed line represents the activity of the neuron in the visual map, the solid line the activity of the neuron in the focus map, the dash-dot line the activity of the neuron in the working memory and the dotted line the activity of the neuron in the anticipation map. Please read the text for explanations on these curves.

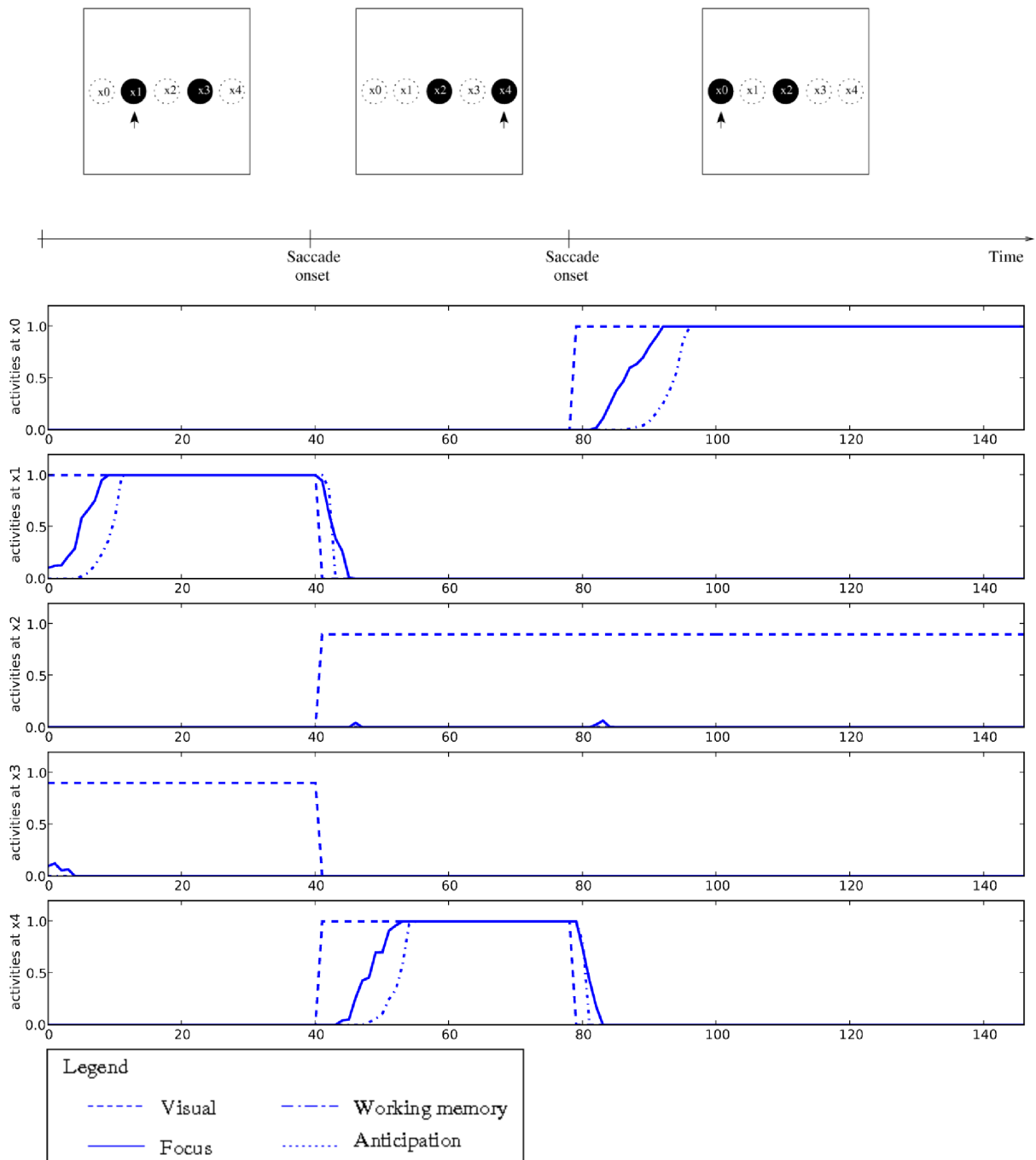


FIG. E.7 – The experiment is the same as on figure E.6 except that the anticipatory signal is disabled.

Appendix

Dynamic of the Neurons

Each sigma neuron *loc* in a map computes a numerical differential equation given by equation (E.7), which is a numerized version of that proposed in [Amari, 1977] and [Taylor, 1999] :

$$\begin{aligned}
 act_{loc}(t+1) = \sigma(act_{loc}(t)) &+ \frac{1}{\tau}(-(act_{loc}(t) - baseline) + \frac{1}{\alpha} \sum_{aff} w_{aff} act_{aff}(t) \\
 &+ \frac{1}{\alpha} \sum_{lat} w_{lat} act_{lat}(t)))
 \end{aligned} \tag{E.7}$$

Each sigma-pi neuron *loc* in a map computes a numerical differential equation given by equation (E.8) :

$$\begin{aligned}
 act_{loc}(t+1) = \sigma(act_{loc}(t)) &+ \frac{1}{\tau}(-(act_{loc}(t) - baseline) + \frac{1}{\alpha} \sum_{lat} w_{lat} act_{lat}(t) \\
 &+ \frac{1}{\alpha} \sum_{(i,j) \in E_{loc}} w_{sigmap_i} act_{aff_i}(t) act_{aff_j}(t))
 \end{aligned} \tag{E.8}$$

where $\sigma(x)$ is a semi-linear function assuring that $0 \leq \sigma(x) \leq 1$, τ is the time constant of the equation, α is a weighting factor for external influences, *aff* is a neuron from another map and *lat* is a neuron from the same map. To know how the set of afferent neurons E_{loc} is determined in the case of a sigma-pi map, please refer to the section E.1.3.0 describing the model.

The size, τ , α and baseline parameters of the different maps are given in the following table :

<i>Map</i>	<i>Size</i>	<i>Type</i>	<i>Baseline</i>	τ	α
INPUT	40*40	Sigma	0.0	0.75	6.0
FOCUS	40*40	Sigma	-0.05	0.75	13.0
WM	40*40	Sigma	-0.2	0.6	13
THAL_WM	40*40	Sigma	0.0	0.6	13
ANTICIPATION	40*40	Sigma-Pi	0.0	2.0	5.0

Connections intra-map and inter-map

The lateral weight from neuron *lat* to neuron *loc* is :

$$w_{lat} = Ae^{-\frac{dist(loc,lat)^2}{a^2}} - Be^{-\frac{dist(loc,lat)^2}{b^2}} \text{ with } A, B, a, b \in \mathfrak{R}^+, loc \neq lat . \tag{E.9}$$

where $dist(loc, lat)$ is the distance between *lat* and *loc* in terms of neuronal distance on the map (1 for the nearest neighbour). In the case of a “receptive field”-like connection between two maps,

the afferent weight from neuron *aff* to neuron *loc* is :

$$w_{aff} = Ae^{-\frac{dist(loc,aff)^2}{a^2}} \text{ with } A, a \in \mathfrak{R}^{*+} \quad (\text{E.10})$$

In the case of the sigma-pi connections, all the weights are the same :

$$w_{sigma-pi} = A \text{ with } A \in \mathfrak{R}^{*+} \quad (\text{E.11})$$

The connections in the model are described the following table :

<i>Source Map</i>	<i>Destination Map</i>	<i>Type</i>	<i>A</i>	<i>a</i>	<i>B</i>	<i>b</i>
INPUT	FOCUS	receptive-field	0.25	2.0	-	-
FOCUS	FOCUS	lateral	1.7	4.0	0.65	17.0
INPUT	WM	receptive-field	0.25	2.0	-	-
FOCUS	WM	receptive-field	0.2	2.0	-	-
WM	WM	lateral	2.5	2.0	1.0	4.0
WM	THAL_WM	receptive-field	2.35	1.5	-	-
THAL_WM	WM	receptive-field	2.4	1.5	-	-
ANTICIPATION	ANTICIPATION	lateral	1.6	3.0	1.0	4.0
WM, FOCUS	ANTICIPATION	sigma-pi	0.05	-	-	-
ANTICIPATION	WM	receptive-field	0.2	2	-	-

E.2 From physiological principles to computational models of the cortex

Jeremy Fix, Nicolas Rougier et Frederic Alexandre
Journal of Physiology - Paris, 101(1-3), pp. 32–39, 2007

Abstract

Understanding the brain goes through the assimilation of an increasing amount of biological data going from single cell recording to brain imaging studies and behavioral analysis. The description of cognition at these three levels provides us with a grid of analysis that can be exploited for the design of computational models. Beyond data related to specific tasks to be emulated by models, each of these levels also lays emphasis on principles of computation that must be obeyed to really implement biologically inspired computations. Similarly, the advantages of such a joint approach are twofold : computational models are a powerful tool to experiment brain theories and assess them on the implementation of realistic tasks, such as visual search tasks. They are also a way to explore and exploit an original formalism of asynchronous, distributed and adaptive computations with such precious properties as self-organisation, emergence, robustness and more generally abilities to cope with an intelligent interaction with the world. In this article, we first discuss three levels at which a cortical circuit might be observed to provide a modeler with sufficient information to design a computational model and illustrate this principle with an application to the control of visual attention.

E.2.1 Motivations

Building models and frameworks to compute in a biologically inspired way is fruitful for both neuroscience and computer science. On one hand, it leads to simulations that allow a better understanding of the complex relations between structure and function in the brain. Particularly, it is possible to investigate the validity of hypotheses onto these relations. On the other hand, this approach allows to explore a formalism of computation that is hardly used in computer science, based on distributed, asynchronous and adaptive local automata and to learn to master properties such as emergence, unsupervised learning, multimodal processing, robustness, etc. The most critical issue in this process is to get the pertinent information from neuroscience and to select or design the adequate computational principles. The information can be extracted from raw data recorded in nervous systems or in behaving animals. It can also be more elaborated and derive from a more conceptualized source, like a functional model. The computational mechanisms can be derived from a solid mathematical framework (if available) and benefit from its properties (stability, convergence proof). Else, it can be *ad hoc* mechanisms, suitable for experimental investigations, the theoretical framework of which remains to be built. To implement such a complex task as endowing an autonomous robot with visual search behavior, the interplay between neuroscience and computer science involves several levels of description.

The microscopic level

The microscopic level requires to identify the adequate elementary unit of computation depending on the purpose of the model. For tasks in which the goal is to understand the inner neuronal functioning, either at the level of a single cell or at the level of communication and synchronization between two neurons, spiking neuron models are generally preferred. In tasks like visiomotor coordination involving global patterns of cerebral activity and behavioral assessment, we rely on the mean firing rate of neurons or even on the behavior of elementary circuits of neurons that can be found in structures like the cerebral cortex [Burnod, 1989]. Choosing such an intermediate level of description is also fundamental from a computational point of view since handling the temporal behavior of a neuron at the level of the spike is a very consuming task for simulations and is not compatible with the simultaneous evaluation of millions of neurons. Fortunately, mean firing rate models neuronal circuits, as proposed for example by the Continuum Neural Field Theory [Amari, 1977, Taylor, 1999] have proved to be efficient and faithful, compared to cellular recording of population of neurons (like Local Field Potential). Such automata aim at explaining the behavior of the cortical circuitry and generally lay emphasis on the variety of inputs and outputs which are integrated in cortical circuits [Bullier, 2001]. Whereas thalamic inputs are generally implemented with a classical integrative model emulating stimulus-specific units [Ballard et al., 1997], cortico-cortical relations are represented as performing a gating effect, implemented with multiplicative connections, and representing feedback as a modulatory activity onto the perceptive flow [Reynolds et al., 2000]. Then, the implementation of a cortical area is only specified by the nature of feed-forward and feedback loops feeding a map of interconnected units. The behavior of the whole is only a consequence of patterns of events which are presented in the flows and of the interplay of the units. In the simulation, everything is a matter of local numerical computations.

The mesoscopic level

The mesoscopic level is that of cerebral regions, homogeneous at a structural as well as functional level. In the cortex, cortical areas have been detected for a long time, by pure observation of the cytoarchitecture (as soon as the beginning of the 20th century by Brodmann). From that time, a huge quantity of work has been done to relate these areas to a functional role and to gather them in information flows. This has benefited from great progresses in visualization and brain activity measurement techniques (e.g. fMRI, antidromic methods). Sensory and motor poles, and the nature of processing between them have been intensively discussed. Particularly, in the visual case, two main processing flows have been identified from the occipital visual region [Ungerleider et Mishkin, 1982] : one toward the limbic temporal region (ventral pathway) dedicated to visual stimuli identification and the other toward the proprioceptive and parietal regions (dorsal pathway), the role of which is still intensively discussed [Milner et Goodale, 1995], from pure spatial localization to body involvement in visual objects seen as tools. Both temporal and parietal representations are the internal and external sensory representations used by the frontal lobe, seen as the motor pole, responsible for the temporal organization of behavior [Fuster, 1997].

This simplified picture has to be made more complex in several ways. Firstly, instead of sequential processing flows, parallel and redundant processing is reported, in dozens of inter-

connected cortical areas [VanEssen et Maunsell, 1983, Zeki, 1978] (e.g. color, depth, texture in various areas of the temporal lobe ; eye, head and body centered information in the parietal lobe). Secondly, even if this presentation lays emphasis on the feed-forward integration (how to transform visual information into representations of the identity and the location of relevant objects), feedback information seems to play a role at least as important as feed-forward influence [Bullier, 2001] (e.g. receptive fields of neurons in the parietal lobe changing according to body parts orientation [Cohen et Andersen, 2002] ; the features of a target to look modulate the activity of V4 neurons [Desimone et Duncan, 1995]). Thirdly, our misleading functional and symbolic intuition and the weaknesses of brain imaging techniques incites us to imagine a step by step processing, where information follows cycles of processing and builds elaborated representations, whereas the functioning is certainly much more distributed, asynchronous and sparse [Bullier, 2001].

To better understand and master this counter-intuitive functioning mode, computational models and simulations are of very high interest. From a pure structural description (number and size of areas, connectivity schemes between them) and from necessary functional recommendations (local and asynchronous evaluation of units), the local functioning rules of units (as discussed in the previous section) must be confronted here to the achievement of stable patterns of activity, as observed in the living cortex. This is consequently a way of refinement for the functioning rules of the local automaton. The overall activity pattern which is obtained can also be interpreted as a way to validate the behavioral level, as discussed at the macroscopic level.

The macroscopic level

The macroscopic level is concerned with selecting the task or the behavior you are interested in, and defining the adequate set of areas (together with their connectivity) which is supposed to emulate that task or behavior. Modern imaging techniques and their associated statistical processing offer a valuable tool to relate experimental tasks to brain activations but are not completely satisfactory for several reasons. Firstly, the brain imaging technology itself gives some limitations relative to the kind of behaviors and subjects that can be explored (which are de facto stereotyped), to the parts of the brain easy to observe and to their spatial and temporal resolution. More importantly, observing a pattern of activity in the brain does not give a complete information neither about the role of the recorded region in the behavior nor about the kind of information it stores and processes. More generally, the observed pattern of activity does not provide an interpretation of the underlying cognitive processes. Consequently, these data must be correlated with more behavioral, or even psychological, data and also with brain theories that are themselves elaborated from the synthesis and interpretation of a large quantity of experimental results. In this picture, computational models and simulations are complementary ways of investigation, especially interesting to assess the validity of an hypothesis or to technically explore an intuition. Using the ascendant approach through levels of description, as summarized here, also ensures that the model does not obey a too sequential, centralized, human-like analysis : whatever the possible bias toward such an interpretation, the main constraint is that the simulation has to work in a completely distributed way while yielding an emergent behavior with acceptable spatial and temporal characteristics and with comparable underlying distributed patterns of activity.

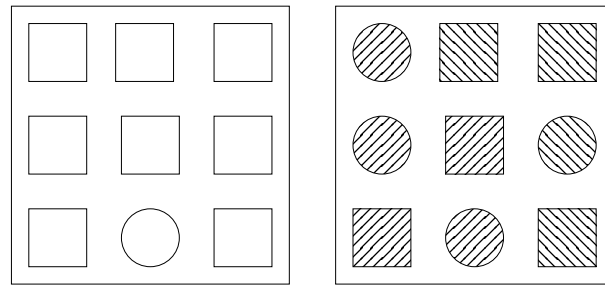


FIG. E.8 – Feature search can be performed very quickly as illustrated on the left part of the figure ; the disc shape literally pops out from the scene. However, as illustrated on the right part of the figure, if the stimuli share at least two features, the pop out effect is suppressed. Hence, finding the disc shape with the stripes going from up-left to down-right requires an active scan of the visual scene.

E.2.2 The computational approach

The computational approach requires in fact to cope with all these three levels at once in order to have working computational models that can explain or predict some experimental results. However, this is a daunting task since we have to simultaneously integrate data from both anatomy, physiology and psychology. This clearly requires to make clear assumptions and choices at several different levels. We can choose for example among elementary models of the neuron, architectures, granularity of models, adaptive algorithm, etc. As an illustration, we would like to introduce very briefly one widely studied cognitive phenomenon (visual attention) and explain what are the choices we did, what those choices implied on the model and what were their consequences regarding the constraints brought by the framework of distributed, asynchronous, and numerical computations we are using.

Psychological and physiological data

In the early eighties, [Treisman et Gelade, 1980] proposed that the brain actually extracts, in parallel, some basic features from the visual information. Among these basic features, that have been reviewed by [Wolfe, 1998], one can find color, shape, orientation or motion. If we consider a visual search behavior, this task is then equivalent to the finding of conjunction of features that best describes the target. In this sense, [Treisman et Gelade, 1980] distinguished two main paradigms (see also [Duncan et Humphreys, 1989] who proposed a classification of visual search efficiency in terms of target-distractors similarities) :

- **Feature search** refers to a search where the target sufficiently differs from the distractors to literally pop out from the search scene
- **Conjunction search** refers to a search where the time to find the target is closely linked to the size of a subset of the search scene, which contains stimuli that are quite similar to the target.

The figure E.8 illustrates these two search modes using two tasks whose common goal is to localize a given target. The second task takes more time than the first one and the strategy

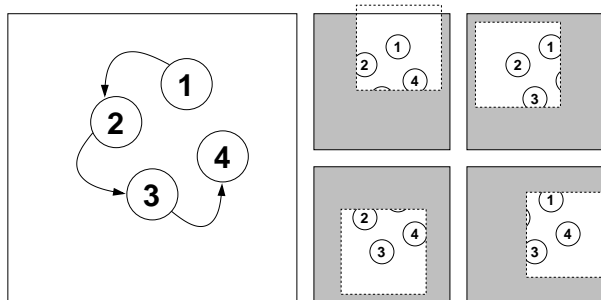


FIG. E.9 – When scanning a visual scene, going for example from stimulus 1 to stimulus 4, as illustrated on the left of the figure, the image received on the retina is radically changed after each eye movement. When the task requires to memorize the positions of the previously focused stimuli, the difficulty is to be able to update their memorised positions after each saccade. The figures on the stimuli are shown only for explanation purpose and do not appear on the screen ; all the stimuli are identical.

generally used to perform the task is to successively scan each circle until finding the target. The reaction time then closely depends on the size of the subset of stimuli composed by the circles.

While the pop-out effect can be explained solely on stimulus-driven activities, it must be emphasized that in general, the selection of a subset of potential targets highly depends on the target to look for. This selection process is one component of the more general concept of visual attention. While the brain is submerged by a high quantity of information, and because its resources are somehow limited, it must perform a selection of the relevant information among what it receives.

In the visual case, this selection mechanism is referred to as visual attention and can take different forms. On the one hand, **feature based attention** characterises the modulation on the processing of visual information by the knowledge of the features of an object of interest [Motter, 1994a]. On the other hand, [Rizzolatti et al., 1987] provided evidences for the influence of saccadic eye movements on directed attention, which led to the premotor theory of attention. [Moore et Fallah, 2001] have also shown that the preparation of an eye movement toward a specific location provides a bias to the cells whose receptive field covers that location. This spatial bias is known as **spatial attention**. Several experiments have provided evidences that our brain can provide such a spatial bias covertly in the absence of the overt deployment of eye movements [Posner et al., 1980], and that the underlying circuits mediating the covert and overt deployment of attention might considerably overlap [Awh et al., 2006].

The first neural correlate of visual attention at the single cell level has been discovered by Moran et al. [Moran et Desimone, 1985] in V4 where neurons were found to respond preferentially for a given feature in their receptive field. When a preferred and a non-preferred stimulus for a neuron are presented at the same time in its receptive field, the response becomes an average between the strong response to the preferred feature and the weak response to the non-preferred one. But when one of the two stimuli is attended, the response of the neuron represents the attended stimulus alone (strong or poor), as if the non-attended were ignored. Attentional modulation of neuronal activity was also observed in other cortical areas. In [Treue et Maunsell, 1996], the

author reported feature-based attentional modulation of visual motion processing in area MT. An increasing literature is also reporting that the modulatory effect of attention is not restricted to the extrastriate cortex but also extends to the early visual areas [Silver et al., 2007].

The observed modulatory effect of attention on the processing of single units raises the intriguing issue of determining its origin(s). As detailed in the introduction, the processing of visual information is supposed to rely on two pathways. On the one hand, the ventral pathway, going from the occipital lobe through the temporal lobe is classically thought to mediate object recognition [Gross, 1994]. Several studies have shown the influence of the intrinsic properties of an object of interest on the processing of single cells [Chelazzi et al., 1998]. This feature-based mechanism could originate from the ventral pathway via massive feedback connections [Rockland et VanHoesen, 1994], and might be generated in the ventrolateral prefrontal cortex to provide a bias corresponding to the features of an object of interest. On the other hand, the dorsal pathway going from the occipital lobe through the parietal lobe is supposed to be involved in producing motor representations of sensory information for the purpose of guiding movements [Cohen et Andersen, 2002, Matelli et Luppino, 2001]. The temporal properties of neurons in the parietal cortex cannot be solely explained by proprioceptive feedbacks as a consequence of a performed movement. Rather, anterior areas might provide more posterior areas with the parameters of an impending movement, then leading to anticipatory activations or remapping, as observed by Colby et al. [Merriam et Colby, 2005, Merriam et al., 2007]. The later have shown that, in the case of saccadic eye movements, neurons in lateral intraparietal area (LIP) exhibit saccade-related activity occurring before, during and/or after a saccade bringing a stimulus in the receptive field of the recorded neurons. These recordings reveal that a circuit, involving parietal areas, is able to predict the future position of currently observed stimuli after an impending eye movement. Moreover, in the case of overt deployment of attention, a crucial issue is to be able to update the position of previously attended stimuli after each saccade (see figure E.9).

Saccadic eye movements are too fast to suppose that a memory of the targets can be continuously updated with the visual input. Hence, a specific mechanism using the memorized locations of the targets and an impending eye movement, predicting the future positions of these targets must exist. The frontal eye field (FEF) might be involved in such a circuit. As shown by [Sommer et Wurtz, 2004a], FEF receives projections from the superior colliculus (SC), relayed by the mediodorsal thalamus, which could convey a corollary discharge of movement commands. Several studies have also shown memory related activity in FEF [Lawrence et al., 2005] as well as predictive responses [Umeno et Goldberg, 1997]. This illustrates that the brain consists in several cooperating areas and that a behavior observed in tasks such as a visual search actually emerges from distributed computations.

Computational approaches to visual attention

In the field of computational neuroscience, several attempts at modeling visual attention have been proposed. The pioneering work of [Koch et Ullman, 1985], relying on the Feature Integration Theory [Treisman et Gelade, 1980], distinguishes several channels extracted from the visual input (color, orientation, intensity), each of them represented in different sets of maps, used to build conspicuity maps to finally lead to a single spatial map representing the behavioral relevance

of each location in the visual field, the so-called saliency map. The selection of a location to attend to is then determined by a winner-take-all operation on the saliency map. A memory of the attended locations finally biases that winner-take-all computation to avoid attending to previously attended locations. This phenomenon reflects one component of the inhibition-of-return introduced by Posner and detailed in the previous section : a cued location facilitates the deployment of attention at that location when the time between the cue and the target is short, but, for longer delays, we observe the reverse effect and, if the target is presented at a cued location, its processing takes longer. The model proposed by Koch and Ullman was the first step to further developments [Itti et Koch, 2001] but, from the past few years, we are observing a slight shift from purely feedforward models to models using both feedforward and feedback projections [Tsotsos et al., 1995], since it is now widely accepted that feedback influences play a crucial role in single unit processing. Among these models, we will focus in the rest of this article on the work of [Hamker, 2004]. This model clearly distinguishes between the ventral and dorsal pathways with a feature-based and a spatial stream processed along two separate pathways. It also emphasizes the role of feedback projections that are supposed to be the cause of attentional effects. The ventral stream provides a feature-based bias corresponding to an object of interest (an object we are looking for in a visual search task for example) and the dorsal stream provides a spatial bias corresponding to a region of interest, that might be the target for an impending eye movement. The main hypothesis is that V4 could be an intermediate layer, being the major source of information carrying along the ventral and dorsal pathways, and the major target of projections from higher cortical areas. The proposed model exhibits good performances in visual search task but one of the limitations is that the model is restricted to the covert deployment of attention, where no eye movement is initiated. We will see in the following sections a possible extension of this approach to deal with saccadic eye movements.

A computational model

The models we propose are built in the framework of local, distributed, asynchronous and numerical computations by considering assemblies of units that we call maps, each unit being connected with other units in the same map by lateral connections and with units from other maps by afferent connections. A unit is a stand-alone computational element, characterised by a numeric activity $u_M(\mathbf{x}, t)$ that is locally updated by computing the influence of input units. The activity of each unit follows the ordinary differential equation (E.12) coming from the Continuum Neural Field Theory [Amari, 1977].

$$\begin{aligned} u_M(x, t + 1) &= u_M(x, t) + \tau \cdot \delta u_M(x, t) \\ \delta u_M(x, t) &= \sum_{y \in M} w_{xy} \cdot u_M(y, t) + I(x) \end{aligned} \tag{E.12}$$

where M and M' are maps of units and $I(x)$ is a function computing the influence of afferent units.

A key point is to determine how the cells combine their inputs to perform their local computations. V4 neurons are a striking example of attentional modulation at the single cell (or small population) level, as explained in the previous section. Let us consider a population of orientation selective cells, receiving afferent connections from lower level areas, these connections being directly modulated by feedback connections coming from higher level areas. These feedback

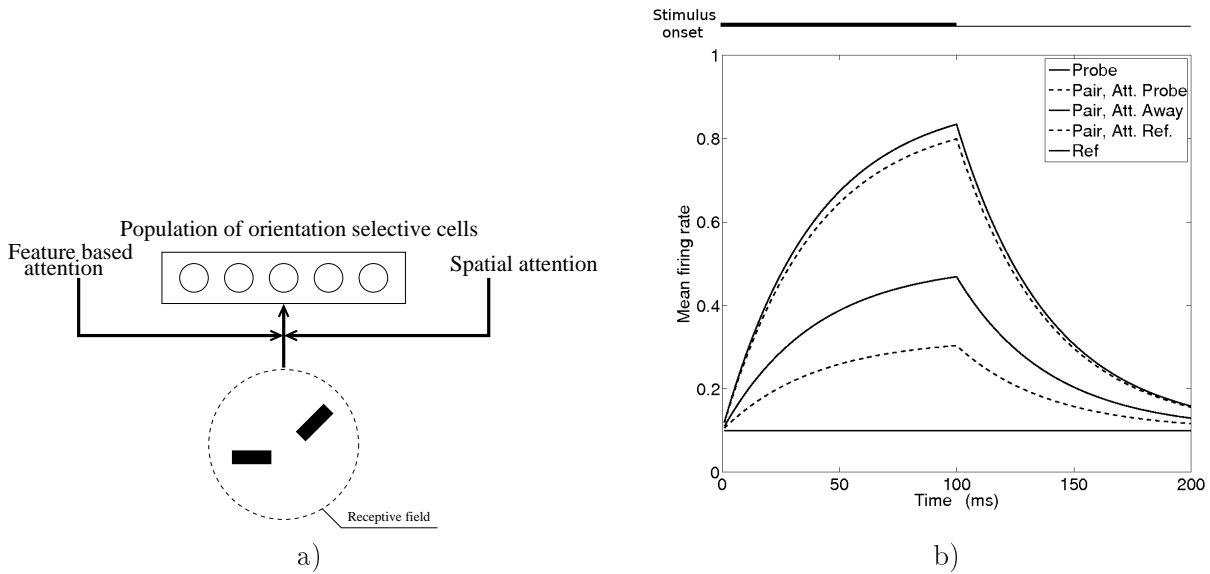


FIG. E.10 – a) A population of orientation selective cells sharing the same receptive field. The afferent connections are modulated by feature-based and spatial attention as proposed in the contrast gain model b) When a preferred and non-preferred stimuli are both presented in the receptive field, the response of the neuron is an average between the responses to the stimuli presented separately. When feature-based attention is directed either toward one or the other stimulus, the cells respond as if only the attended stimulus was present. This effect is even stronger when spatial attention is directed toward the receptive field. The plots are displayed in the same order than the legend.

projections carry information about the features of an object of interest (feature based attention) and a location that might be the target of an impending action (spatial attention) that have been shown to have an influence on the response of V4 neurons. In [Taylor et al., 2006, Reynolds et al., 2000], the authors show that, among different possibilities of integration of the feedback modulation, the contrast gain model seems to be the most suitable (figure E.10a). In this model, if we record the activity of one unit while presenting two stimuli in the receptive field of the population (a preferred and a non-preferred stimulus for the considered unit), we observe two properties (figure E.10b) :

- attending the preferred stimulus drives the activity of the cell toward its response when only the preferred stimulus is presented
- attending the non-preferred stimulus drives the activity of the cell toward its response when only the non-preferred stimulus is presented

These modulatory effects reflect the biased competition mechanism introduced by [Desimone et Duncan, 1995] and illustrate how we can deal with biological data at the single-cell level.

Let us now consider modeling at a higher level, gathering elementary computational units to form maps. These maps combine flows of information and cooperate in a distributed way to allow the emergence of a global behavior. As an illustrating example, let us consider the mechanisms with which the brain might memorize the attended locations and update these positions after each eye movement, in the case of an overt deployment of attention (figure E.9). In [Vitay et Rougier, 2005], we have proposed to connect homogeneous assemblies of units to build a dynamic

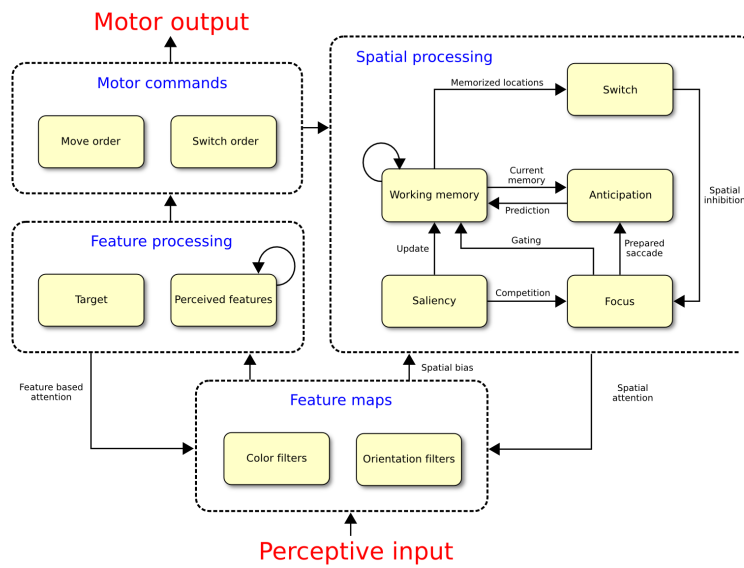


FIG. E.11 – An example of model relying on local, distributed, asynchronous and numerical computations, used to perform a visual search task. Further details can be found in the text below.

working memory circuit. We have extended this model in [Fix et al., 2007b] to take into account the eye movements while performing a visual search task, by adding a mechanism that predicts the consequences of these saccades on the visual perception. We have shown that disrupting this mechanism drastically impairs the performances of the system. At the single cell level, these models are homogeneous and are built with the same basic units. The specificity of each map only comes from the pattern of connections that connect it to the other maps. The structure of these projections defines the architecture at a mesoscopic level.

We can also think about a model as a whole, and use it to perform visual search tasks, measuring psychological variables as, for example, the reaction time. Let us consider the model depicted on figure E.11. This figure illustrates how the models proposed in [Hamker, 2004] and [Fix et al., 2007b] could be combined, leading to one among other possibilities of computational models that gather the psychological and physiological data detailed in the previous sections. The purpose of this article is not to explain deeply the patterns of connections between the maps. Rather, we would like to emphasize how the flows of information are combined to allow the emergence of a behavior in a distributed architecture. The interested reader can find a complete description of the models in [Hamker, 2004] and [Fix et al., 2007b].

The visual input is processed in parallel in different maps, extracting basic features. This distributed representation of the visual input, labeled **Feature Maps**, then feeds two pathways, a spatial non-feature specific one and a feature roughly non-spatial one. The main purpose of the first is to spatially select a location of interest (within the **Saliency** and **Focus** maps), to memorize that that given location has been attended to (the memory consists in a recurrently connected circuit labeled **working memory**), and to anticipate the consequences of an eye movement on this memory, if the movement is triggered (with the **Anticipation** map). A key point of the model is the use of feedback projections of the selected location to the **Feature maps**, biasing this distributed representation toward the features of the stimulus at the attended location.

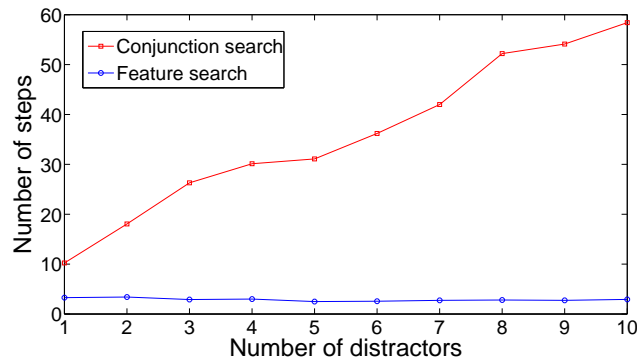


FIG. E.12 – The reaction time, defined as the number of computational steps required to perform the task, increases linearly with the set size in the conjunction search paradigm while keeping constant in the feature search paradigm.

The feature specific pathway then combines this representation with a target template. This template might be a complex combination of basic features and is also projected via feedback connections to the **Feature maps**. The resulting activities in the **Feature processing** maps is then propagated to the decision area so as to provide it with the necessary clues to decide which behavior to adopt. In our case, we distinguish two decisions : one is to switch covertly the locus of attention (covert attention) and the other is to perform an eye movement toward that location (overt attention). When an eye movement is performed, the target is decoded from the **Focus** map. A striking consequence of the distributed nature of the computations is that the memory is fed with an attended location at the same time that the decision to switch covertly or overtly the attention is taken.

If we now use this model to perform a visual search task²⁰ and see it as a black box, we can restrict the measurements to the available ones from the point of view of an external observer, as it would be done by psychologists performing this kind of task with monkeys. We can for example measure the time it takes for the model to perform the task. In a task involving eye movements, we can also record the number of saccades performed by the “subject”, the target of the movements, the scanpath, etc.. The figure E.12 represents the reaction time, function of the set size, in the two paradigms of feature search and conjunction search. In the first case, the task is to detect a blue bar, among green bars. In the second case, the task is to detect a blue bar at 45 degrees among distractors that share at least one feature with the target, namely green bars at 45 degrees or blue bars at 135 degrees. We can then observe a classical set size effect : in a feature search, the time to perform the task does not depend on the number of distractors whereas the time to perform a conjunction search linearly depends on the set size.

The figure E.13 is an illustration of a scanpath obtained during a visual search task in which the model has to perform an eye movement toward each of the black targets, the visual field being limited to the dashed rectangle²¹. The working memory contains all the previously focused

²⁰ Videos of the model performing visual search tasks for the two paradigms of feature and conjunction search are available at <http://www.loria.fr/~rougier/index.php?n=Demos.Demos>

²¹ A video of the model performing a visual search task with explicit eye movements is available at <http://www.loria.fr/~rougier/index.php?n=Demos.Demos>

together, this assumption makes sense and helps us to have a better understanding of the whole. Of course, a question remains on how properties of a functional model would cope with a more detailed model of neurons. Would it change fundamentally or would it be rather a refinement of the existing properties : the strength of computational models is to have the opportunity to refine this level of description, to compare it with more precise observations, without drawing again the whole system.

At the mesoscopic level, modeling meets neuroscience on the analysis of implicated populations and of their underlying behavior. Similarly to the refinement process in neuroscience that corresponds to iteratively better understand the functional role of a cortical map in a task, computational models can also enrich their description of maps of computing units, seen as the crossroads of feed-forward, feed-back and lateral information flows. At this level, adding learning rules, designed as the way to describe the mutual influence of these flows, is certainly the most important task to consider in the near future.

The behavior of computational models at the macroscopic level is intended to have a deep impact in the behavioral neuroscience and to offer them a new behaving entity on which to apply their measurement and analysis. This can be particularly interesting if the simulations are embedded in such autonomous agents as robots, giving a direct access to an embodied cognition.

Bibliographie

- [Ahmad, 1992] Ahmad, S. (1992). Visitation : a neural model of covert visual attention. In *Advances in Neural Information Processing Systems*, pages 420–427. San Mateo, CA :Morgan Kaufman.
- [Albin et al., 1989] Albin, R., Young, A., et Penney, J. (1989). The functional anatomy of basal ganglia disorders. *Trends Neuroscience*, 12(10) :366–75.
- [Alexander et al., 1986] Alexander, G., DeLong, M., et Strick, P. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review Neurosciences*, 9 :357–81.
- [Alexandre, 2008] Alexandre, F. (2008). Cortical basis of communication : local computation, coordination and attention. *Neural Networks*, *accepté*.
- [Alexandre et al., 2008] Alexandre, F., Fix, J., Hutt, A., Rougier, N., et Vieville, T. (2008). On practical neural field parameters adjustment. *2nd conférence en Neurosciences Computationnelles*.
- [Allport, 1993] Allport, D. (1993). *Attention and performance XIV*, chapter Attention and control : have we been asking the wrong questions? A critical review of twenty five years. Cambridge, MA :MIT Press.
- [Amador et al., 1998] Amador, N., Schlag-Rey, M., et Schlag, J. (1998). Primate antisaccades. i. behavioral characteristics. *Journal of Neurophysiology*, 80(4) :1775–86.
- [Amador et al., 2000] Amador, N., Schlag-Rey, M., et Schlag, J. (2000). Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *Journal of Neurophysiology*, 84(4) :2166–70.
- [Amador et al., 2004] Amador, N., Schlag-Rey, M., et Schlag, J. (2004). Primate antisaccade. ii. supplementary eye field neuronal activity predicts correct performance. *Journal of Neurophysiology*, 91(4) :1672–89.
- [Amari, 1977] Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2) :77–87.
- [Andersen et al., 1990a] Andersen, R., Asanuma, C., Essick, C., et Siegel, R. (1990a). Cortico-cortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobe. *Journal of Comparative Neurology*, 296 :65–113.
- [Andersen et al., 1990b] Andersen, R., Bracewell, R., Barash, S., Gnadt, J., et Fogassi, L. (1990b). Eye position effects on visual, memory, and saccade-related activity in areas lip and 7a of macaque. *Journal of Neuroscience*, 10(4) :1176–96.
- [Andersen et al., 1985] Andersen, R., Essick, G., et Siegel, R. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230(4724) :456–8.

- [Andersen et Mountcastle, 1983] Andersen, R. et Mountcastle, V. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience*, 3(3) :532–48.
- [Anderson et VanEssen, 1987] Anderson, C. et VanEssen, D. (1987). Shifter circuits : a computational strategy for dynamic aspects of visual processing. *Proc Natl Acad Sci U S A*, 84(17) :6297–301.
- [Averbeck et al., 2006] Averbeck, B., Latham, P., et Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Review Neuroscience*, 7(5) :358–66.
- [Awh et al., 2006] Awh, E., Armstrong, K., et Moore, T. (2006). Visual and oculomotor selection : links, causes and implications for spatial attention. *Trends Cognitive Sciences*, 10(3) :124–30.
- [Badler et Keller, 2002] Badler, J. et Keller, E. (2002). Decoding of a motor command vector from distributed activity in superior colliculus. *Biological Cybernetics*, 86(3) :179–89.
- [Bahill et al., 1975] Bahill, T., Clark, M., et Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, 24 :191–204.
- [Ballard et al., 1997] Ballard, D., Hayhoe, M., Pook, P., et Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4) :723–42 ; discussion 743–67.
- [Barbas et Mesulam, 1981] Barbas, H. et Mesulam, M. (1981). Organization of afferent input to subdivisions of area 8 in the rhesus monkey. *Journal of Comparative Neurology*, 200(3) :407–31.
- [Barbas et Mesulam, 1985] Barbas, H. et Mesulam, M. (1985). Cortical afferent input to the principalis region of the rhesus monkey. *Neuroscience*, 15(3) :619–37.
- [Barto et al., 1983] Barto, A., Sutton, R., et Anderson, C. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man and Cybernetics*, 13(5) :834–846.
- [Batista et al., 1999] Batista, A., Buneo, C., Snyder, L., et Andersen, R. (1999). Reach plans in eye-centered coordinates. *Science*, 285(5425) :257–60.
- [Berry, 2008] Berry, H. (2008). *Modélisation de systèmes biologiques complexes : Exemples en biologie cellulaire computationnelle, neurosciences computationnelles et applications à l'informatique*. PhD thesis, Université Paris-Sud 11.
- [Beurle, 1956] Beurle, R. (1956). Properties of a mass of cells capable of regenerating pulses. *Philosophical Transactions Series B*, 240 :55–94.
- [Blaser et al., 2000] Blaser, E., Pylyshyn, Z., et Holcombe, A. (2000). Tracking an object through feature space. *Nature*, 408(6809) :196–9.
- [Boniface, 2000] Boniface, Y. (2000). *Etude et développement d'une bibliothèque d'adaptation du parallélisme neuromimétique au parallélisme MIMD*. PhD thesis, Université Henri Poincaré, Nancy 1.
- [Braun et al., 1992] Braun, D., Weber, H., Mergner, T., et Schulte-Mönting, J. (1992). Saccadic reaction times in patients with frontal and parietal lesions. *Brain*, 115 :1359–1386.
- [Bressloff et Coombes, 1997] Bressloff, P. et Coombes, S. (1997). Physics of the extended neuron. *Int. J. Mod. Phys. B*, 11 :2343–2392.
- [Bricolo et al., 2002] Bricolo, E., Gianesini, T., Fanini, A., Bundesen, C., et Chelazzi, L. (2002). Serial attention mechanisms in visual search : a direct behavioral demonstration. *Journal of Cognitive Neuroscience*, 14(7) :980–93.

-
- [Broadbent, 1958] Broadbent, D. (1958). *Perception and communication*. New York : Pergamon Press.
- [Brodmann, 1909] Brodmann, K. (1909). *Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey*. Leipzig :Barth.
- [Brown et al., 2004] Brown, J., Bullock, D., et Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, 17(4) :471–510.
- [Bruce et Goldberg, 1985] Bruce, C. et Goldberg, M. (1985). Primate frontal eye fields. i. single neurons discharging before saccades. *Journal of Neurophysiology*, 53(3) :603–35.
- [Bruce et al., 1985] Bruce, C., Goldberg, M., Bushnell, M., et Stanton, G. (1985). Primate frontal eye fields. ii. physiological and anatomical correlates of electrically evoked eye movements. *Journal of Neurophysiology*, 54(3) :714–34.
- [Bullier, 2001] Bullier, J. (2001). Integrated model of visual processing. *Brain Res Brain Res Rev*, 36(2-3) :96–107.
- [Burnod, 1989] Burnod, Y. (1989). *An Adaptive Neural Network : The Cerebral Cortex*. Prentice Hall.
- [Burr, 2004] Burr, D. (2004). Eye movements : keeping vision stable. *Current Biology*, 14(5) :R195–7.
- [Burr et Ross, 1982] Burr, D. et Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Research*, 22(4) :479–84.
- [Busettoni et Mays, 2005a] Busettoni, C. et Mays, L. (2005a). Saccade-vergence interactions in macaques. i. test of the omnipause multiply model. *Journal of Neurophysiology*, 94(4) :2295–311.
- [Busettoni et Mays, 2005b] Busettoni, C. et Mays, L. (2005b). Saccade-vergence interactions in macaques. ii. vergence enhancement as the product of a local feedback vergence motor error and a weighted saccadic burst. *Journal of Neurophysiology*, 94(4) :2312–30.
- [Büttner et Büttner-Ennever, 2005] Büttner, U. et Büttner-Ennever, J. (2005). Present concepts of oculomotor organization. *Progress Brain Research*, 151 :1–42.
- [Carpenter et Grossberg, 1987] Carpenter, G. et Grossberg, S. (1987). Invariant pattern recognition and recall by an attentive self-organizing art architecture in a nonstationary world. *Proceeding of Neural Network*, 2 :737–745.
- [Carpenter et Grossberg, 2003] Carpenter, G. et Grossberg, S. (2003). Adaptive resonance theory. In *The Handbook of Brain Theory and Neural Networks, Second Edition*, pages 87–90. MIT Press.
- [Carpenter, 1988] Carpenter, R. (1988). *Movements of the eyes, 2nd Edition*. Pion Ltd London.
- [Cassanello et Ferrera, 2007] Cassanello, C. et Ferrera, V. (2007). Computing vector differences using a gain field-like mechanism in monkey frontal eye field. *Journal of Physiology*, 582(Pt 2) :647–64.
- [Cavada, 2001] Cavada, C. (2001). The visual parietal areas in the macaque monkey : current structural knowledge and ignorance. *Neuroimage*, 14(1 Pt 2) :S21–6.
- [Cavanaugh et Wurtz, 2004] Cavanaugh, J. et Wurtz, R. (2004). Subcortical modulation of attention counters change blindness. *Journal of Neuroscience*, 24(50) :11236–43.
- [Cave et Bichot, 1999] Cave, K. et Bichot, N. (1999). Visuospatial attention : beyond a spotlight model. *Psychonomic Bulletin and Review*, 6(2) :204–23.

- [Chance et al., 2002] Chance, F., Abbott, L., et Reyes, A. (2002). Gain modulation from background synaptic input. *Neuron*, 35(4) :773–82.
- [Chelazzi et al., 1998] Chelazzi, L., Duncan, J., Miller, E., et Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, 80(6) :2918–40.
- [Chelazzi et al., 1993] Chelazzi, L., Miller, E., Duncan, J., et Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, 363(6427) :345–7.
- [Cherry, 1953] Cherry, C. (1953). Some experiments on the recognition of speech with one and two ears. *J. Acoust. Soc. Amer.*, 25 :975–979.
- [Chollet et al., 1991] Chollet, F., DiPiero, V., Wise, R., Brooks, D., Dolan, R., et Frackowiak, R. (1991). The functional anatomy of motor recovery after stroke in humans : a study with positron emission tomography. *Annals of Neurology*, 29 :63–71.
- [Clarke, 1957] Clarke, F. (1957). Rapid light adaptation of localised areas of the extra-foveal retina. *Opt Acta (Lond)*, 4 :69–77.
- [Clarke, 1960] Clarke, F. (1960). A study of troxler’s effect. *Opt Acta (Lond)*, 7 :219–236.
- [Cohen et Andersen, 2002] Cohen, Y. et Andersen, R. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Review Neuroscience*, 3(7) :553–62.
- [Colby et Duhamel, 1991] Colby, C. et Duhamel, J. (1991). Heterogeneity of extrastriate visual areas and multiple parietal areas in the macaque monkey. *Neuropsychologia*, 29(6) :517–37.
- [Colby et al., 1995] Colby, C., Duhamel, J., et Goldberg, M. (1995). Oculocentric spatial representation in parietal cortex. *Cerebral Cortex*, 5(5) :470–81.
- [Colby et al., 1996] Colby, C., Duhamel, J., et Goldberg, M. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology*, 76(5) :2841–52.
- [Colby et al., 1988] Colby, C., Gattass, R., Olson, C., et Gross, C. (1988). Topographical organization of cortical afferents to extrastriate visual area po in the macaque : a dual tracer study. *Journal of Comparative Neurology*, 269(3) :392–413.
- [Colby et Goldberg, 1999] Colby, C. et Goldberg, M. (1999). Space and attention in parietal cortex. *Annual Review Neurosciences*, 22 :319–49.
- [Collewyn et al., 1995] Collewyn, H., Erkelens, C., et Steinman, R. (1995). Voluntary binocular gaze-shifts in the plane of regard : dynamics of version and vergence. *Vision Research*, 35(23-24) :3335–58.
- [Constantinidis et al., 2001] Constantinidis, C., Franowicz, M., et Goldman-Rakic, P. (2001). The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nature Neuroscience*, 4(3) :311–6.
- [Constantinidis et Wang, 2004] Constantinidis, C. et Wang, X. (2004). A neural circuit basis for spatial working memory. *Neuroscientist*, 10(6) :553–65.
- [Coombes, 2005] Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biological Cybernetics*, 93(2) :91–108.
- [Corbetta et al., 1998] Corbetta, M., Akbudak, E., Conturo, T., Snyder, A., Ollinger, J., Drury, H., Linenweber, M., Petersen, S., Raichle, M., VanEssen, D., et Shulman, G. (1998). A common network of functional areas for attention and eye movements. *Neuron*, 21(4) :761–73.
- [Corbetta et Shulman, 2002] Corbetta, M. et Shulman, G. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Review Neuroscience*, 3(3) :201–15.

-
- [Cornforth et al., 2005] Cornforth, D., Green, D., et Newth, D. (2005). Ordered asynchronous processes in multi-agent systems. *Physica D : Nonlinear Phenomena*, 204(1-2) :70–82.
- [Craighero et al., 1999] Craighero, L., Fadiga, L., Rizzolatti, G., et Umiltà, C. (1999). Action for perception : a motor-visual attentional effect. *J Exp Psychol Hum Percept Perform*, 25(6) :1673–92.
- [Craighero et Rizzolatti, 2005] Craighero, L. et Rizzolatti, G. (2005). The premotor theory of attention. In *Neurobiology of Attention*, pages 181–186. Elsevier.
- [Curthoys et al., 1984] Curthoys, I., Markham, C., et Furuya, N. (1984). Direct projection of pause neurons to nystagmus-related excitatory burst neurons in the cat pontine reticular formation. *Experimental Neurology*, 83(2) :414–22.
- [Cutzu et Tsotsos, 2003] Cutzu, F. et Tsotsos, J. (2003). The selective tuning model of attention : psychophysical evidence for a suppressive annulus around an attended item. *Vision Research*, 43(2) :205–19.
- [Dayan et Abbott, 2001] Dayan, P. et Abbott, L. (2001). *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, Cambridge MA.
- [Deco et al., 2005] Deco, G., Rolls, E., et Zihl, J. (2005). A neurodynamical model of visual attention. In *Neurobiology of Attention*, pages 593–599.
- [Deng et al., 1986] Deng, S., Goldberg, M., Segraves, M., Ungerleider, L., et Mishkin, M. (1986). The effect of unilateral ablation of the frontal eye fields on saccadic performance in the monkey. In *Adaptive Processes in Visual and Oculomotor Systems*, pages 201–208. Pergamon Press, Oxford.
- [Deniau et Chevalier, 1985] Deniau, J. et Chevalier, G. (1985). Disinhibition as a basic process in the expression of striatal functions. ii. the striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Research*, 334 :227–233.
- [Desimone et al., 1984] Desimone, R., Albright, T., Gross, C., et Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, 4(8) :2051–62.
- [Desimone et Duncan, 1995] Desimone, R. et Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review Neurosciences*, 18 :193–222.
- [Deutsch et Deutsch, 1963] Deutsch, J. et Deutsch, D. (1963). Attention : Some theoretical considerations. *Psychological Review*, 70 :80–90.
- [Dias et Segraves, 1999] Dias, E. et Segraves, M. (1999). Muscimol-induced inactivation of monkey frontal eye field : effects on visually and memory-guided saccades. *Journal of Neurophysiology*, 81(5) :2191–214.
- [Ditchburn et al., 1959] Ditchburn, R., Fender, D., et Mayne, S. (1959). Vision with controlled movements of the retinal image. *Journal of Physiology*, 145(1) :98–107.
- [Dodge, 1919] Dodge, R. (1919). Five types of eye movement in the horizontal meridian plane of the field of regard. *American Journal of Physiology*, 8 :307–329.
- [Dolson, 1997] Dolson, D. (1997). Attentive object recognition in the selective tuning network. Master’s thesis, Univ. of Toronto.
- [Doma et Hallett, 1988] Doma, H. et Hallett, P. (1988). Dependence of saccadic eye-movements on stimulus luminance, and an effect of task. *Vision Research*, 28(8) :915–24.
- [Dominey et Arbib, 1992] Dominey, P. et Arbib, M. (1992). A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex*, 2(2) :153–75.

- [Droulez et Berthoz, 1991] Droulez, J. et Berthoz, A. (1991). A neural network model of sensoritopic maps with predictive short-term memory properties. *Proc Natl Acad Sci U S A*, 88(21) :9653–7.
- [Dubner et Zeki, 1971] Dubner, R. et Zeki, S. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, 35(2) :528–32.
- [Duhamel et al., 1997] Duhamel, J., Bremmer, F., BenHamed, S., et Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature*, 389(6653) :845–8.
- [Duhamel et al., 1992] Duhamel, J., Colby, C., et Goldberg, M. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255(5040) :90–2.
- [Duncan et Humphreys, 1989] Duncan, J. et Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3) :433–58.
- [Dupuy, 1999] Dupuy, J. (1999). *Aux origines des sciences cognitives*. La découverte.
- [Ermentrout, 1998] Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics*, 61 :353–430.
- [Fates, 2008] Fates, N. (2008). Asynchronism induces second order phase transitions in elementary cellular automata. *Journal of Cellular Automata*.
- [Felleman et VanEssen, 1987] Felleman, D. et VanEssen, D. (1987). Receptive field properties of neurons in area v3 of macaque monkey extrastriate cortex. *Journal of Neurophysiology*, 57(4) :889–920.
- [Felleman et VanEssen, 1991] Felleman, D. et VanEssen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1) :1–47.
- [Ferrier, 1875] Ferrier, D. (1875). Experiments on the brains of monkeys. *Proceedings of the Royal Society of London*, (23) :409–430.
- [Findlay et Brown, 2006a] Findlay, J. et Brown, V. (2006a). Eye scanning of multi-element displays : I. scanpath planning. *Vision Research*, 46(1-2) :179–95.
- [Findlay et Brown, 2006b] Findlay, J. et Brown, V. (2006b). Eye scanning of multi-element displays : II. saccade planning. *Vision Research*, 46(1-2) :216–27.
- [Fix, 2007] Fix, J. (2007). A computational approach to the control of voluntary saccadic eye movements. In *1st International Conference on Cognitive Neurodynamics (ICCN 2007), Shanghai (Chine)*.
- [Fix et al., 2007a] Fix, J., Rougier, N., et Alexandre, F. (2007a). From physiological principles to computational models of the cortex. *Journal of Physiology - Paris*, 101(1-3) :32–9.
- [Fix et al., 2007b] Fix, J., Vitay, J., et Rougier, N. (2007b). A distributed computational model of spatial memory anticipation during a visual search task. In *Anticipatory Behavior in Adaptive Learning Systems*, pages 170–188. Springer-Verlag.
- [Frank et al., 2001] Frank, M., Loughry, B., et O’Reilly, R. (2001). Interactions between frontal cortex and basal ganglia in working memory : a computational model. *Cogn Affect Behav Neurosci*, 1(2) :137–60.
- [Frezza-Buet, 1999] Frezza-Buet, H. (1999). *Un modèle de cortex pour le comportement motivé d’un agent neuromimétique autonome*. PhD thesis, Université Henri Poincaré, Nancy 1.

-
- [Frintrop, 2005a] Frintrop, S. (2005a). Chap. 3, state of the art of computational attention systems. In *VOCUS : A Visual Attention System for Object Detection and Goal-directed Search*, volume 3899 of *Lecture Notes in Computer Science*. Springer-Verlag.
- [Frintrop, 2005b] Frintrop, S. (2005b). *VOCUS : A Visual Attention System for Object Detection and Goal-directed Search*. PhD thesis, Univ. of Bonn.
- [Fritzke, 1995] Fritzke, B. (1995). A growing neural gas network learns topologies. *Advances in Neural Information Processing Systems 7*, pages 625–632.
- [Funahashi et al., 1989] Funahashi, S., Bruce, C., et Goldman-Rakic, P. (1989). Mnemonic coding of visual space in the monkey’s dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2) :331–49.
- [Funahashi et Takeda, 2002] Funahashi, S. et Takeda, K. (2002). Information processes in the primate prefrontal cortex in relation to working memory processes. *Review Neuroscience*, 13(4) :313–45.
- [Fuster, 1997] Fuster, J. (1997). *The Prefrontal Cortex : Anatomy, Physiology, and neuropsychology of the frontal lobe*. Lippincott Williams & Wilkins ; 3rd edition.
- [Gallese et al., 1994] Gallese, V., Murata, A., Kaseda, M., Niki, N., et Sakata, H. (1994). Deficit of hand preshaping after muscimol injection in monkey parietal cortex. *Neuroreport*, 5(12) :1525–9.
- [Galletti et Battaglini, 1989] Galletti, C. et Battaglini, P. (1989). Gaze-dependent visual neurons in area v3a of monkey prestriate cortex. *Journal of Neuroscience*, 9(4) :1112–25.
- [Gattass et al., 1988] Gattass, R., Sousa, A., et Gross, C. (1988). Visuotopic organization and extent of v3 and v4 of the macaque. *Journal of Neuroscience*, 8(6) :1831–45.
- [Gaymard et al., 2003] Gaymard, B., Lynch, J., Ploner, C., Condy, C., et Rivaud-Péchéux, S. (2003). The parieto-collicular pathway : anatomical location and contribution to saccade generation. *European Journal of Neuroscience*, 17(7) :1518–26.
- [Gegenfurtner et al., 1997] Gegenfurtner, K., Kiper, D., et Levitt, J. (1997). Functional properties of neurons in macaque area v3. *Journal of Neurophysiology*, 77(4) :1906–23.
- [Georgopoulos et al., 1982] Georgopoulos, A., Kalaska, J., Caminiti, R., et Massey, J. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, 2(11) :1527–37.
- [Gerstner et Kistler, 2002] Gerstner, W. et Kistler, W. (2002). *Spiking Neuron Models*. Cambridge University Press.
- [Ghosh et Nag, 2000] Ghosh, J. et Nag, A. (2000). An overview of radial basis function networks. *Radial Basis Function Neural Network Theory and Applications*.
- [Girard, 2003] Girard, B. (2003). *Intégration de la navigation et de la sélection de l’action dans une architecture de contrôle inspirée des ganglions de la base*. PhD thesis, Université Paris 6.
- [Girau, 2007] Girau, B. (2007). *Du parallélisme connexionniste à une pratique de calcul distribué numérique bio-inspiré*. PhD thesis, Nancy Université.
- [Goldberg, 2000] Goldberg, M. (2000). The control of gaze. In *Principles of Neural Science*, number Chap 39, pages 782–800. McGraw-Hill.
- [Goldman-Rakic, 1987] Goldman-Rakic, P. (1987). Circuitry of primate prefrontal cortex and the regulation of behavior by representational memory. *Handbook of Physiology*, 5(1) :373–417.
- [Golub et VanLoan, 1996] Golub, G. et VanLoan, C. (1996). *Matrix computations*. The Johns Hopkins University Press.

- [Goodale et Milner, 1992] Goodale, M. et Milner, D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1) :20–25.
- [Gottlieb, 2007] Gottlieb, J. (2007). From thought to action : the parietal cortex as a bridge between perception, action, and cognition. *Neuron*, 53(1) :9–16.
- [Gottlieb et al., 1998] Gottlieb, J., Kusunoki, M., et Goldberg, M. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391(6666) :481–4.
- [Gray et Wedderburn, 1960] Gray, J. et Wedderburn, A. (1960). Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 12 :180–184.
- [Graziano, 2006] Graziano, M. (2006). Progress in understanding spatial coordinate systems in the primate brain. *Neuron*, 51(1) :7–9.
- [Graziano et Gross, 1998] Graziano, M. et Gross, C. (1998). Visual responses with and without fixation : neurons in premotor cortex encode spatial locations independently of eye position. *Experimental Brain Research*, 118(3) :373–80.
- [Graziano et al., 1997] Graziano, M., Hu, X., et Gross, C. (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology*, 77(5) :2268–92.
- [Griffith, 1963] Griffith, J. (1963). A field theory of neural nets : I. derivation of field equations. *Bull Math Biophys*, 25 :111–20.
- [Griffith, 1965] Griffith, J. (1965). A field theory of neural nets. ii. properties of the field equations. *Bull Math Biophys*, 27(2) :187–95.
- [Gross, 1973] Gross, C. (1973). Visual functions of inferotemporal cortex. In *Handbook of Sensory Physiology*, pages 451–482. Springer-Verlag.
- [Gross, 1994] Gross, C. (1994). How inferior temporal cortex became a visual area. *Cerebral Cortex*, 4(5) :455–69.
- [Gross, 2002] Gross, C. (2002). Genealogy of the “grandmother cell”. *Neuroscientist*, 8(5) :512–8.
- [Grossberg, 1987] Grossberg, S. (1987). Competitive learning : From interactive activation to adaptive resonance. *Cognitive Science*, 11 :23–63.
- [Hallett, 1978] Hallett, P. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, 18(10) :1279–96.
- [Hallett et Lightstone, 1976] Hallett, P. et Lightstone, A. (1976). Saccadic eye movements to flashed targets. *Vision Research*, 16(1) :107–14.
- [Hamker, 2004] Hamker, F. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research*, 44(5) :501–21.
- [Hamker, 2005a] Hamker, F. (2005a). The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision. *Computer Vision and Image Understanding*, 100 :64–106.
- [Hamker, 2005b] Hamker, F. (2005b). Modeling attention : From computational neuroscience to computer vision. In *WAPCV 2004 - LNCS 3368*, pages 118–132. Springer-Verlag.
- [Hamker, 2005c] Hamker, F. (2005c). The reentry hypothesis : the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas v4, it for attention and eye movement. *Cerebral Cortex*, 15(4) :431–47.
- [Hanes et al., 1998] Hanes, D., Patterson, W., et Schall, J. (1998). Role of frontal eye fields in countermanding saccades : visual, movement, and fixation activity. *Journal of Neurophysiology*, 79(2) :817–34.

-
- [Hanes et Wurtz, 2001] Hanes, D. et Wurtz, R. (2001). Interaction of the frontal eye field and superior colliculus for saccade generation. *Journal of Neurophysiology*, 85(2) :804–15.
- [Hebb, 1949] Hebb, D. (1949). *The Organization of Behavior : A Neuropsychological Theory*. Wiley, New York.
- [Hegde et VanEssen, 2000] Hegde, J. et VanEssen, D. (2000). Selectivity for complex shapes in primate visual area v2. *Journal of Neuroscience*, 20(5) :RC61.
- [Heinke et Humphreys, 2005] Heinke, D. et Humphreys, G. (2005). Selective attention for identification model : Simulating visual neglect. *Computer Vision and Image Understanding*, 100 :172–197.
- [Hikosaka, 2007] Hikosaka, O. (2007). Basal ganglia mechanisms of reward-oriented eye movement. *Annals New York Academy Sciences*, 1104 :229–49.
- [Hikosaka et al., 2006] Hikosaka, O., Nakamura, K., et Nakahara, H. (2006). Basal ganglia orient eyes to reward. *Journal of Neurophysiology*, 95(2) :567–84.
- [Hikosaka et al., 1993] Hikosaka, O., Sakamoto, M., et Miyashita, N. (1993). Effects of caudate nucleus stimulation on substantia nigra cell activity in monkey. *Experimental Brain Research*, 95(3) :457–72.
- [Hikosaka et al., 2000] Hikosaka, O., Takikawa, Y., et Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological Review*, 80(3) :953–78.
- [Hikosaka et Wurtz, 1983a] Hikosaka, O. et Wurtz, R. (1983a). Visual and oculomotor functions of monkey substantia nigra pars reticulata. i. relation of visual and auditory responses to saccades. *Journal of Neurophysiology*, 49(5) :1230–53.
- [Hikosaka et Wurtz, 1983b] Hikosaka, O. et Wurtz, R. (1983b). Visual and oculomotor functions of monkey substantia nigra pars reticulata. ii. visual responses related to fixation of gaze. *Journal of Neurophysiology*, 49(5) :1254–67.
- [Hikosaka et Wurtz, 1983c] Hikosaka, O. et Wurtz, R. (1983c). Visual and oculomotor functions of monkey substantia nigra pars reticulata. iii. memory-contingent visual and saccade responses. *Journal of Neurophysiology*, 49(5) :1268–84.
- [Hikosaka et Wurtz, 1983d] Hikosaka, O. et Wurtz, R. (1983d). Visual and oculomotor functions of monkey substantia nigra pars reticulata. iv. relation of substantia nigra to superior colliculus. *Journal of Neurophysiology*, 49(5) :1285–301.
- [Hikosaka et Wurtz, 1985a] Hikosaka, O. et Wurtz, R. (1985a). Modification of saccadic eye movements by gaba-related substances. i. effect of muscimol and bicuculline in monkey superior colliculus. *Journal of Neurophysiology*, 53(1) :266–91.
- [Hikosaka et Wurtz, 1985b] Hikosaka, O. et Wurtz, R. (1985b). Modification of saccadic eye movements by gaba-related substances. ii. effects of muscimol in monkey substantia nigra pars reticulata. *Journal of Neurophysiology*, 53(1) :292–308.
- [Hoffman et Subramaniam, 1995] Hoffman, J. et Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics*, 57(6) :787–95.
- [Hopfield, 1982] Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A*, 79(8) :2554–8.
- [Horn, 2005] Horn, A. (2005). The reticular formation. *Progress Brain Research*, 151 :127–155.
- [Hubel, 1988] Hubel, D. (1988). *Eye, brain and vision*. Scientific American Library.

- [Hudson et al., 1997] Hudson, P., vanDenHerik, H., et Postma, E. (1997). Scan : A scalable model of attentional selection. *Neural Network*, 10(6) :993–1015.
- [Huerta et Kaas, 1990] Huerta, M. et Kaas, J. (1990). Supplementary eye field as defined by intracortical microstimulation : connections in macaques. *Journal of Comparative Neurology*, 293 :299–330.
- [Huerta et al., 1987] Huerta, M., Krubitzer, L., et Kaas, J. (1987). Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys. ii. cortical connections. *Journal of Comparative Neurology*, 265(3) :332–61.
- [Isoda et Tanji, 2002] Isoda, M. et Tanji, J. (2002). Cellular activity in the supplementary eye field during sequential performance of multiple saccades. *Journal of Neurophysiology*, 88(6) :3541–5.
- [Ito et al., 1995] Ito, M., Tamura, H., Fujita, I., et Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, 73(1) :218–26.
- [Itti, 2000] Itti, L. (2000). *Models of Bottom-Up and Top-Down Visual Attention*. PhD thesis, California Institute of Technology.
- [Itti et Koch, 2001] Itti, L. et Koch, C. (2001). Computational modeling of visual attention. *Nature Review Neuroscience*, 2(3) :194–203.
- [Itti et al., 1998] Itti, L., Koch, C., et Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11).
- [James, 1890] James, W. (1890). *The Principles of Psychology*. New York : Holt.
- [Jeannerod, 2001] Jeannerod, M. (2001). Neural simulation of action : a unifying mechanism for motor cognition. *Neuroimage*, 14(1 Pt 2) :S103–9.
- [Jonides, 1981] Jonides, J. (1981). *Attention and performance IX*, chapter Voluntary vs. automatic control over the mind’s eye movement. Erlbaum.
- [Jouffrais et Boussaoud, 1999] Jouffrais, C. et Boussaoud, D. (1999). Neuronal activity related to eye-hand coordination in the primate premotor cortex. *Experimental Brain Research*, 128(1-2) :205–9.
- [King et Palmer, 1985] King, A. et Palmer, A. (1985). Integration of visual and auditory information in bimodal neurons in the guinea-pig superior colliculus. *Experimental Brain Research*, 60(3) :492–500.
- [King et al., 1986] King, W., Lisberger, S., et Fuchs, A. (1986). Oblique saccadic eye movements of primates. *Journal of Neurophysiology*, 56(3) :769–84.
- [Klam, 2003] Klam, F. (2003). *Etude de la représentation de l’espace dans le cortex pariétal*. PhD thesis, Université Pierre & Marie Curie.
- [Klein, 2000] Klein, R. (2000). Inhibition of return. *Trends Cognitive Sciences*, 4(4) :138–147.
- [Kleiser et al., 2004] Kleiser, R., Seitz, R., et Krekelberg, B. (2004). Neural correlates of saccadic suppression in humans. *Current Biology*, 14(5) :386–90.
- [Kobayashi et al., 2002] Kobayashi, S., Lauwereyns, J., Koizumi, M., Sakagami, M., et Hikosaka, O. (2002). Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex. *Journal of Neurophysiology*, 87(3) :1488–98.
- [Koch et Ullman, 1985] Koch, C. et Ullman, S. (1985). Shifts in selective visual attention : towards the underlying neural circuitry. *Human Neurobiology*, 4(4) :219–27.

-
- [Kohonen, 1982] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43 :59–69.
- [Kohonen, 1995] Kohonen, T. (1995). *Self-organizing maps*. Springer-Verlag.
- [Kovacs et al., 1995] Kovacs, G., Vogels, R., et Orban, G. (1995). Selectivity of macaque inferior temporal neurons for partially occluded shapes. *Journal of Neuroscience*, 15(3) :1984–97.
- [Kowler et al., 1995] Kowler, E., Anderson, E., Doshier, B., et Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, 35(13) :1897–916.
- [Kowler et Steinman, 1980] Kowler, E. et Steinman, R. (1980). Small saccades serve no useful purpose : reply to a letter by r. w. ditchburn. *Vision Research*, 20(3) :273–6.
- [Krauzlis, 2005] Krauzlis, R. (2005). The control of voluntary eye movements : new perspectives. *Neuroscientist*, 11(2) :124–37.
- [Kristjánsson et Nakayama, 2002] Kristjánsson, A. et Nakayama, K. (2002). The attentional blink in space and time. *Vision Research*, 42(17) :2039–50.
- [Kumar, 2005] Kumar, A. (2005). *Studies of saccade and vergence eye movement interactions in human subjects*. PhD thesis, Case Western Reserve University.
- [Kumar et al., 2002] Kumar, A., Han, Y., Garbutt, S., et Leigh, R. (2002). Properties of anticipatory vergence responses. *Invest Ophthalmol Vis Sci*, 43(8) :2626–32.
- [Kustov et Robinson, 1996] Kustov, A. et Robinson, D. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, 384(6604) :74–7.
- [Lanyon et Denham, 2004] Lanyon, L. et Denham, S. (2004). A model of active visual search with object-based attention guiding scan paths. *Neural Networks*, 17(5-6) :873–97.
- [Lauwereyns et al., 2002] Lauwereyns, J., Watanabe, K., Coe, B., et Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature*, 418(6896) :413–7.
- [Lawrence et al., 2005] Lawrence, B., White, R., et Snyder, L. (2005). Delay-period activity in visual, visuomotor, and movement neurons in the frontal eye field. *Journal of Neurophysiology*, 94(2) :1498–1508.
- [Lee et al., 1988] Lee, C., Rohrer, W., et Sparks, D. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, 332(6162) :357–60.
- [Leigh et Zee, 1999] Leigh, R. et Zee, D. (1999). *The Neurology of Eye Movements - 3e Ed*. FA Davis Company, Philadelphia.
- [Leigh et Zee, 2006] Leigh, R. et Zee, D. (2006). *The Neurology of Eye Movements - 4e Ed*. Oxford University Press.
- [Lewis et VanEssen, 2000] Lewis, J. et VanEssen, D. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *Journal of Comparative Neurology*, 428(1) :112–37.
- [Li et Matin, 1997] Li, W. et Matin, L. (1997). Saccadic suppression of displacement : separate influences of saccade size and of target retinal eccentricity. *Vision Research*, 37(13) :1779–97.
- [Li, 2002] Li, Z. (2002). A saliency map in primary visual cortex. *Trends Cognitive Sciences*, 6(1) :9–16.
- [Lindner et Ilg, 2006] Lindner, A. et Ilg, U. (2006). Suppression of optokinesis during smooth pursuit eye movements revisited : the role of extra-retinal information. *Vision Research*, 46(6-7) :761–7.

- [Luck et al., 1997] Luck, S., Chelazzi, L., Hillyard, S., et Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex. *Journal of Neurophysiology*, 77(1) :24–42.
- [Lynch et McLaren, 1989] Lynch, J. et McLaren, J. (1989). Deficits of visual attention and saccadic eye movements after lesions of parietooccipital cortex in monkeys. *Journal of Neurophysiology*, 61(1) :74–90.
- [Lynch et Tian, 2005] Lynch, J. et Tian, J.-R. (2005). Cortico-cortical networks and cortico-subcortical loops for the higher control of eye movements. *Progress Brain Research*, 151 :461–501.
- [Martinez-Conde et al., 2004] Martinez-Conde, S., Macknik, S., et Hubel, D. (2004). The role of fixational eye movements in visual perception. *Nature Review Neuroscience*, 5(3) :229–40.
- [Martinez-Conde et al., 2006] Martinez-Conde, S., Macknik, S., Troncoso, X., et Dyar, T. (2006). Microsaccades counteract visual fading during fixation. *Neuron*, 49(2) :297–305.
- [Martinez-Trujillo et al., 2004] Martinez-Trujillo, J., Medendorp, W., Wang, H., et Crawford, J. (2004). Frames of reference for eye-head gaze commands in primate supplementary eye fields. *Neuron*, 44(6) :1057–66.
- [Matelli et Luppino, 2001] Matelli, M. et Luppino, G. (2001). Parietofrontal circuits for action and space perception in the macaque monkey. *Neuroimage*, 14(1 Pt 2) :S27–32.
- [Maunsell et vanEssen, 1983] Maunsell, J. et vanEssen, D. (1983). The connections of the middle temporal visual area (mt) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3(12) :2563–86.
- [May, 2005] May, P. (2005). The mammalian superior colliculus : laminar structure and connections. *Progress Brain Research*, 151 :321–78.
- [Mays et Sparks, 1980] Mays, L. et Sparks, D. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *Journal of Neurophysiology*, 43(1) :207–32.
- [Mazer et Gallant, 2003] Mazer, J. et Gallant, J. (2003). Goal-related activity in v4 during free viewing visual search. evidence for a ventral stream visual salience map. *Neuron*, 40(6) :1241–50.
- [McAdams et Maunsell, 1999] McAdams, C. et Maunsell, J. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *Journal of Neuroscience*, 19(1) :431–41.
- [McCulloch et Pitts, 1943] McCulloch, W. et Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5 :115–133.
- [McLeod et al., 1988] McLeod, P., Driver, J., et Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, 332(6160) :154–5.
- [McPeck et Keller, 2004] McPeck, R. et Keller, E. (2004). Deficits in saccade target selection after inactivation of superior colliculus. *Nature Neuroscience*, 7(7) :757–63.
- [Medler, 1998] Medler, D. (1998). A brief history of connectionism. *Neural Computing Surveys*, pages 61–101.
- [Mehaffey et al., 2005] Mehaffey, W., Doiron, B., Maler, L., et Turner, R. (2005). Deterministic multiplicative gain control with active dendrites. *Journal of Neuroscience*, 25(43) :9968–77.
- [Mel, 1993] Mel, B. (1993). Synaptic integration in an excitable dendritic tree. *Journal of Neurophysiology*, 70(3) :1086–101.

-
- [Mel, 1999] Mel, B. (1999). Why have dendrites? a computational perspective. In *Dendrites*, pages 271–289. Oxford University Press.
- [Merigan et Maunsell, 1993] Merigan, W. et Maunsell, J. (1993). How parallel are the primate visual pathways? *Annual Review Neurosciences*, 16 :369–402.
- [Merriam et Colby, 2005] Merriam, E. et Colby, C. (2005). Active vision in parietal and extrastriate cortex. *Neuroscientist*, 11(5) :484–93.
- [Merriam et al., 2007] Merriam, E., Genovese, C., et Colby, C. (2007). Remapping in human visual cortex. *Journal of Neurophysiology*, 97(2) :1738–55.
- [Middleton et Strick, 2002] Middleton, F. et Strick, P. (2002). Basal-ganglia 'projections' to the prefrontal cortex of the primate. *Cerebral Cortex*, 12(9) :926–35.
- [Milanese, 1993] Milanese, R. (1993). *Detecting Salient Regions in an Image : From Biological Evidence to Computer Implementation*. PhD thesis, Univ. of Geneva.
- [Miles, 1998] Miles, F. (1998). The neural processing of 3-d visual information : evidence from eye movements. *European Journal of Neuroscience*, 10(3) :811–22.
- [Milner et Goodale, 1995] Milner, D. et Goodale, M. (1995). *The visual brain in action*. Oxford University Press.
- [Minsky et Papert, 1969] Minsky, M. et Papert, S. (1969). *Perceptrons : An introduction to Computational Geometry*. MIT Press.
- [Miyashita, 1993] Miyashita, Y. (1993). Inferior temporal cortex : where visual perception meets memory. *Annual Review Neurosciences*, 16 :245–63.
- [Monakow et al., 1978] Monakow, K., Akert, K., et Künzle, H. (1978). Projections of the precentral motor cortex and other cortical areas of the frontal lobe to the subthalamic nucleus in the monkey. *Experimental Brain Research*, 33(3-4) :395–403.
- [Montagnini et Chelazzi, 2005] Montagnini, A. et Chelazzi, L. (2005). The urgency to look : prompt saccades to the benefit of perception. *Vision Research*, 45(27) :3391–401.
- [Moody et Darken, 1989] Moody, J. et Darken, C. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1 :281–294.
- [Moore et Armstrong, 2003] Moore, T. et Armstrong, K. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421(6921) :370–3.
- [Moore et Fallah, 2001] Moore, T. et Fallah, M. (2001). Control of eye movements and spatial attention. *Proc Natl Acad Sci U S A*, 98(3) :1273–6.
- [Moore et Fallah, 2004] Moore, T. et Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of Neurophysiology*, 91(1) :152–62.
- [Moore et al., 1998] Moore, T., Tolias, A., et Schiller, P. (1998). Visual representations during saccadic eye movements. *Proc Natl Acad Sci U S A*, 95(15) :8981–4.
- [Moran et Desimone, 1985] Moran, J. et Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715) :782–4.
- [Moschovakis et al., 1996] Moschovakis, A., Scudder, C., et Highstein, S. (1996). The microscopic anatomy and physiology of the mammalian saccadic system. *Progress in Neurobiology*, 50(2-3) :133–254.
- [Motter, 1994a] Motter, B. (1994a). Neural correlates of attentive selection for color or luminance in extrastriate area v4. *Journal of Neuroscience*, 14(4) :2178–89.

- [Motter, 1994b] Motter, B. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area v4. *Journal of Neuroscience*, 14(4) :2190–9.
- [Mountcastle, 1955] Mountcastle, V. (1955). Topographic organization and modality representation in first somatic area of cat's cerebral cortex by method of single unit analysis. *American Journal of Physiology*, 183 :464.
- [Mountcastle, 1957] Mountcastle, V. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, 20(4) :408–34.
- [Mozer et Sitton, 1998] Mozer, M. et Sitton, M. (1998). Computational modeling of spatial attention. In *Attention*, pages 341–396. Psychology Press, East Sussex, UK.
- [Müller, 1843] Müller, J. (1843). *Elements of Physiology*. Taylor & Walton, London.
- [Müller et al., 2005] Müller, J., Philiastides, M., et Newsome, W. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proc Natl Acad Sci U S A*, 102(3) :524–9.
- [Munoz et al., 1991] Munoz, D., Pélisson, D., et Guitton, D. (1991). Movement of neural activity on the superior colliculus motor map during gaze shifts. *Science*, 251(4999) :1358–60.
- [Munoz et Wurtz, 1995a] Munoz, D. et Wurtz, R. (1995a). Saccade-related activity in monkey superior colliculus. i. characteristics of burst and buildup cells. *Journal of Neurophysiology*, 73(6) :2313–33.
- [Munoz et Wurtz, 1995b] Munoz, D. et Wurtz, R. (1995b). Saccade-related activity in monkey superior colliculus. ii. spread of activity during saccades. *Journal of Neurophysiology*, 73(6) :2334–48.
- [Müri et al., 1996] Müri, R., Vermersch, A., Rivaud, S., Gaymard, B., et Pierrot-Deseilligny, C. (1996). Effects of single-pulse transcranial magnetic stimulation over the prefrontal and posterior parietal cortices during memory-guided saccades in humans. *Journal of Neurophysiology*, 76(3) :2102–6.
- [Nakamura et Colby, 2002] Nakamura, K. et Colby, C. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proc Natl Acad Sci U S A*, 99(6) :4026–31.
- [Nambu et al., 1996] Nambu, A., Takada, M., Inase, M., et Tokuno, H. (1996). Dual somatotopical representations in the primate subthalamic nucleus : evidence for ordered but reversed body-map transformations from the primary motor cortex and the supplementary motor area. *Journal of Neuroscience*, 16(8) :2671–83.
- [Navalpakkam et Itti, 2005] Navalpakkam, V. et Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45(2) :205–31.
- [Navalpakkam et Itti, 2006] Navalpakkam, V. et Itti, L. (2006). Top-down attention selection is fine grained. *Journal of Vision*, 6(11) :1180–93.
- [Noe, 2005] Noe, A. (2005). *Action in Perception*. MIT Press.
- [Oja, 1982] Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. Math. Biol.*, 15(3) :267–73.
- [Olshausen et al., 1993] Olshausen, B., Anderson, C., et VanEssen, D. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11) :4700–19.
- [O'Regan et Noë, 2001] O'Regan, J. et Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral Brain Sciences*, 24(5) :939–73 ; discussion 973–1031.

-
- [Park et al., 2006] Park, J., Schlag-Rey, M., et Schlag, J. (2006). Frames of reference for saccadic command tested by saccade collision in the supplementary eye field. *Journal of Neurophysiology*, 95(1) :159–70.
- [Parkhurst et al., 2002] Parkhurst, D., Law, K., et Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1) :107–23.
- [Parthasarathy et al., 1992] Parthasarathy, H., Schall, J., et Graybiel, A. (1992). Distributed but convergent ordering of corticostriatal projections : analysis of the frontal eye field and the supplementary eye field in the macaque monkey. *Journal of Neuroscience*, 12(11) :4468–88.
- [Penney et Young, 1983] Penney, J. et Young, A. (1983). Speculations on the functional anatomy of basal ganglia disorders. *Annual Review Neurosciences*, 6 :73–94.
- [Petrides, 1994] Petrides, M. (1994). Frontal lobes and working memory : evidence from investigations of the effects of cortical excisions in nonhuman primates. *Handbook of Neuropsychology*, 9 :59–82.
- [Pierrot-Deseilligny et al., 2004] Pierrot-Deseilligny, C., Milea, D., et Müri, R. (2004). Eye movement control by the cerebral cortex. *Current Opinion in Neurology*, 17(1) :17–25.
- [Pierrot-Deseilligny et al., 2003] Pierrot-Deseilligny, C., Müri, R., Ploner, C., Gaymard, B., Demeret, S., et Rivaud-Pechoux, S. (2003). Decisional role of the dorsolateral prefrontal cortex in ocular motor behaviour. *Brain*, 126(Pt 6) :1460–73.
- [Pierrot-Deseilligny et al., 2002] Pierrot-Deseilligny, C., Ploner, C., Muri, R., Gaymard, B., et Rivaud-Pechoux, S. (2002). Effects of cortical lesions on saccadic : eye movements in humans. *Ann N Y Acad Sci*, 956 :216–29.
- [Pierrot-Deseilligny et al., 1991a] Pierrot-Deseilligny, C., Rivaud-Pechoux, S., Gaymard, B., et Agid, Y. (1991a). Cortical control of memory-guided saccades in man. *Experimental Brain Research*, 83(3) :607–17.
- [Pierrot-Deseilligny et al., 1991b] Pierrot-Deseilligny, C., Rivaud-Pechoux, S., Gaymard, B., et Agid, Y. (1991b). Cortical control of reflexive visually-guided saccades. *Brain*, 114 (Pt 3) :1473–85.
- [Pinto et Ermentrout, 2001] Pinto, D. et Ermentrout, B. (2001). Spatially structured activity in synaptically coupled neuronal networks : I. traveling fronts and pulses. *SIAM J. Appl. Math.*, 62(1) :206–225.
- [Ploner et al., 1999] Ploner, C., Rivaud-Péchoux, S., Gaymard, B., Agid, Y., et Pierrot-Deseilligny, C. (1999). Errors of memory-guided saccades in humans with lesions of the frontal eye field and the dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 82(2) :1086–90.
- [Posner et Cohen, 1984] Posner, M. et Cohen, Y. (1984). *Attention and performance X*, chapter Components of visual orienting, pages 531–556. Lawrence Erlbaum Associates.
- [Posner et Gilbert, 1999] Posner, M. et Gilbert, C. (1999). Attention and primary visual cortex. *Proc Natl Acad Sci U S A*, 96(6) :2585–7.
- [Posner et Petersen, 1990] Posner, M. et Petersen, S. (1990). The attention system of the human brain. *Annual Review Neurosciences*, 13 :25–42.
- [Posner et al., 1980] Posner, M., Snyder, C., et Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, 109(2) :160–74.
- [Pouget et al., 2000] Pouget, A., Dayan, P., et Zemel, R. (2000). Information processing with population codes. *Nature Review Neuroscience*, 1(2) :125–32.

- [Pouget et al., 2002] Pouget, A., Deneve, S., et Duhamel, J. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Review Neuroscience*, 3(9) :741–7.
- [Pouget et Sejnowski, 1997] Pouget, A. et Sejnowski, T. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, 9 :222–237.
- [Pouget et Sejnowski, 2001] Pouget, A. et Sejnowski, T. (2001). Simulating a lesion in a basis function model of spatial representations : comparison with hemineglect. *Psychological Review*, 108(3) :653–73.
- [Pouget et Sejnowski, 2002] Pouget, A. et Sejnowski, T. (2002). Dynamic remapping. In *The Handbook of Brain Theory and Neural Networks, 2nd Edition*, pages 368–371. Boston MIT Press.
- [Prechtl et al., 1997] Prechtl, J., Cohen, L., Pesaran, B., Mitra, P., et Kleinfeld, D. (1997). Visual stimuli induce waves of electrical activity in turtle cortex. *Proc Natl Acad Sci U S A*, 94(14) :7621–6.
- [Press et al., 2007] Press, W., Teukolsky, S., Vetterling, W., et Flannery, B. (2007). *Numerical Recipes, Third edition*. Cambridge University Press.
- [Purves et al., 2005] Purves, D., Augusting, G., Fitzpatrick, D., Hall, W., LaManta, A., et McNamara, J. (2005). *Neurosciences*. De Boeck & Larcier.
- [Rafal et al., 1989] Rafal, R., Calabresi, P., Brennan, C., et Sciolto, T. (1989). Saccade preparation inhibits reorienting to recently attended locations. *J Exp Psychol Hum Percept Perform*, 15(4) :673–85.
- [Raymond et al., 1992] Raymond, J., Shapiro, K., et Arnell, K. (1992). Temporary suppression of visual processing in an rsvp task : an attentional blink? . *J Exp Psychol Hum Percept Perform*, 18(3) :849–60.
- [Redgrave, 2007] Redgrave, P. (2007). Basal ganglia. *Scholarpedia*, 2(6) :1825.
- [Rensink et al., 1997] Rensink, R., O’Regan, J., et Clark, J. (1997). To see or not to see : the need for attention to perceive changes in scenes. *Psychological Science*, 8(5) :368–373.
- [Reynolds et Chelazzi, 2004] Reynolds, J. et Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review Neurosciences*, 27 :611–47.
- [Reynolds et al., 1999] Reynolds, J., Chelazzi, L., et Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas v2 and v4. *Journal of Neuroscience*, 19(5) :1736–53.
- [Reynolds et Desimone, 1999] Reynolds, J. et Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, 24(1) :19–29, 111–25.
- [Reynolds et al., 2000] Reynolds, J., Pasternak, T., et Desimone, R. (2000). Attention increases sensitivity of v4 neurons. *Neuron*, 26(3) :703–14.
- [Riesenhuber et Poggio, 1999] Riesenhuber, M. et Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11) :1019–25.
- [Rizzolatti et al., 1998] Rizzolatti, G., Luppino, G., et Matelli, M. (1998). The organization of the cortical motor system : new concepts. *Electroencephalography and Clinical Neurophysiology*, 106(4) :283–96.
- [Rizzolatti et al., 1987] Rizzolatti, G., Riggio, L., Dascola, I., et Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians : evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25(1A) :31–40.

-
- [Robinson, 1972] Robinson, D. (1972). Eye movements evoked by collicular stimulation in the alert monkey. *Vision Research*, 12(11) :1795–808.
- [Robinson et Petersen, 1992] Robinson, D. et Petersen, S. (1992). The pulvinar and visual salience. *Trends Neuroscience*, 15(4) :127–32.
- [Rochel, 2004] Rochel, O. (2004). *Une approche événementielle pour la modélisation et la simulation de réseaux de neurones impulsifs*. PhD thesis, Université Henri Poincaré, Nancy 1.
- [Rockland et VanHoesen, 1994] Rockland, K. et VanHoesen, G. (1994). Direct temporal-occipital feedback connections to striate cortex (v1) in the macaque monkey. *Cerebral Cortex*, 4(3) :300–13.
- [Rolls, 1991] Rolls, E. (1991). Neural organization of higher visual functions. *Current Opinion Neurobiology*, 1(2) :274–8.
- [Rolls, 2000] Rolls, E. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron*, 27(2) :205–18.
- [Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6) :386–408.
- [Rosenzweig et al., 1998] Rosenzweig, M., Leiman, A., Breedlove, S., Bonaventure, N., et Will, B. (1998). *Psychobiologie*. De Boeck.
- [Ross et al., 2001] Ross, J., Morrone, M., Goldberg, M., et Burr, D. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, 24(2) :113–21.
- [Rougier, 2000] Rougier, N. (2000). *Modèles de mémoires pour la navigation autonome*. PhD thesis, Université Henri Poincaré, Nancy 1.
- [Rougier et Hutt, 2008] Rougier, N. et Hutt, A. (2008). Synchronous and asynchronous evaluation of dynamic neural fields. *Journal of Difference Equations and Applications*.
- [Rougier et Vitay, 2006] Rougier, N. et Vitay, J. (2006). Emergence of attention within a neural population. *Neural Network*, 19(5) :573–81.
- [Rumelhart et al., 1987] Rumelhart, D., Hinton, G., et McClelland, J. (1987). *Parallel Distributed Processing, Vol. 1*, chapter A general framework for parallel distributed processing. MIT Press.
- [Sakata et al., 1997] Sakata, H., Taira, M., Kusunoki, M., Murata, A., et Tanaka, Y. (1997). The tins lecture. the parietal association cortex in depth perception and visual control of hand action. *Trends Neuroscience*, 20(8) :350–7.
- [Sakata et al., 1995] Sakata, H., Taira, M., Murata, A., et Mine, S. (1995). Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cerebral Cortex*, 5(5) :429–38.
- [Salinas et Abbott, 1994] Salinas, E. et Abbott, L. (1994). Vector reconstruction from firing rates. *Journal of Computational Neuroscience*, 1(1-2) :89–107.
- [Salinas et Abbott, 1995] Salinas, E. et Abbott, L. (1995). Transfer of coded information from sensory to motor networks. *Journal of Neuroscience*, 15(10) :6461–74.
- [Salinas et Abbott, 1996] Salinas, E. et Abbott, L. (1996). A model of multiplicative neural responses in parietal cortex. *Proc Natl Acad Sci U S A*, 93(21) :11956–61.
- [Salinas et Sejnowski, 2001] Salinas, E. et Sejnowski, T. (2001). Gain modulation in the central nervous system : where behavior, neurophysiology, and computation meet. *Neuroscientist*, 7(5) :430–40.

- [Salinas et Thier, 2000] Salinas, E. et Thier, P. (2000). Gain modulation : a major computational principle of the central nervous system. *Neuron*, 27(1) :15–21.
- [Schall, 1995] Schall, J. (1995). Neural basis of saccade target selection. *Review Neuroscience*, 6(1) :63–85.
- [Schall et al., 2000] Schall, J., Hanes, D., et Taylor, T. (2000). Neural control of behavior : countermanding eye movements. *Psychological Research*, 63(3-4) :299–307.
- [Schall et al., 1993] Schall, J., Morel, A., et Kaas, J. (1993). Topography of supplementary eye field afferents to frontal eye field in macaque : implications for mapping between saccade coordinate systems. *Vision Neuroscience*, 10(2) :385–93.
- [Scherrer, 2003] Scherrer, B. (2003). *Apprentissage de représentation et auto-organisation modulaire pour un agent autonome*. PhD thesis, Université Henri Poincaré, Nancy 1.
- [Schiller et Chou, 1998] Schiller, P. et Chou, I. (1998). The effects of frontal eye field and dorsomedial frontal cortex lesions on visually guided eye movements. *Nature Neuroscience*, 1(3) :248–53.
- [Schiller et Tehovnik, 2005] Schiller, P. et Tehovnik, E. (2005). Neural mechanisms underlying target selection with saccadic eye movements. *Progress Brain Research*, 149 :157–71.
- [Schiller et al., 1979] Schiller, P., True, S., et Conway, J. (1979). Effects of frontal eye field and superior colliculus ablations on eye movements. *Science*, 206(4418) :590–2.
- [Schlag et Schlag-Rey, 1985] Schlag, J. et Schlag-Rey, M. (1985). Unit activity related to spontaneous saccades in frontal dorsomedial cortex of monkey. *Experimental Brain Research*, 58(1) :208–11.
- [Schlag et Schlag-Rey, 1987] Schlag, J. et Schlag-Rey, M. (1987). Evidence for a supplementary eye field. *Journal of Neurophysiology*, 57(1) :179–200.
- [Schlag et al., 1992] Schlag, J., Schlag-Rey, M., et Pigarev, I. (1992). Supplementary eye field : influence of eye position on neural signals of fixation. *Experimental Brain Research*, 90(2) :302–6.
- [Schlag-Rey et al., 1992] Schlag-Rey, M., Schlag, J., et Dassonville, P. (1992). How the frontal eye field can impose a saccade goal on superior colliculus neurons. *Journal of Neurophysiology*, 67(4) :1003–5.
- [Schwabe et al., 2006] Schwabe, L., Obermayer, K., Angelucci, A., et Bressloff, P. (2006). The role of feedback in shaping the extra-classical receptive field of cortical neurons : a recurrent network model. *Journal of Neuroscience*, 26(36) :9117–29.
- [Segraves, 1992] Segraves, M. (1992). Activity of monkey frontal eye field neurons projecting to oculomotor regions of the pons. *Journal of Neurophysiology*, 68(6) :1967–85.
- [Segraves et Goldberg, 1987] Segraves, M. et Goldberg, M. (1987). Functional properties of corticotectal neurons in the monkey's frontal eye field. *Journal of Neurophysiology*, 58(6) :1387–419.
- [Selemon et Goldman-Rakic, 1985] Selemon, L. et Goldman-Rakic, P. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *Journal of Neuroscience*, 5(3) :776–94.
- [Selemon et Goldman-Rakic, 1988] Selemon, L. et Goldman-Rakic, P. (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey : evidence for a distributed neural network subserving spatially guided behavior. *Journal of Neuroscience*, 8(11) :4049–68.

-
- [Serre et al., 2007] Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., et Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress Brain Research*, 165 :33–56.
- [Sherrington, 1894] Sherrington, C. (1894). Experimental note on two movements of the eye. *Journal of Physiology - London*, 17(1-2) :27–9.
- [Shook et al., 1991] Shook, B., Schlag-Rey, M., et Schlag, J. (1991). Primate supplementary eye field. ii. comparative aspects of connections with the thalamus, corpus striatum, and related forebrain nuclei. *Journal of Comparative Neurology*, 307(4) :562–83.
- [Silver et al., 2007] Silver, M., Ress, D., et Heeger, D. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *Journal of Neurophysiology*, 97(1) :229–37.
- [Simons et Chabris, 1999] Simons, D. et Chabris, C. (1999). Gorillas in our midst : sustained inattentive blindness for dynamic events. *Perception*, 28(9) :1059–74.
- [Simons, 2000] Simons, J. (2000). Current approaches to change blindness. *Visual Cognition*, 7(1-2-3) :1–15.
- [Smit et al., 1990] Smit, A., VanOpstal, A., et VanGisbergen, J. (1990). Component stretching in fast and slow oblique saccades in the human. *Experimental Brain Research*, 81(2) :325–34.
- [Soetedjo et al., 2002] Soetedjo, R., Kaneko, C., et Fuchs, A. (2002). Evidence against a moving hill in the superior colliculus during saccadic eye movements in the monkey. *Journal of Neurophysiology*, 87(6) :2778–89.
- [Sommer et Wurtz, 2000] Sommer, M. et Wurtz, R. (2000). Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus. *Journal of Neurophysiology*, 83(4) :1979–2001.
- [Sommer et Wurtz, 2004a] Sommer, M. et Wurtz, R. (2004a). What the brain stem tells the frontal cortex. i. oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. *Journal of Neurophysiology*, 91(3) :1381–402.
- [Sommer et Wurtz, 2004b] Sommer, M. et Wurtz, R. (2004b). What the brain stem tells the frontal cortex. ii. role of the sc-md-fef pathway in corollary discharge. *Journal of Neurophysiology*, 91(3) :1403–23.
- [Sommer et Wurtz, 2006] Sommer, M. et Wurtz, R. (2006). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature*, 444(7117) :374–7.
- [Sparks, 1978] Sparks, D. (1978). Functional properties of neurons in the monkey superior colliculus : coupling of neuronal activity and saccade onset. *Brain Research*, 156(1) :1–16.
- [Sparks, 1993] Sparks, D. (1993). Are gaze shifts controlled by a 'moving hill' of activity in the superior colliculus? *Trends in Neuroscience*, 16(6) :214–218.
- [Sparks, 2002] Sparks, D. (2002). The brainstem control of saccadic eye movements. *Nature Review Neuroscience*, 3(12) :952–64.
- [Sparks, 2004] Sparks, D. (2004). Commands for eye and head movements in the primate superior colliculus. In *The superior colliculus : new approaches for studying sensorimotor integration*, pages 303–318.
- [Sparks et al., 1976] Sparks, D., Holland, R., et Guthrie, B. (1976). Size and distribution of movement fields in the monkey superior colliculus. *Brain Research*, 113(1) :21–34.
- [Sperry, 1950] Sperry, R. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol*, 43(6) :482–9.

- [Stanton et al., 1995] Stanton, G., Bruce, C., et Goldberg, M. (1995). Topography of projections to posterior cortical areas from the macaque frontal eye fields. *Journal of Comparative Neurology*, 353(2) :291–305.
- [Stanton et al., 1988a] Stanton, G., Goldberg, M., et Bruce, C. (1988a). Frontal eye field efferents in the macaque monkey : I. subcortical pathways and topography of striatal and thalamic terminal fields. *Journal of Comparative Neurology*, 271(4) :473–92.
- [Stanton et al., 1988b] Stanton, G., Goldberg, M., et Bruce, C. (1988b). Frontal eye field efferents in the macaque monkey : II. topography of terminal fields in midbrain and pons. *Journal of Comparative Neurology*, 271(4) :493–506.
- [Stuphorn et al., 2000] Stuphorn, V., Taylor, T., et Schall, J. (2000). Performance monitoring by the supplementary eye field. *Nature*, 408(6814) :857–60.
- [Sutton et Barto, 1998] Sutton, R. et Barto, A. (1998). *Reinforcement learning : An Introduction*. MIT Press, Cambridge MA.
- [Tanaka, 1996] Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Reviews Neuroscience*, 19 :109–39.
- [Taylor, 1999] Taylor, J. (1999). Neural bubble dynamics in two dimensions. *Biological Cybernetics*, 80 :5167–5174.
- [Taylor et al., 2006] Taylor, N., Hartley, M., et Taylor, J. (2006). The micro-structure of attention. *Neural Network*, 19(9) :1347–70.
- [Tehovnik et al., 2000] Tehovnik, E., Sommer, M., Chou, I., Slocum, W., et Schiller, P. (2000). Eye fields in the frontal lobes of primates. *Brain Res Brain Res Rev*, 32(2-3) :413–48.
- [Thompson et Bichot, 2005] Thompson, K. et Bichot, N. (2005). A visual salience map in the primate frontal eye field. *Progress Brain Research*, 147 :251–62.
- [Tian et Lynch, 1996] Tian, J. et Lynch, J. (1996). Corticocortical input to the smooth and saccadic eye movement subregions of the frontal eye field in cebus monkeys. *Journal of Neurophysiology*, 76(4) :2754–71.
- [Tiesinga et al., 2004] Tiesinga, P., Fellous, J., Salinas, E., José, J., et Sejnowski, T. (2004). Synchronization as a mechanism for attentional gain modulation. *Neurocomputing*, 58-60 :641–646.
- [Tobler et Müri, 2002] Tobler, P. et Müri, R. (2002). Role of human frontal and supplementary eye fields in double step saccades. *Neuroreport*, 13(2) :253–5.
- [Treisman et Gelade, 1980] Treisman, A. et Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1) :97–136.
- [Treisman et Sato, 1990] Treisman, A. et Sato, S. (1990). Conjunction search revisited. *J Exp Psychol Hum Percept Perform*, 16(3) :459–78.
- [Treisman et Schmidt, 1982] Treisman, A. et Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14(1) :107–41.
- [Treue et Martínez-Trujillo, 1999] Treue, S. et Martínez-Trujillo, J. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(6736) :575–9.
- [Treue et Maunsell, 1996] Treue, S. et Maunsell, J. (1996). Attentional modulation of visual motion processing in cortical areas mt and mst. *Nature*, 382(6591) :539–41.
- [Tsotsos, 1995] Tsotsos, J. (1995). Towards a computational model of visual attention. In *Early Vision and Beyond*, pages 207–218. MIT Press, Bradford Books, Cambridge MA.

-
- [Tsotsos et al., 2001] Tsotsos, J., Culhane, S., et Cutzu, F. (2001). From foundational principles to a hierarchical selection circuit for attention. In *Visual Attention and Cortical Circuits*, pages 285–306. MIT Press.
- [Tsotsos et al., 1995] Tsotsos, J., Culhane, S., Lai, W., et Davis, N. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, 78 :507–545.
- [Tsotsos et al., 2005] Tsotsos, J., Liu, Y., Martínez-Trujillo, J., Pomplun, M., Simine, E., et Zhou, K. (2005). Attending to visual motion. *Computer Vision and Image Understanding*, 100 :3–40.
- [Tsunoda et al., 2001] Tsunoda, K., Yamane, Y., Nishizaki, M., et Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, 4(8) :832–8.
- [Umeno et Goldberg, 1997] Umeno, M. et Goldberg, M. (1997). Spatial processing in the monkey frontal eye field. i. predictive visual responses. *Journal of Neurophysiology*, 78(3) :1373–83.
- [Umeno et Goldberg, 2001] Umeno, M. et Goldberg, M. (2001). Spatial processing in the monkey frontal eye field. ii. memory responses. *Journal of Neurophysiology*, 86(5) :2344–52.
- [Ungerleider et Mishkin, 1982] Ungerleider, L. et Mishkin, M. (1982). *Analysis of Visual Behavior*, chapter Two cortical visual systems, pages 549–586. MIT Press.
- [VanEssen et al., 1992] VanEssen, D., Anderson, C., et Felleman, D. (1992). Information processing in the primate visual system : an integrated systems perspective. *Science*, 255(5043) :419–23.
- [VanEssen et Maunsell, 1983] VanEssen, D. et Maunsell, J. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, 6 :370–375.
- [vanGisbergen et al., 1987] vanGisbergen, J., VanOpstal, A., et Tax, A. (1987). Collicular ensemble coding of saccades based on vector summation. *Neuroscience*, 21(2) :541–55.
- [vanHemmen et Schwartz, 2008] vanHemmen, J. et Schwartz, A. (2008). Population vector code : a geometric universal as actuator. *Biological Cybernetics*, 98(6) :509–18.
- [Varela, 1996] Varela, F. (1996). *Invitation aux sciences cognitives*. Points. Editions du Seuil, editions du seuil edition.
- [Verheijen, 1961] Verheijen, F. (1961). A simple after image method demonstrating the involuntary multidirectional eye movements during fixation. *Opt Acta (Lond)*, 8 :309–11.
- [Vitay, 2006] Vitay, J. (2006). *Emergence de fonctions sensorimotrices sur un substrat neuronal numérique distribué*. PhD thesis, Université Henri Poincaré, Nancy 1.
- [Vitay et Rougier, 2005] Vitay, J. et Rougier, N. (2005). Using neural dynamics to switch attention. In *International Joint Conference on Neural Networks (IJCNN 2005)*.
- [Vogels et al., 2005] Vogels, T., Rajan, K., et Abbott, L. (2005). Neural network dynamics. *Annual Review Neurosciences*, 28 :357–76.
- [vonHolst et Mittelstaedt, 1950] vonHolst, E. et Mittelstaedt, H. (1950). Das reafferenzprinzip. *Naturwissenschaften*, 37 :464–476.
- [Walker et al., 1995] Walker, M., Fitzgibbon, E., et Goldberg, M. (1995). Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *Journal of Neurophysiology*, 73(5) :1988–2003.
- [Wallman et Pettigrew, 1985] Wallman, J. et Pettigrew, J. (1985). Conjugate and disjunctive saccades in two avian species with contrasting oculomotor strategies. *Journal of Neuroscience*, 5(6) :1418–28.

- [Walther et al., 2002] Walther, D., Itti, L., Riesenhuber, M., Poggio, T., et Koch, C. (2002). Attentional selection for object recognition - a gentle way. In *Lecture Notes in Computer Science*, volume 2525, pages 472–479.
- [Wang et al., 2007] Wang, X., Zhang, M., Cohen, I., et Goldberg, M. (2007). The proprioceptive representation of eye position in monkey primary somatosensory cortex. *Nature Neuroscience*, 10(5) :640–6.
- [Ward et Frackowiak, 2006] Ward, N. et Frackowiak, R. (2006). The functional anatomy of cerebral reorganisation after focal brain injury. *Journal of Physiology-Paris*, 99(4-6) :425–36.
- [Ward et al., 1996] Ward, R., Duncan, J., et Shapiro, R. (1996). The slow time-course of visual attention. *Cognitive Psychology*, 30(1) :79–109.
- [Wardak et al., 2006] Wardak, C., Ibos, G., Duhamel, J., et Olivier, E. (2006). Contribution of the monkey frontal eye field to covert visual attention. *Journal of Neuroscience*, 26(16) :4228–35.
- [Williford et Maunsell, 2006] Williford, T. et Maunsell, J. (2006). Effects of spatial attention on contrast response functions in macaque area v4. *Journal of Neurophysiology*, 96(1) :40–54.
- [Wilson et Cowan, 1972] Wilson, H. et Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J*, 12(1) :1–24.
- [Wilson et Cowan, 1973] Wilson, H. et Cowan, J. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2) :55–80.
- [Winer et Schreiner, 2005] Winer, J. et Schreiner, C. (2005). *The Inferior Colliculus*. Springer.
- [Wolfe, 1994] Wolfe, J. (1994). Guided search 2.0 : A revised model of visual search. *Psychonomic Bulletin Review*, 1(2) :202–238.
- [Wolfe, 1998] Wolfe, J. (1998). *Attention*, chapter Visual Search. University College London Press.
- [Wolfe, 2000] Wolfe, J. (2000). *Seeing, 2nd ed.*, chapter Visual Attention, pages 335–386. San Diego, CA : Academic Press.
- [Wolfe, 2001] Wolfe, J. (2001). Asymmetries in visual search : an introduction. *Perception and Psychophysics*, 63(3) :381–9.
- [Wolfe, 2005] Wolfe, J. (2005). Guidance of visual search by preattentive information. In *Neurobiology of Attention*, pages 101–104. Elsevier.
- [Wolfe et al., 1989] Wolfe, J., Cave, K., et Franzel, S. (1989). Guided search : an alternative to the feature integration model for visual search. *J Exp Psychol Hum Percept Perform*, 15(3) :419–33.
- [Wolfe et Gancarz, 1996] Wolfe, J. et Gancarz, G. (1996). Guided search 3.0 : A mode of visual search catches up with jay enoch 40 years later. In *Basic and Clinical Applications of Vision Science*, pages 189–192. Dordrecht, Netherlands : Kluwer Academic.
- [Wolfram, 1984] Wolfram, S. (1984). University and complexity in cellular automata. *Physica D*, 10 :1–35.
- [Wolfram, 2002] Wolfram, S. (2002). *A New Kind of Science*. Champaign, IL : Wolfram Media.
- [Woodman et Luck, 2003] Woodman, G. et Luck, S. (2003). Serial deployment of attention during visual search. *J Exp Psychol Hum Percept Perform*, 29(1) :121–38.
- [Wurtz, 2008] Wurtz, R. (2008). Neuronal mechanisms of visual stability. *Vision Research*.

-
- [Xing et Andersen, 2000] Xing, J. et Andersen, R. (2000). Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames. *Journal of Cognitive Neuroscience*, 12(4) :601–14.
- [Yarbus, 1967] Yarbus, A. (1967). *Eye movements and vision*. Plenum Press.
- [Yeterian et Pandya, 1991] Yeterian, E. et Pandya, D. (1991). Prefrontostriatal connections in relation to cortical architectonic organization in rhesus monkeys. *Journal of Comparative Neurology*, 312(1) :43–67.
- [Zaharescu et al., 2004] Zaharescu, A., Rothenstein, A., et Tsotsos, J. (2004). Towards a biologically plausible active visual search model. In *Proc. ECCV WAPCV2004*.
- [Zambarbieri, 2002] Zambarbieri, D. (2002). The latency of saccades toward auditory targets in humans. *Progress Brain Research*, 140 :51–9.
- [Zeki, 1973] Zeki, S. (1973). Colour coding in rhesus monkey prestriate cortex. *Brain Research*, 53(2) :422–7.
- [Zeki, 1978] Zeki, S. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, 274(5670) :423–8.
- [Zhang, 1996] Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble : a theory. *Journal of Neuroscience*, 16(6) :2112–26.
- [Zhaoping, 2002] Zhaoping, L. (2002). A saliency map in primary visual cortex. *Trends Cognitive Sciences*, 6(1) :9–16.
- [Zipser et Andersen, 1988] Zipser, D. et Andersen, R. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331(6158) :679–84.
- [Zuidam et Collewyn, 1979] Zuidam, I. et Collewyn, H. (1979). Vergence eye movements of the rabbit in visuomotor behavior. *Vision Research*, 19(2) :185–94.