



Une approche neuro-dynamique de conception des processus d'auto-organisation

Lucian Alecu

► To cite this version:

Lucian Alecu. Une approche neuro-dynamique de conception des processus d'auto-organisation. Autre [cs.OH]. Université Henri Poincaré - Nancy 1, 2011. Français. NNT: 2011NAN10031 . tel-01746178v1

HAL Id: tel-01746178

<https://hal.univ-lorraine.fr/tel-01746178v1>

Submitted on 29 Mar 2018 (v1), last revised 7 Jul 2011 (v2)

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.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Une approche neuro-dynamique de conception des processus d'auto-organisation

THÈSE

présentée et soutenue publiquement le 30 juin 2011

pour l'obtention du

Doctorat de l'université Henri Poincaré – Nancy 1

(spécialité informatique)

par

Lucian Alecu

Composition du jury

<i>Président :</i>	Sylvain CONTASSOT-VIVIER :	Professeur, Université Henri Poincaré, Nancy 1
<i>Rapporteurs :</i>	Mathias QUOY :	Professeur, Université de Cergy-Pontoise
	Arnaud REVEL :	Professeur, Université de la Rochelle
<i>Examineurs :</i>	Frédéric ALEXANDRE :	Directeur de Recherche INRIA, Professeur, Université Henri Poincaré, Nancy 1 (directeur de thèse)
	Hervé FREZZA-BUET :	Enseignant-chercheur, École Supérieure d'Électricité (co-directeur de thèse)
	Hugues BERRY :	Chercheur, INRIA Rhone-Alpes

Mis en page avec la classe thloria.

Remerciements

L'itinéraire scientifique qui m'a amené à ce manuscrit aurait été inconcevable sans mon co-directeur de thèse, Hervé Frezza-Buet. C'est lors de mon stage effectué pour le projet de fin d'études universitaires qu'Hervé m'a fait découvrir la « boîte des mystères cérébraux » et il m'a insufflé le goût pour l'informatique bio-inspirée. Grâce à lui, quelque temps plus tard, nous avons pu entamer ensemble une vraie aventure dans le « pays des modèles neuronaux de traitement d'information ». Ses compétences professionnelles remarquables, son approche pédagogique, son caractère chaleureux, sincère et convivial, ont transformé toute cette aventure dans une odyssée inoubliable, à travers laquelle chaque expérience m'a enrichi à la fois du point de vue scientifique et humain. Pour tout ceci, pour la confiance qu'il m'a donnée dans les moments essentiels de la thèse et son soutien continu, ainsi que pour le « terroir » de la langue française qu'il m'a fait découvrir, je lui suis profondément reconnaissant.

Ce parcours de recherche a également été rendu possible grâce à mon directeur de thèse, Frédéric Alexandre. Sa clarté d'exposition, son approche pédagogique et sa rigueur scientifique m'ont apporté chaque fois de nouvelles inspirations et m'ont induit de nouvelles perspectives vis-à-vis de mes travaux et de mes connaissances. Ses capacités tout à fait remarquables de synthèse et d'argumentation scientifique ont fait développer ma capacité à explorer et à me focaliser dans le vaste, dense et très dynamique univers de la recherche, et m'ont aidé considérablement à remettre en question et à raffiner périodiquement mes propres visions et approches de recherche. Pour tous ces aspects, ainsi que pour ses qualités humaines manifestées avec une grande modestie, je lui suis aussi profondément reconnaissant.

Je souhaite exprimer également ma gratitude envers tous les membres du jury, à la fois pour leur intérêt vis-à-vis de mes travaux de recherche, pour l'effort investi pour la lecture du manuscrit, ainsi que pour les remarques et les discussions très stimulantes lors de la soutenance, concernant le contexte et les perspectives de mes travaux.

Je tiens à remercier Constantinos et Bassem pour l'ambiance fraternelle et joviale qu'ils ont maintenue en permanence au bureau ou en dehors des heures de travail, ainsi que pour le passionnant partage des expériences socio-culturelles, culinaires et scientifiques.

Je remercie aussi Jérémy et Matthieu pour les idées que nous avons échangées, ainsi que pour leurs encouragements pendant les moments clés de la thèse.

Je tiens à remercier les membres de la « grande famille » des équipes du Loria Nancy et de Supélec à Metz, au sein desquelles j'ai effectué mes travaux. Je pense spécialement à Thomas, Mauricio, Mathieu, Nico, Axel, Laurent, Yann, Bernard et Jean-Charles de Nancy, à Thierry de Sophia-Antipolis, ainsi qu'à Gilles, Sébastien, Jean-Baptiste, Olivier, Virginie, Senthil, Joel, Patrick, Claudine, les deux Stéphane et les deux Frédéric de Metz.

Je remercie également la Région Lorraine et Supélec pour avoir financé ce travail, ainsi que toutes les personnes que je n'ai pas mentionnées, mais qui ont enrichi mon expérience au niveau professionnel ou humain.

Enfin, je remercie profondément mes parents, pour leur dévouement incessant à travers les années, ainsi que pour m'avoir accompagné et soutenu à chaque instant au cours de ce fascinant périple.

À mes parents.

Table des matières

1

Introduction

1.1	Information et connaissance	1
1.2	De l'épistémologie à l'informatique	2
1.3	Visions et approches générales de recherche	3
1.4	Itinéraire de recherche en résumé	9

2

Processus et principes de calcul corticaux

2.1	Propriétés physiologiques et fonctionnelles du cortex	12
2.1.1	Ségrégation anatomique du traitement cortical	13
2.1.2	Plasticité corticale	13
2.1.3	Cartes corticales	15
2.1.4	Colonnes corticales	16
2.2	Paradigmes de modélisation neuronale	19
2.2.1	Réseaux feedforward	21
2.2.2	Réseaux récurrents	21
2.2.3	Réseaux d'unités excitatrices/inhibitrices	21
2.2.4	Réseaux de type mémoire associative	22
2.2.5	Réseaux stochastiques	23
2.2.6	Réservoirs	24
2.3	De la dynamique au calcul	25
2.3.1	Quelques notions de la théorie des systèmes dynamiques	25
2.3.2	Le calcul comme un effet émergent d'un système dynamique	26
2.4	Modèles d'apprentissage cortical	28
2.4.1	Paradigmes d'apprentissage dans les modèles neuronaux	29
2.4.2	Apprentissage Hebbien	29
2.4.3	Cartes d'auto-organisation	30

2.5	Principes de calcul cortical	33
2.6	Architecture d'apprentissage compétitif d'inspiration corticale	35
2.6.1	Architecture neuronale d'auto-organisation basée sur la dynamique des champs neuronaux	37
2.6.2	Comparaison avec d'autres modèles de compétition neuronale	39
2.7	Résumé	40

3

Champs neuronaux dynamiques

3.1	Concepts généraux	44
3.1.1	Nécessité d'une modélisation mésoscopique de l'activité corticale	44
3.1.2	Approches de modélisation mésoscopique de l'activité corticale	44
3.1.3	Description générale de la dynamique d'un champ neuronal	46
3.1.4	Critiques du paradigme de modélisation des champs neuronaux dynamiques	47
3.2	Modèles théoriques et comportements dynamiques	49
3.2.1	Modèle de Wilson et Cowan	49
3.2.2	Modèle d'Amari	52
3.2.3	Extension bidimensionnelle des modèles de champs neuronaux	57
3.2.4	Modèle d'Erlhagen	58
3.2.5	Modèle de Pinto et Ermentrout	59
3.2.6	Autres modèles de champ	61
3.3	Méthodes d'analyse dynamique des modèles de champ	70
3.4	Méthodes de simulation des modèles théoriques	71
3.5	Discussion	73

4

Applications des champs neuronaux dynamiques

4.1	Modélisation de l'activité corticale au niveau physiologique	76
4.1.1	Modélisation de l'activité du cortex visuel primaire, induite par une sti- mulation visuelle simple	76
4.1.2	Modélisation de la préférence d'orientation visuelle	76
4.1.3	Modélisation des phénomènes neurophysiologiques anormaux	77
4.2	Modélisation des processus cognitifs élémentaires	78
4.2.1	Modélisation des mécanismes neuronaux de mémoire	78
4.2.2	Modélisation des processus d'attention	81
4.2.3	Vers la modélisation des mécanismes neuronaux liés aux phénomènes de la conscience	82

4.3	Modélisation des traitements sensorimoteurs et des systèmes situés	84
4.3.1	Problématiques et approches de modélisation	84
4.3.2	Modélisation à l'aide d'un système dynamique dual	86
4.3.3	Modélisation par duplication des commandes motrices	88
4.3.4	Modélisation par utilisation d'une carte cognitive de cellules de transition	90
4.3.5	Modélisation de la coordination des tâches robotiques	91
4.3.6	Modélisation intégrant l'apprentissage par renforcement	93
4.4	Apprentissage d'architectures neuronales basé sur la dynamique des champs neu- ronaux	96
4.4.1	Apprentissage des connexions afférentes	97
4.4.2	Apprentissage multimodal	98
4.5	Discussion	99

5

Méthodologie d'étude empirique des champs neuronaux dynamiques

5.1	Scénarios pertinents pour évaluer la dynamique d'un champ	102
5.1.1	Scénarios d'attention	104
5.1.2	Scénarios d'exploration	104
5.1.3	Scénarios auxiliaires	104
5.1.4	Scénario d'auto-organisation unidimensionnelle	105
5.1.5	Scénario d'auto-organisation bidimensionnelle	106
5.2	Évaluation des capacités dynamiques d'un champ neuronal	107
5.2.1	Propriétés dynamiques souhaitables et critères d'optimalité	107
5.2.2	Instrument d'évaluation	110
5.2.3	Implantation de l'instrument d'évaluation. Champs de référence	113
5.2.4	Pertinence des critères d'optimalité	116
5.2.5	Plate-forme d'analyse empirique de la dynamique des champs neuronaux	119
5.3	Analyse du comportement dynamique des modèles classiques	120
5.3.1	Scénarios d'attention	121
5.3.2	Scénarios d'exploration	122
5.3.3	Scénarios auxiliaires	122
5.3.4	Scénario d'auto-organisation unidimensionnelle	130
5.3.5	Scénario d'auto-organisation bidimensionnelle	132
5.3.6	Discussion	134
5.4	Méthodologie d'ajustement fin des paramètres d'un modèle de champ	137
5.5	Discussion	141

6	
Champs neuronaux dynamiques pour l'auto-organisation	
6.1	Champs neuronaux dynamiques rétro-inhibés (CNDRI) 144
6.2	Résolution des tâches d'auto-organisation avec des champs neuronaux dynamiques 148
6.2.1	Auto-organisation unidimensionnelle 148
6.2.2	Auto-organisation bidimensionnelle 149
6.2.3	Discussion 149
6.3	Dynamiques émergentes de CNDRI 151
6.3.1	Scénarios d'attention 151
6.3.2	Scénarios d'exploration 152
6.3.3	Discussion 156
6.4	Les paramètres du modèle CNDRI 156
6.4.1	Procédure de paramétrage des équations du modèle CNDRI 156
6.4.2	Influence des paramètres des équations du modèle CNDRI 160
6.5	Discussion 169
7	
Conclusions	
7.1	Résumé et analyse des résultats 173
7.2	Perspectives 175
7.2.1	Perspectives envisagées pour nos approches 175
7.2.2	Perspectives envisagées pour le domaine 176
7.2.3	Défis de l'avenir 177
Annexes	181
A Liste de publications	181
Bibliographie	183

Chapitre 1

Introduction

1.1 Information et connaissance

Étymologiquement, le verbe *connaître*, provenant de l'ancien perse *xsna*, signifie « se rendre compte » ou « prendre connaissance », et sous-entend à la fois une action et un sujet : un regard (donc un point de vue) effectué par un individu, à l'égard d'un univers inconnu. Cependant, le jeu de mots que ce terme cache (*co* – *naître*) pourrait illustrer encore mieux le caractère dynamique et subjectif de la connaissance. Ainsi, connaître pourrait être vu comme l'acte de donner du sens à ce qu'on appelle généralement l'*information*. Cette action d'interpréter (attribuer un sens à) une information donne son caractère subjectif. Connaître signifie créer des connexions entre différents objets (physiques ou mentaux) de la réalité. De ce point de vue, la connaissance désigne une mise en relation, et représente un processus dynamique réalisé par un sujet. Ce simple regard vis-à-vis du concept de la connaissance est, en lui-même, un exemple de ce processus de création de sémantique.

Si elle est souvent considérée comme une abstraction de la réalité physique (c'est-à-dire une méta réalité), la connaissance est en elle-même l'expression et l'exemple d'une réalité physique. Pour prendre forme, et pour être transmise, la connaissance recourt à un support physique. De cette perspective, la connaissance est une fois de plus, une mise en relation de la réalité avec elle-même.

Cependant, la connaissance n'est pas uniquement le résultat d'un processus d'association et d'intégration d'objets de la réalité, mais aussi le résultat d'une exploration et d'une séparation de la réalité, deux activités destinées à découvrir et définir ces objets. Si l'origine et la nature de la connaissance font l'objet des études épistémologiques, la phénoménologie intrinsèque des processus cognitifs (de création et de manipulation de la connaissance) fait l'objet des études des *sciences cognitives*.

Placées au carrefour d'une étude pluridisciplinaire, les sciences cognitives visent à intégrer différents résultats des recherches biologiques, neuroscientifiques, psychologiques, linguistiques, et informatiques, à la fois pour proposer des modèles théoriques pour décrire les processus cognitifs (ce qui en anglais porte le nom de *theory of mind*), et pour développer des systèmes physiques capables d'implémenter et d'employer ces modèles théoriques.

Dans ce cadre d'étude, la démarche scientifique y participe en explorant la réalité et en créant des objets abstraits (concepts, modèles, formalismes, etc.) qui pourraient décrire au mieux la structure de connexion entre certains objets de la réalité. Cette entreprise vise à valider à la fois la cohérence interne de cette méta réalité qu'elle conçoit, et sa corrélation avec l'expérience de réalité qu'elle observe.

« *Like everything metaphysical the harmony between thought and reality is to be found in the grammar of the language.* » (L. Wittgenstein, *Philosophical Investigations*)

Si chaque connaissance fait preuve de son caractère relatif (subjectif), le même caractère demeure quand il s'agit de fonder une théorie de la connaissance.

« *All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our arguments; nor it belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which our arguments have their life.* » (L. Wittgenstein, *On Certainty*)

Ainsi, nous pouvons affirmer que le cœur de tout système théorique repose sur une vision ontologique de la réalité. Peut-être plus que dans beaucoup d'autres disciplines, dans le cas des sciences cognitives, les débats philosophiques formulent non seulement les hypothèses des systèmes théoriques, mais sillonnent également les visions et les approches de recherche. Loin d'être conçue comme un « simple instrument de calcul », l'*informatique* elle-même est apparue comme résultat d'un tel débat philosophique.

1.2 Del'épistémologie à l'informatique

Depuis les temps les plus anciens de l'histoire de la pensée humaine, les philosophes ont vivement confronté leurs visions épistémologiques, c'est-à-dire leurs visions concernant la nature et les frontières de la connaissance. Dans un ouvrage de référence de la philosophie, *Critique de la raison pure*, I. Kant argumente l'impossibilité de connaître la réalité en tant que telle, c'est-à-dire au-delà de nos moyens et modèles (perceptifs ou cognitifs) relatifs aux phénomènes du réel. G. Deleuze, résume l'idée de cette critique kantienne, par la phrase suivante :

« *Nous reconnaissons les choses, nous ne les connaissons pas.* » (G. Deleuze, *Proust et les signes*)

Cette vision délimite ainsi la réalité de la connaissance, et met en cause l'objectivité absolue de cette dernière. Comme conséquence de cet argument, au cours des années, le discours épistémologique s'est redirigé vers la problématique de l'acquisition et du traitement de l'information pour créer et opérer avec des connaissances.

Sous un aspect formel, le noyau de ce thème inclut un fameux problème étudié par Leibniz, et connu sous le nom allemand *Entscheidungsproblem*. En résumé, ce problème cherche à trouver une description logique (c'est-à-dire un algorithme) capable de décider si une affirmation quelconque (donc une connaissance), décrite dans un langage formel, est vraie ou fausse.

S'attelant à ce problème, K. Gödel montre qu'un système axiomatique cohérent du point de vue de la logique classique, capable d'exprimer les principes arithmétiques élémentaires, est incapable de valider toutes les affirmations concernant l'arithmétique des nombres naturels. De plus, si un tel système est capable de valider certaines propositions logiques, il devient impossible de prouver sa propre cohérence logique. Ces théorèmes de Gödel sur la non décidabilité de certaines affirmations logiques mettent en évidence les limites non seulement de la connaissance, mais aussi de la méthode scientifique même (la réfutation d'un théorème par une approche inductive).

Intéressés par un autre problème mathématique, celui de la *calculabilité* d'une valeur numérique d'une fonction, A. Church [20], et A. Turing [117] conçoivent indépendamment deux systèmes de manipulation de symboles, respectant les principes arithmétiques élémentaires. Utilisant chacun son propre formalisme, les deux chercheurs montrent que ce problème se réduit au

Entscheidungsproblem ci-dessus mentionné. Dans les deux cas, les auteurs confirment les résultats de Gödel : il est impossible de construire un mécanisme capable de décider la valeur de vérité de n'importe quelle affirmation logique. Malgré cette conclusion déconcertante, les résultats de Church et Turing posent les bases d'une toute nouvelle discipline, c'est-à-dire celle du traitement de l'information, ou plus brièvement, *l'informatique*. Les formalismes développés par les deux auteurs, portant aujourd'hui respectivement les noms de λ -calcul et *machine de Turing*, représentent les deux principaux formalismes implémentés par les modèles et les architectures de calcul utilisés à présent dans le traitement de l'information.

Si les démarches de ces pionniers de l'informatique exhibent le caractère limitatif de la connaissance, l'informatique s'est imposée comme discipline pour sa capacité de fournir un moyen pratique d'explorer et de créer l'univers de la connaissance. Au début, elle a été utilisée comme routine automatique de réfutation logique. Ensuite, elle est devenue un outil capable de simuler différents aspects d'un modèle théorique, afin d'explorer les conséquences de son fonctionnement dans un certain contexte. Enfin, plus récemment, l'informatique est devenue un instrument de création de connaissance. Bien que les sciences cognitives utilisent l'informatique sous tous les trois aspects mentionnés ici, son apport principal est celui d'outil de création de connaissance. Actuellement, les différentes approches informatiques de ce domaine visent avec prédilection la conception des modèles de traitement d'information, matérialisées sous la forme des solutions logicielles ou robotiques autonomes et adaptatives, capables de s'atteler aux problèmes complexes de la connaissance.

1.3 Visions et approches générales de recherche

Si la connaissance, telle que nous l'avons présentée, représente un processus de création de liaisons entre les objets d'une réalité, elle pourrait laisser entendre le fait que l'homme, en tant que créateur de connaissance, est uniquement une entité de traitement d'information. Au premier regard, ce concept de dichotomie de la réalité (concept ayant une forte influence dans l'histoire de la pensée humaine) pourrait être considéré naturel et même désirable dans le cadre d'une analyse théorique (afin de pouvoir bien définir les limites des objets de la réalité et opérer proprement avec ces objets). Cependant, les conséquences de cette vision générale dualiste du monde, prise comme hypothèse initiale pour expliquer de nombreux phénomènes, sont visibles non seulement dans l'évolution des approches scientifiques et des discours philosophiques (voir par exemple les séparations : masse - énergie, organique - inorganique, corps - esprit, etc.) mais aussi, sur un plan global, dans la séparation de l'homme de son environnement et son effet dévastateur envers la biosphère. Dans beaucoup de cas, cette hypothèse s'est rendue fallacieuse, et elle est aujourd'hui très critiquée au milieu des débats philosophiques (voir le dualisme cartésien).

L'apparition des machines de calcul et de l'informatique comme discipline d'étude du traitement de l'information a permis le développement des modèles théoriques fondés sur cette hypothèse (c'est-à-dire celle qui considère la connaissance exclusivement comme processus de traitement de l'information, entièrement dissocié de la réalité extérieure). Grâce aux pionniers de l'informatique, comme A. Turing, J. McCarthy ou M. Minsky, cette hypothèse a pu être implantée sous la forme des systèmes formels de manipulation de symboles. Cette entreprise est devenue une des premières tentatives de s'atteler à la problématique des processus cognitifs, et elle est devenue l'approche fondatrice du domaine général de recherche de *l'Intelligence Artificielle*.

En lignes générales, cette approche, connue sous le nom anglais de *computational theory of*

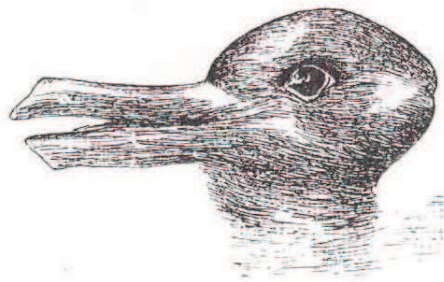


FIG. 1.1 – L'image du « canard-lapin » de J. Jastrow, utilisée par Wittgenstein dans *Philosophical Investigations* pour analyser le concept de l'ambiguïté.

*mind*¹, considère que tout processus cognitif peut s'exprimer sous la forme d'un *calcul*, c'est-à-dire par une chaîne d'opérations élémentaires effectuées sur un ensemble (fini) de symboles. Appliquée dans le cadre du système logique du premier ordre, elle est devenue ce qu'on appelle aujourd'hui l'approche *symbolique*.

Malgré ses avantages, cette approche s'est vue assez rapidement limitée dans son but de décrire la nature intrinsèque des processus cognitifs. Le cadre formel assuré par cette approche computationnelle ne réussit pas à expliquer par exemple l'ambiguïté issue d'un traitement cognitif. L'ambiguïté est une preuve de la subjectivité de l'acte de connaissance, or le traitement formel symbolique se veut non ambigu. Ce paradigme de calcul tient fort à implémenter un traitement déterministe : à partir d'une même entrée (information), on souhaite arriver toujours à une même sortie (connaissance). Un des avantages de ce paradigme de calcul (traitement basé sur une chaîne inductive d'opérations) est que chaque résultat (intermédiaire ou final) peut être analysé, validé ou réfuté dans un cadre logique formel. Cependant, appliqué à l'analyse des images (en particulier au domaine de reconnaissance faciale), ce paradigme échoue à fournir une solution robuste pour le processus cognitif de reconnaissance visuelle. Appliquer la même chaîne d'opérations à une forme quelque peu changée ou déformée d'une image conduit souvent à un résultat différent, ce qui peut se traduire par une incapacité à reconnaître l'image initiale. De plus, dans ce paradigme, il est impossible que la même image puisse *signifier* une fois une chose et une autre fois une autre chose. Ainsi, ce paradigme ne réussit pas à expliquer l'ambiguïté (la bistabilité) d'une image telle que celle du « canard-lapin » (voir la figure 1.1). L'image est reprise par Kihlstrom dans [71] pour illustrer le fait que la perception n'est pas le produit exclusif d'un stimulus, mais celui d'une activité mentale. Comme nous pouvons le constater, bien que l'information initiale (le flot perceptif) reste inchangée, elle peut engendrer des résultats (des connaissances) différents.

Les débats sur l'invariance ou la multistabilité d'un processus cognitif, mais aussi sur la manière d'intégrer les composantes élémentaires du flot perceptif dans la connaissance (*binding problem*), représentent des sujets centraux de l'étude de la psychologie Gestalt. Même si cette théorie est plutôt une analyse phénoménologique de la nature des processus cognitifs, elle met en évidence la nature émergente de ces processus, c'est-à-dire la création de nouveaux objets mentaux à partir des opérations et interactions élémentaires avec la réalité.

Dans ce contexte, l'approche symbolique laisse entendre que, au cours du traitement de

¹Dans la suite de ce document, pour faciliter la communication, nous allons utiliser l'anglicisme *computationnelle* pour faire référence à cette théorie.

l'information, certains modules de traitement réalisent une interprétation de l'information, afin de résoudre l'ambiguïté. Cependant, cet argument ne fait que déplacer le problème d'élucider l'ambiguïté d'une information à une étape ultérieure que celle initiale (par exemple celle perceptive). Pour son analogie avec l'image induite vis-à-vis du processus mental humain, cet argument fallacieux a été nommé *l'argument de l'homuncule*.

Comme alternative à cette approche symbolique, plusieurs chercheurs ont proposé une nouvelle approche computationnelle, fondée sur l'hypothèse que tout processus cognitif est le résultat émergent d'une multitude d'interactions entre différentes entités élémentaires de traitement d'information. À présent, cette approche, initiée par les travaux de F. Rosenblatt, D. E. Rumelhart ou J. J. Hopfield sur les réseaux neuronaux, porte le nom de *sub-symbolique* ou *connexionniste*. L'aspect qui distingue essentiellement cette approche de celle symbolique réside dans la dynamique du traitement de l'information : le flot de ce traitement n'est pas unidirectionnel ou hiérarchisé, les entités de traitement s'influencent souvent réciproquement (directement ou indirectement), via leurs connexions.

Les résultats expérimentaux et théoriques des recherches neurophysiologiques concernant l'anatomie et le fonctionnement des systèmes nerveux ont trouvé rapidement leur écho dans ce paradigme de traitement de l'information. Cette forme du paradigme computationnel connexionniste, qui intègre ou modélise certains aspects de ces résultats, définit ce qu'on appelle aujourd'hui le domaine des *neurosciences computationnelles*.

Bien qu'il étende la portée de la modélisation computationnelle des processus cognitifs, ce paradigme ne réussit pas, par lui-même, à aborder l'aspect relationnel dynamique de l'acte de connaissance, c'est-à-dire l'interaction sujet—réalité : l'action effectuée par un sujet pour saisir et traiter le flot d'information. Or, comme l'affirme A. Berthoz, la perception — l'étape initiale même d'un processus cognitif — sous-entend un processus actif :

« La perception n'est pas seulement une interprétation des messages sensoriels : elle est contrainte par l'action, elle est simulation interne de l'action, elle est jugement et prise de décision, elle est anticipation des conséquences de l'action. » (A. Berthoz, *Le Sens du Mouvement*)

Ainsi, tout processus cognitif dépend d'une interaction permanente avec l'environnement, et sous-entend une perspective (un positionnement pré-cognitif) du sujet. Cet aspect de « la perception [comme] une sélection active » (A. Berthoz) avait été exprimé à l'égard du niveau phénoménologique de l'action par le philosophe M. Merleau-Ponty, qui déclarait : « la vision est palpation par le regard ». De cette perspective, le processus cognitif est un processus incarné et relationnel. Cette vision a conduit des chercheurs comme T. van Gelder ou I. Harvey à affirmer que les processus cognitifs ont une nature intrinsèquement dynamique, plutôt qu'une nature purement computationnelle [120, 56].

« In this vision, the cognitive system is not just the encapsulated brain ; rather, since the nervous system, body, and environment are all constantly changing and simultaneously influencing each other, the true cognitive system is a single unified system embracing all three. The cognitive system does not interact with the body and the external world by means of the occasional static symbolic inputs and outputs ; rather, interaction between the inner and the outer is best thought of as a matter of coupling, such that both sets of processes continually influencing each other's direction of change. At the level at which the mechanisms are best described, cognitive processing is not sequential and cyclic, for all aspects of the cognitive system are undergoing change all the time. Any sequential character in cognitive performance is the high-level, overall trajectory of change in a system whose rules of evolution

specify not sequential change but rather simultaneous mutual coevolution. » (T. van Gelder, *What Might Cognition Be, If Not Computation ?*)

Comme T. van Gelder l'affirme dans [120], un des défis majeurs des processus cognitifs concerne la manière dont ils traitent l'aspect temporel de l'interaction entre le sujet et l'environnement. Si le paradigme computationnel symbolique offre comme possibles outils de modélisation le traitement séquentiel, modulaire ou hiérarchisé de représentations symboliques, il ne réussit pas à saisir proprement les notions de la durée ou du rythme, intrinsèquement liées à la nature temporelle d'un processus cognitif. Cet argument mène T. van Gelder à affirmer alors que le processus cognitif est le résultat émergent du fonctionnement d'un *système dynamique*.

Néanmoins, le paradigme computationnel connexionniste est un cas particulier de système dynamique. Si on considère que chacune des entités élémentaires implémente la dynamique d'un processus, l'interaction de l'ensemble de ces entités forme un système dynamique complexe. Notons que dans ce contexte, l'échange informationnel entre deux entités n'a plus une nature purement symbolique, mais plutôt une nature numérique (quantitative). Remarquons également que ce paradigme n'a plus besoin de faire recours à l'argument de l'homuncule pour expliquer la procédure du traitement de l'information. Le résultat d'un système implémentant ce paradigme d'information n'est pas une représentation symbolique d'un flot d'information, mais l'effet émergent des interactions entre les entités élémentaires qui le composent.

Dans cette perspective, ce que nous appelons aujourd'hui *esprit* pourrait être considéré, pour citer D. Siegel, comme étant un « processus relationnel incarné, qui régule le flot de l'énergie et de l'information ». Par conséquent, l'*intelligence* pourrait être vue comme étant une qualité qui caractérise la relation causale des interactions entre une entité et son environnement (y compris l'entité elle-même). Au sens le plus large, l'intelligence est liée à la capacité d'une entité à s'adapter à la dynamique de son environnement. Pour cette raison, le long de ce document, toute référence aux *systèmes intelligents* sous-entendra la capacité de ces systèmes à interagir avec un environnement a priori inconnu, et à s'adapter en permanence aux défis des réalités qu'ils découvrent.

Si l'intelligence a été longtemps considérée comme uniquement un attribut du comportement humain, on convient aujourd'hui à l'attribuer à une gamme assez large d'espèces vivantes. Notons dans ce contexte que dans le sens le plus général, la vie est un processus dynamique résultant de l'interaction d'une multitude de phénomènes de transfert énergétique, une forme d'organisation de la matière capable d'échapper à l'immuabilité du minéral pour créer son propre équilibre dynamique. À travers le temps, l'évolution, processus intrinsèque de la vie, a permis aux organismes vivants d'explorer le champ des possibilités décrit par les lois physiques, de développer et d'optimiser des mécanismes permettant d'accomplir différents types d'équilibres entre l'organisme vivant et son environnement. Au cours de l'évolution, ces mécanismes se sont spécialisés de manière interdépendante, sous la forme d'outils d'exploration (organes de perception et d'action), mais aussi d'outils de traitement, de régulation et de contrôle. Dans ce contexte, l'intelligence désigne la capacité d'un organisme à s'adapter à la dynamique des interactions qu'il subit avec l'environnement. De ce point de vue, l'intelligence ne définit pas une qualité irréductiblement liée à la nature organique d'un sujet, se référant à la dynamique qui caractérise l'interaction d'une entité avec le flot d'énergie et d'information. Cette vue a permis aux chercheurs du domaine de l'intelligence artificielle d'énoncer ce que aujourd'hui on appelle la proposition de Dartmouth :

« Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it. » (Dartmouth proposal, 1956)

Grâce à sa formulation pratique, cette hypothèse s'est vue adoptée rapidement par une grande partie de la communauté scientifique, fondant ainsi la base épistémologique de la majorité des

travaux informatiques dans le domaine de l'intelligence artificielle. Cependant, une certaine minorité de cette communauté a contesté cette conjecture. Parmi les exposants de ce débat, nous notons J. Searle, qui avait formulé l'argument suivant (de la *chambre Chinoise*) pour combattre la proposition de Dartmouth : si une personne qui ne connaît pas du tout la langue chinoise, suit pas à pas une procédure de tri d'un ensemble de caractères chinois, qui garantit la formulation des phrases correctes du point de vue syntaxique et sémantique, elle ne pourra pas affirmer à la fin de cette procédure d'avoir *compris* la signification de ces phrases. Sous son aspect fonctionnel, l'argument de J. Searle souhaite mettre en évidence la différence entre la simulation et l'existence proprement dite de la pensée intelligente, autrement dit la différence entre le processus de traitement de l'information et celui de compréhension de l'information. Intituler alors un système *intelligent* serait-il le résultat d'une « campagne publicitaire » ou d'un malentendu des notions sous-jacentes aux concepts de l'intelligence et de l'esprit ?

« The question of whether a computer is playing chess, or doing long division, or translating Chinese, is like the question of whether robots can murder or airplanes can fly – or people ; after all, the "flight" of the Olympic long jump champion is only an order of magnitude short of that of the chicken champion (so I'm told). These are questions of decision, not fact ; decision as to whether to adopt a certain metaphoric extension of common usage. » (N. Chomsky, *Powers and Prospects*)

Aujourd'hui, affirmer que les processus mentaux peuvent être entièrement émulés (cf. la proposition de Dartmouth) par d'autres systèmes physiques (des machines de calcul en particulier), constitue l'hypothèse faible de l'intelligence artificielle, alors que la position de J. Searle de distinguer les concepts de l'intelligence et la compréhension, définit l'hypothèse forte de l'intelligence artificielle. Le sujet si la nature des processus cognitifs (chez les espèces vivantes) est ontologiquement différente de la nature des processus cognitifs décrits par les théories actuelles des sciences cognitives reste un aspect fondamental qui continue à être vivement débattu par les théoriciens de ce vaste domaine de recherche, afin de décider s'il est nécessaire de continuer ce que Wittgenstein appelait « le combat contre l'ensorcellement du langage ».

Malgré ces fondations conceptuelles, qui ne font pas l'objet d'un consensus sur le plan ontologique, l'informatique s'est démarquée dans le contexte des sciences cognitives en abordant toujours un point de vue pratique. Ce point de vue, exprimé par A. Turing, un des pionniers de cette discipline, considère que tout système artificiel dont son comportement (mis en évidence dans un certain contexte et vis-à-vis d'une certaine problématique) est qualitativement indiscernable du comportement humain, a le droit d'être considéré aussi intelligent [118]. Comme nous pouvons le constater, cette perspective n'impose pas vraiment une définition, elle proposant plutôt un critère d'évaluation du comportement global d'un système physique, et non pas de son fonctionnement intrinsèque. Si ce point de vue anthropomorphique (construire des systèmes physiques dont le comportement est similaire à celui d'une personne humaine) a conduit les chercheurs à proposer le paradigme de traitement symbolique de l'information (similaire au langage logico-mathématique développé par l'homme), grâce aux avancements des neurosciences, cette perspective a pu se focaliser sur la manière dont le cerveau traite l'information et interagit avec l'environnement.

Sous aspect historique, l'intelligence a évolué d'une phase d'utilisation de ressources naturelles, à une phase de construction et d'utilisation d'outils physiques, et enfin à une phase de conception et d'utilisation d'outils mentaux. À l'égard de ce parcours, le temps est venu, comme déclarait M. Pawlyn, pour que l'homme utilise les connaissances acquises « par la nature pendant sa phase continue de recherche et du développement, effectuée depuis les derniers milliards d'années », plutôt que ses ressources brutes. Si la vision newtonienne de décortiquer les « principes

de la nature » a été intensément appliquée pour étudier la phénoménologie de l'organisation élémentaire de l'univers (même en abandonnant la méthode newtonienne), cette vision pourrait guider également l'exploration des formes d'organisation plus complexes de la matière du même univers, afin de comprendre le microcosme dans lequel nous vivons ou que nous-mêmes sommes.

Notre approche de recherche s'inscrit dans cette démarche générale de s'inspirer des solutions biologiques développées par la nature au cours de l'évolution, pour concevoir des modèles informatiques de traitement d'information et d'interaction avec un environnement, modèles qui soient capables d'explorer et de s'adapter en permanence à une réalité a priori inconnue.

En particulier, la source principale d'inspiration pour nos recherches est constituée par la structure corticale. Les arguments qui nous ont conduit à focaliser notre intérêt de recherche et de modélisation sur cette structure cérébrale reposent principalement sur le rôle déterminant du cortex dans l'activité des processus cognitifs, ainsi que son caractère structurel générique au niveau anatomique et fonctionnel.

Cependant, comme nous allons l'examiner dans le chapitre suivant, déceler les mécanismes neurophysiologiques qui régissent au niveau cortical constitue un défi majeur pour les recherches neuroscientifiques actuelles. Du point de vue de la démarche informatique, notre premier objectif de recherche vise à examiner les résultats expérimentaux accumulés par la littérature du domaine et mettre en évidence les principes fonctionnels de nature corticale qui pourraient servir à l'implantation des systèmes artificiels de traitement d'information. Ensuite, nos travaux s'appuient sur le processus cortical d'auto-organisation, processus intrinsèque au fonctionnement cortical, responsable pour l'émergence et l'organisation de différentes cartes corticales.

Si les études expérimentales indiquent une forte influence de ce processus d'organisation corticale dans le traitement d'information de nature uni- et multi- modale, les mécanismes neuronaux sous-jacents qui le définissent restent largement inconnus.

Par conséquent, le deuxième objectif visé par notre recherche concerne la conception d'un modèle de calcul neuronal capable d'exhiber des propriétés similaires aux propriétés qui caractérisent la structure corticale, et notamment le processus d'auto-organisation de nature corticale. Étant placée dans un contexte informatique, nous notons que notre démarche de modélisation ne suit pas comme condition nécessaire l'argument de plausibilité biologique, son but étant plutôt celui de s'inspirer des résultats neuroscientifiques présents pour intégrer certaines caractéristiques et principes de fonctionnement cortical dans un cadre formel étendu de calcul neuronal. À cet égard, l'approche que nous proposons ici décrit le travail de conception d'un modèle de traitement générique, distribué et évolutif de l'information, modèle qui pourrait être utilisé ensuite pour le développement des applications logicielles, ainsi que des systèmes informatiques situés.

Notre démarche de modélisation repose sur le formalisme des champs neuronaux dynamiques, un paradigme connexionniste et dynamique de modélisation neuronale de l'activité corticale qui a été appliqué avec beaucoup de succès dans la modélisation d'une multitude de mécanismes et de processus de nature corticale. Cependant, comme nous allons le constater, l'utilisation de ce paradigme dans le contexte de modélisation que nous avons décrit ici pose un défi supplémentaire et formule un nouvel objectif pour nos travaux : trouver un modèle de champ approprié pour l'implantation pratique du modèle de calcul neuronal conçu pour décrire le processus d'auto-organisation.

Pour conclure, ce mémoire décrit une approche de conception computationnelle, connexionniste et dynamique, qui vise à modéliser le processus d'auto-organisation corticale, un processus essentiel dans le traitement neuronal de l'information au niveau cortical. La section suivante résume le parcours des travaux effectués pour accomplir les objectifs de recherche décrits ci-dessus.

1.4 Itinéraire de recherche en résumé

Si les débats vis-à-vis des limites théoriques des approches que nous avons présentées ci-dessus (computationnelle en général et dynamique en particulier) pour expliquer la nature des processus cognitifs sont loin de diminuer en portée ou intensité, les limites pratiques de ces approches sont aussi loin d'avoir été atteintes. Les résultats des recherches effectuées dans ce cadre théorique peuvent être utilisés dans le développement des systèmes situés et des solutions logicielles plus adaptés à un spectre de problèmes plus large. Par conséquent, nous considérons que les bases théoriques des paradigmes exposés ici méritent d'être explorées en profondeur. Nos travaux de recherche, présentés dans ce mémoire, décrivent une telle entreprise.

Comme nous l'avons précisé dans le paragraphe 1.3, notre contexte général de recherche concerne la conception des modèles computationnels bio-inspirés, capables de participer au développement des solutions logicielles ou robotiques, adaptatives et autonomes. L'objectif particulier que nous visons dans ce contexte se réfère à la modélisation des processus corticaux d'auto-organisation. Les problématiques sous-jacentes à ce sujet sont abordées dans ce mémoire comme suit.

Dans le deuxième chapitre, nous présentons les éléments fondamentaux concernant la structure et le fonctionnement du tissu cortical, structure cérébrale qui constitue pour nos démarches de modélisation la source principale d'inspiration biologique. Ensuite, nous présentons un ensemble d'outils théoriques, développés dans le cadre des paradigmes connexionnistes et dynamiques de l'approche computationnelle de modélisation des processus neuronaux. Basés à la fois sur ces résultats expérimentaux et théoriques, nous mettons en évidence un certain nombre de principes de modélisation neuronale, compatibles avec le fonctionnement cortical. Au final de ce chapitre nous proposons une architecture neuronale d'auto-organisation, inspirée du traitement cortical de l'information, et respectant les principes généraux énoncés.

Afin d'implémenter cette architecture dans le cadre d'un paradigme connexionniste et dynamique, dans le troisième chapitre nous présentons un formalisme général de calcul neuronal dynamique, conçu pour modéliser l'activité corticale observée au niveau mésoscopique. Les différents modèles de référence de ce formalisme général de modélisation corticale, appelé *modèle des champs neuronaux dynamiques*, sont analysés dans ce chapitre, mettant en évidence les caractéristiques structurales et comportementales de chacun de ces modèles.

Les notions théoriques présentées dans le troisième chapitre nous permettent d'analyser dans le chapitre suivant comment le modèle des champs neuronaux peut être utilisé pour concevoir des architectures neuronales capables de modéliser différents aspects de processus cognitifs élémentaires, comme l'attention, la décision, la planification des actions, etc. Le spectre varié de ces applications nous aide à examiner différentes approches qui pourraient être utilisées dans notre propre démarche de modélisation.

Ainsi, dans le cinquième chapitre, nous présentons un cadre étendu d'analyse du comportement dynamique des modèles de champs neuronaux, examinant la capacité des modèles de référence à être utilisés dans certains contextes applicatifs. Cette démarche se concrétise dans la proposition d'une plate-forme empirique d'analyse dynamique de ces modèles neuronaux, et est utilisée pour implémenter une procédure d'ajustement de paramètres d'un modèle quelconque de champ, pour améliorer son comportement dynamique vis-à-vis d'un contexte applicatif donné. Les conclusions de notre analyse sur la capacité des champs classiques à implémenter l'architecture neuronale d'inspiration corticale que nous avons introduite dans le deuxième chapitre nous conduisent à considérer la conception d'un nouveau modèle de champ neuronal, capable d'accomplir l'objectif que nous nous sommes établi.

Ainsi, dans le sixième chapitre, nous proposons un nouveau formalisme de type champ neu-

ronal, inspiré des modèles classiques de champs neuronaux, mais introduisant un mécanisme neuronal de régulation locale, appelé *mécanisme de rétro-inhibition*. Dans ce chapitre, nous détaillons notre démarche de modélisation, examinant ensuite les performances du nouveau modèle dans les scénarios d’auto-organisation. Comme nous allons le constater dans ce chapitre, les résultats de cette analyse nous confirment la capacité de ce nouveau modèle à implémenter avec succès l’architecture neuronale d’auto-organisation proposée au début de ce mémoire.

Enfin, dans le dernier chapitre, nous résumons les résultats théoriques obtenus à travers notre démarche de recherche, mettant en évidence également les perspectives de ces travaux de modélisation neuronale d’inspiration corticale.

Chapitre 2

Processus et principes de calcul corticaux

Dans l'introduction de ce mémoire, nous avons évoqué les objectifs généraux de recherche visés par les sciences cognitives. Dans ce contexte, nous avons situé l'informatique comme étant à la fois un outil utilisé pour faciliter l'analyse des données expérimentales, et un support utilisé pour développer des solutions caractérisées par un grand degré d'autonomie et une grande capacité d'adaptation. Nous avons convenu de désigner de tels systèmes de traitement d'information comme *intelligents*. Dans notre contexte de recherche, nous sommes intéressé par ce deuxième axe de recherche, ce qui constitue le spectre d'étude de l'intelligence artificielle.

Parmi la multitude d'approches présentées dans le chapitre précédent, notre thèse se focalise sur la conception de systèmes de traitement d'information de type neuronal connexionniste, abordant une approche à la fois computationnelle et dynamique. Dans cette démarche scientifique, les connaissances mises en évidence par les neurosciences, concernant les processus neuronaux de nature corticale, constituent notre principale source d'inspiration.

Ce chapitre commence par présenter un résumé des faits expérimentaux, des hypothèses théoriques, et des modèles formels concernant l'anatomie et le fonctionnement de la structure corticale. Extrait de ces découvertes, un nombre de principes fonctionnels du traitement d'information sont énoncés, ceux-ci imposant pour la suite les lignes directrices pour la conception de nos propres modèles neuronaux théoriques. Ensuite, l'apprentissage au niveau neuronal est mis en évidence comme étant un processus déterminant dans le développement des capacités comportementales d'adaptation d'un système neuronal, pris dans l'ensemble de ces processus de traitement d'information. Les principaux types d'apprentissage neuronal sont énumérés et décrits en lignes générales. Certaines approches théoriques de modélisation de ceux-ci sont ensuite passées en revue. Dans ce contexte, nous évoquons notre intérêt pour la conception d'un mécanisme neuronal général, capable d'émuler le processus d'auto-organisation corticale.

Dans la dernière section du chapitre, une architecture de calcul générique d'inspiration corticale est proposée. Sa conception permet une description formelle unifiée, ainsi que la comparaison de plusieurs modèles d'apprentissage neuronal. Au final, les processus d'auto-organisation de nature corticale sont décrits avec cette architecture. Prenant en compte dans l'implantation de cette architecture des principes fonctionnels de calcul neuronal énumérés précédemment, nous pouvons reformuler l'objectif de nos recherches dans un langage computationnel et dynamique.

2.1 Propriétés physiologiques et fonctionnelles du cortex

Le cerveau est généralement considéré comme étant la source génératrice du comportement intelligent chez les organismes vivants évolués. Toutefois, l'intelligence n'est pas une caractéristique réservée exclusivement aux espèces disposant d'un cerveau. Les plantes, les micro-organismes, ou même les molécules individuelles² seraient capables de manifester un tel comportement. En outre, nous pouvons constater souvent que le simple fait de disposer d'un cerveau ne garantit pas l'engendrement d'un comportement intelligent. Nous pourrions nous demander ainsi pourquoi essayer de concevoir des systèmes artificiels intelligents, basés sur les principes fonctionnels cérébraux, alors que les connaissances scientifiques actuelles sur le fonctionnement du cerveau restent largement incomplètes ? À part les raisons (et les désirs) anthropomorphiques, souvent manifestées dans les activités créatrices humaines (culturelles ou scientifiques), nous pouvons énumérer d'autres arguments. Ainsi, en poursuivant une démarche constructiviste du cerveau (construire pour comprendre), nous serions capable de valider et d'enrichir les connaissances sur son mode de fonctionnement. Du point de vue épistémologique, ceci suivrait également la devise socratique de la connaissance de soi. Un autre argument, d'ordre pragmatique, se réfère au fait que le développement évolutif du cerveau est en directe corrélation avec les capacités adaptatives des espèces : plus le cerveau est développé, plus l'espèce qui le possède est capable d'adapter son comportement, face aux divers facteurs externes. La série des arguments peut bien sûr s'élargir, mais nous considérons les aspects exprimés ici suffisamment pertinents pour accorder attention à ce type d'approche scientifique. Bien que cette démarche ne vise pas à reconstruire le cerveau en le simulant en laboratoire (le projet ambitieux *Blue Brain* [78] ayant toutefois exactement ce but), elle propose une piste de recherche, en s'inspirant des solutions que la nature a conçues et a validées au cours de l'évolution. De ce point de vue, la source principale d'inspiration dans notre travail de recherche a été le cortex cérébral.

Le cortex cérébral est un tissu neuronal multi-couches (six couches chez l'homme), couvrant la surface extérieure des hémisphères cérébraux. Chez l'homme, le tissu cortical a une épaisseur variant entre 3 et 4 mm, une superficie d'environ 2600 cm², un nombre approximatif de 28×10^9 neurones (d'un total estimé de 10^{11} neurones dans le cerveau humain), et un nombre de connexions synaptiques d'environ 10^{12} (d'un total estimé à 10^{14} synapses dans le cerveau humain) [87]. Du point de vue évolutif, le cortex cérébral est une des structures anatomiques les plus récentes apparues en cours de l'évolution ; elle n'est mise en évidence que chez les mammifères.

Regardant ces aspects, l'on pourrait nous demander les raisons pour lesquelles nous avons choisi directement le cortex comme source d'inspiration pour nos travaux, sans avoir pris en compte les mécanismes cérébraux « primaires » (c'est-à-dire les mécanismes cérébraux fondamentaux, communs à toutes les espèces possédant un cerveau). On pourrait également nous accuser de vouloir doter, à tout prix, nos modèles théoriques de traitement d'information de certaines « caractéristiques anthropomorphiques ». Par conséquent, on pourrait nous recommander de commencer nos investigations scientifiques avec les mécanismes cérébraux des espèces moins évoluées, qui seraient plus faciles à comprendre et à modéliser théoriquement. Face à ces recommandations pertinentes, nous mentionnons que ces sources d'inspiration font déjà l'objet d'une multitude de recherches expérimentales et théoriques. Dans ce contexte, nous notons par

²Par exemple, Lagzi et al. ont montré dans [74] comment la chimiotaxie des molécules peut être utilisée pour trouver la solution d'un labyrinthe. Dans leur expérience, les parois du labyrinthe sont couverts par une substance chimique (les endroits différents du labyrinthe ont des valeurs différentes de pH, et l'endroit avec le pH minimum désigne la sortie du labyrinthe). En contact avec ces parois, une molécule d'une autre nature chimique placée dans le labyrinthe se déplace vers l'endroit ayant le pH minimum. Selon les résultats expérimentaux, les déplacements effectués par ce type de molécules sont souvent (quasi) optimaux.

exemple les résultats expérimentaux concernant l'étude du système nerveux du nématode *Caenorhabditis elegans*. Grâce à ces travaux, il a été récemment montré que le système nerveux chez les individus de cette espèce comprend précisément 302 neurones [73]. En outre, toutes leurs connexions synaptiques ont pu être cartographiées [122]. Toutefois, si ces résultats permettent de comprendre, par exemple, la coordination motrice nécessaire pour effectuer un déplacement dans l'environnement [108], ils ne peuvent pas directement nous fournir des explications sur les aspects cognitifs du comportement chez les espèces plus évoluées. Comme nous allons le détailler dans la suite, plusieurs caractéristiques fonctionnelles, propres à la structure corticale, ont conduit les chercheurs à lui attribuer un rôle fondamental dans la génération et la dynamique du comportement intelligent. Ces aspects constituent pour nos recherches un point de départ essentiel.

De nombreuses études ont mis en évidence le rôle déterminant du cortex cérébral dans le traitement de l'information visuelle [66], auditive [9], olfactive [126], tactile ou sensorimotrice, dans la représentation et la planification des actions [16], ainsi que dans la génération du langage [45] ou de la conscience [31]. Si cette multitude de processus mentaux dans lesquels le cortex est impliqué a constitué une des principales motivations pour la poursuite d'une recherche approfondie sur son fonctionnement, elle a constitué également une source d'inspiration pour la conception des modèles de traitement d'information.

2.1.1 Ségrégation anatomique du traitement cortical

Les expériences neurophysiologiques ont montré que, pendant la première période de la vie (chez l'homme, les premières années après la naissance), le cortex cérébral subit des transformations fonctionnelles importantes, résultant de l'expérience perçue par l'individu dans son interaction avec l'environnement extérieur. Ainsi, le cortex subit un processus de spécialisation fonctionnelle, formant différentes aires locales de traitement d'information. Selon la spécificité de l'information traitée, différentes aires corticales ont été mises en évidence et classifiées par Brodmann (voir sur la figure 2.1 certaines aires fonctionnelles du cortex). Comment une structure anatomique quasi homogène dans toute sa surface réussit à rendre possible un tel processus de spécialisation fonctionnelle reste un sujet d'intenses recherches. Toutefois, d'un point de vue informatique, cette propriété constitue une importante source d'inspiration. Nous pouvons postuler l'existence d'un mécanisme intrinsèque, de nature cortical, responsable d'un traitement générique d'information. Mettre en évidence, comprendre et modéliser différents aspects de ce mécanisme, représente une problématique de recherche partagée par la physiologie, les neurosciences et l'informatique, problématique que nous nous approprions également.

2.1.2 Plasticité corticale

On pourrait considérer comme hypothèse que le processus de spécialisation corticale est encodé génétiquement, et ce que nous observons après la naissance est exclusivement le déroulement ontogénique de cet encodage. Toutefois, l'encodage génétique de toute l'information nécessaire pour « retenir » la topologie des connexions neuronales, donnée par les 10^{12} synapses estimées dans la structure corticale humaine, dépasserait toute la capacité informationnelle de codage du génome. Il serait donc plus plausible à considérer que le code génétique ne contient qu'une série limitée « d'instructions générales », nécessaires pour ce processus. En effet, l'environnement et l'expérience ressentie par l'individu influencent en grande partie le processus de ségrégation (spécialisation) corticale. Par exemple, Hubel décrit dans [64] plusieurs expériences neurophysiologiques qui montrent les différences significatives entre la connectivité neuronale du

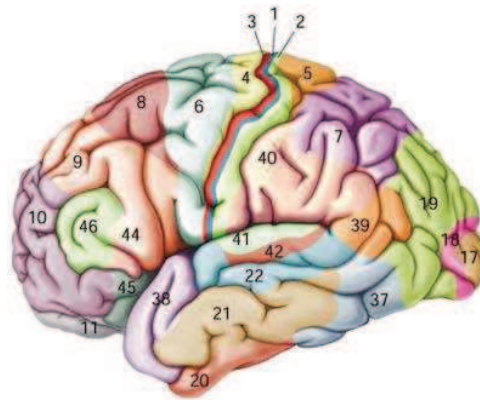


FIG. 2.1 – Les régions occupées par différentes aires Brodmann du cortex cérébral humain. Par exemple, les aires 1, 2 et 3 définissent le cortex primaire somatosensoriel, responsable pour le traitement des perceptions liées au sens tactile ; l'aire 4 représente le cortex moteur primaire, impliqué dans le traitement du mouvement des muscles et des articulations ; les aires Brodmann 17 (appelée aussi V1, ou cortex visuel primaire), 18 (appelée aussi V2, ou cortex visuel secondaire), 19 (appelée aussi V3, ou cortex visuel associatif) abritent les premières étapes du traitement de l'information visuelle ; les aires 44 et 45 définissent ensemble l'aire de Broca, responsable pour une partie importante du traitement du langage parlé, etc. Image extraite du site web *The Brain From Top to Down* (<http://thebrain.mcgill.ca>).

cortex visuel primaire chez les chatons ou les macaques nouveau-nés, qui ont eu un de leurs yeux suturé après la naissance (ou leurs stimuli visuels altérés), par rapport aux individus suivant un développement neuronal normal. Ces résultats montrent que la structure corticale n'est pas déterminée entièrement génétiquement (ou « câblée ») pour un individu, elle est plutôt modelée par plusieurs facteurs externes. D'autres expériences ont également montré que ce processus de spécialisation, loin de s'arrêter peu après la naissance, continue d'opérer pendant toute la vie de l'individu. Ainsi, des études comme [33, 107, 44, 96] ont montré une réorganisation assez rapide du cortex, en cas de lésions cérébrales, ou après un changement significatif subi par l'individu au niveau perceptif (par exemple suite à l'amputation d'un membre ou suite à un entraînement soutenu pour mieux discriminer les sources et les qualia des perceptions). Est également bien connu le fait que, par rapport aux autres personnes, les personnes aveugles « recrutent » certaines parties des aires corticales du traitement visuel pour les utiliser pour des traitements perceptifs d'une autre nature (par exemple sensorimotrice, auditive, olfactive etc.) [34]. Cette propriété remarquable du cortex, à la fois de se spécialiser sous aspect fonctionnel et de se réadapter aux changements signalés via des sources extra-corticales, définit la *plasticité corticale*. Du point de vue informatique, elle soulève une fois de plus l'intérêt pour la modélisation d'un mécanisme générique et adaptatif de traitement d'information.

La plasticité corticale se traduit par des changements au niveau structurel ou neurophysiologique des connexions neuronales. Ainsi, le processus cortical de spécialisation et d'adaptation peut engendrer la création ou la suppression des connexions entre différents neurones, ainsi que des changements concernant le flot des neurotransmetteurs au niveau synaptique de ces connexions. Ce dernier type de changement, qui se réfère à l'influence manifestée par l'intensité d'activation d'un neurone vis-à-vis d'autres neurones avec lesquels il est connecté, est

généralement modélisé par les *poids des connexions synaptiques*. Bien évidemment, un poids de connexion synaptique nul entre deux neurones équivaut à l'absence de connexion entre ceux-ci. Cette abstraction permet ainsi d'inclure aussi dans la modélisation le premier aspect de la plasticité corticale évoqué ci-dessus. Le processus d'adapter (en permanence) les poids des connexions synaptiques neuronales, au cours de l'interaction de l'individu avec l'environnement perçu, porte le nom de *plasticité synaptique*. Dans un cadre plus général, les changements continus au niveau neuronal, qui conduisent à une adaptation fonctionnelle de la structure corticale, portent le nom d'*apprentissage neuronal*. Différents modèles théoriques ont été proposés au cours du temps pour décrire les mécanismes intrinsèques de l'apprentissage neuronal de nature corticale. Un bref passage en revue de ceux-ci est présenté dans la section 2.4 de ce chapitre.

2.1.3 Cartes corticales

Afin d'explorer les bases neuronales de l'apprentissage cortical, il est nécessaire d'examiner dans un premier temps comment les neurones répondent aux stimuli perceptifs élémentaires. Autrement dit, il est nécessaire d'étudier la corrélation entre un stimulus et les propriétés des champs récepteurs et des motifs d'activation des neurones. Par *champ récepteur* d'un neurone on comprend la zone de l'espace dans laquelle la présence d'un stimulus peut influencer son activité. Par exemple, le champ récepteur d'un neurone du cortex visuel primaire représente la zone du champ visuel qui, une fois stimulée, déclenche une réponse d'activation de la part du neurone. Grâce aux avancées technologiques, qui ont rendu possible l'enregistrement in vivo ou in vitro des activités électriques des cellules neuronales individuelles, l'activité de différents neurones du tissu cortical a pu être enregistrée et analysée en détail, dans plusieurs cas de stimulation perceptive.

Dans une série de travaux expérimentaux de référence [63, 65, 66], Hubel et Wiesel ont mis en évidence le fait que des neurones du cortex visuel primaire (l'aire Brodmann 17 ou V1) répondent d'une manière sélective aux différents stimuli visuels. De plus, les neurones montrent une spécificité d'activation selon le contexte de stimulation ou les propriétés des stimuli. Par exemple, une première expérience met en évidence l'existence des neurones dans V1 qui répondent avec une forte fréquence d'activation seulement quand la rétine est stimulée avec une tache lumineuse ayant une orientation particulière. Dans une autre expérience, Hubel et Wiesel montrent que l'activation des mêmes neurones dépend aussi de la provenance du stimulus visuel dans le champ visuel (ainsi, certains neurones ne s'activent que si le stimulus visuel provient par exemple de l'œil gauche). Ces neurones ne sont pas séparés spatialement, mais ils sont répartis sur toute la surface du cortex visuel primaire. La distribution spatiale de la sélectivité des neurones corticaux, vis-à-vis d'une certaine *modalité* (caractéristique ou propriété fonctionnelle) d'un stimulus, porte le nom de *carte corticale* de cette modalité. Selon la modalité du stimulus, différentes cartes corticales ont été mises en évidence : carte d'orientation (voir la figure 2.2 (a)), carte de dominance oculaire (voir la figure 2.2 (b)), carte de sélectivité de couleurs, carte de préférence de direction du mouvement du stimulus visuel, etc. Une caractéristique supplémentaire de cette organisation en carte corticale est le fait que la relation topologique entre deux stimuli visuels distincts, situés dans l'espace bidimensionnel de la rétine, est conservée dans l'espace bidimensionnel du tissu cortical. Autrement dit, deux stimuli rétinien voisins entraînent l'activation de neurones voisins dans la structure corticale. Cette propriété de projection topographique de l'espace visuel rétinien sur l'espace bidimensionnel cortical visuel porte le nom de *rétinotopie*.

Explorant horizontalement (le long de la surface corticale) les réponses d'activation des neurones corticaux de la zone V1, Hubel et Wiesel remarquent que la sélectivité de ces neurones change graduellement, et non pas d'une manière aléatoire. Comme nous pouvons l'observer dans

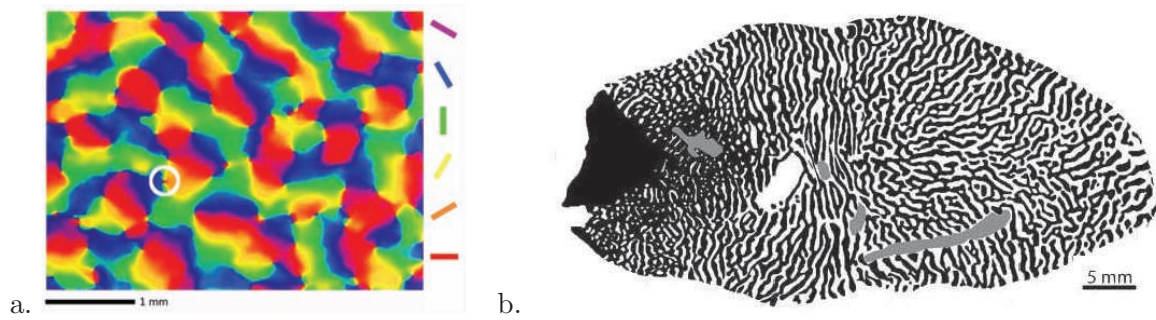


FIG. 2.2 – Cartes corticales d’auto-organisation, mises en évidence dans le cortex visuel primaire d’un macaque. a. *Carte de préférence d’orientation*. Les couleurs de la carte correspondent aux préférences de chaque neurone vis-à-vis de l’orientation d’une barre lumineuse qui stimule son champ récepteur. Si dans leur majorité, les neurones montrent une seule préférence d’orientation, nous remarquons néanmoins l’existence d’ensembles de neurones (appelés en anglais *pinwheel formations* ou *saddle points*), qui regroupent dans leur voisinage des neurones ayant tout le spectre de préférences d’orientation visuelle (un tel exemple est marqué par le cercle blanc sur cette figure). b. *Carte de dominance oculaire*. Les couleurs (blanc ou noir) correspondent aux préférences d’activation des neurones de V1, vis-à-vis du positionnement d’un stimulus lumineux fixe, alternativement dans le champ visuel de l’œil gauche ou droit. La zone entièrement blanche correspond au point aveugle, et la zone entièrement noire correspond à la périphérie du champ visuel. Images extraites du site web http://www.scholarpedia.org/article/Visual_map.

la figure 2.2 (a), la préférence d’orientation des neurones du cortex visuel primaire varie généralement peu d’un neurone à l’autre, avec l’exception notable de certains groupes de neurones où, dans le même voisinage anatomique, se trouvent des neurones ayant n’importe quelle préférence d’orientation. En fait, c’est cette représentation ordonnée (régulière) de la sélectivité neuronale le long de la structure corticale qui donne le caractère de *carte* à ce type de spécialisation corticale.

2.1.4 Colonne corticales

D’un autre point de vue, examinant verticalement (en profondeur) le tissu cortical, Hubel et Wiesel mettent en évidence le fait que les neurones des différentes couches corticales, qui se projettent sur la même position de l’espace bidimensionnel du tissu cortical, maintiennent dans leur majorité la même sélectivité (préférence) vis-à-vis du stimulus visuel actif dans leur champ récepteur. Ainsi, cette conservation verticale de la spécialisation neuronale exhibe une organisation en *colonnes corticales* de ce tissu nerveux (voir la figure 2.3). Le même type d’organisation en colonnes avait été évoquée précédemment par Mountcastle pour le cortex somatosensoriel [86], et depuis ces recherches, elle a été confirmée également pour beaucoup d’autres aires corticales. Aujourd’hui, ces formations portent le nom de *mini-colonne corticales*, alors que l’ensemble compact de mini-colonnes qui exhibent toutes les valeurs de préférence vis-à-vis de la même modalité du stimulus porte le nom de *hyper-colonne corticale*. Selon Mountcastle [87], une mini-colonne contient généralement 80 – 100 neurones, alors qu’une hyper-colonne contient environ 50 – 100 mini-colonnes corticales, couvrant un diamètre d’environ 500 – 750 μm .

Inspirés par ces résultats expérimentaux, Hubel et Wiesel ont postulé que chaque petite zone

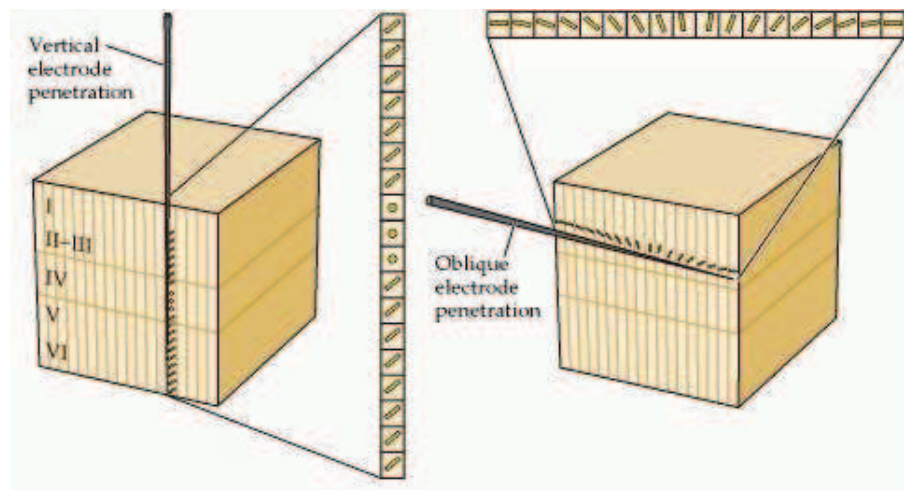


FIG. 2.3 – Illustration graphique de l'organisation en colonnes du cortex visuel primaire. Si une exploration horizontale par l'introduction d'une électrode met en évidence une variation graduelle de la préférence d'orientation des neurones, une exploration verticale du tissu cortical met en évidence la même préférence d'orientation. Les petits cercles sur la figure gauche symbolisent l'absence de préférence d'orientation pour les neurones de la couche IV du cortex visuel primaire. Image extraite de [94].

du cortex visuel primaire contient les « ingrédients » élémentaires nécessaires pour le traitement de la scène visuelle perçue par les deux rétines, qui sont ensuite utilisés par les étages corticaux supérieurs du traitement visuel. Cette vision, illustrée schématiquement dans la figure 2.4, porte le nom du *modèle du cube de glace* (en anglais *ice cube model*). Toutefois, suite à des recherches plus récentes, cette vision a été écartée, le traitement cortical n'étant plus considéré comme un « empilement hiérarchique » de plusieurs étages de traitement (ce qui en anglais est nommé par *top-down*), mais plutôt comme une interaction complexe entre différentes zones corticales. Ainsi, cette nouvelle vision considère le traitement cortical comme étant un traitement de type *bottom-up*, où le rôle déterminant est attribué aux boucles de feedback entre les différentes zones corticales [14].

Les résultats de Hubel et Wiesel ont ouvert la voie à une multitude de recherches sur le fonctionnement cortical. L'existence des cartes corticales dans d'autres aires corticales (auditives, olfactives, somatosensorielles, sensorimotrices, associative, etc.) est également confirmée aujourd'hui [127, 96, 6]. Toutes ces cartes corticales conservent la topologie de l'espace des stimuli perceptifs externes et montrent une organisation sous forme de colonnes corticales. Grâce à cette capacité du cortex à s'organiser selon différentes propriétés (modalités) du flot d'information perceptive, les cartes corticales sont aussi appelées *cartes d'auto-organisation corticale*.

L'organisation verticale en colonnes corticales et l'organisation horizontale en cartes topographiques montrent un caractère générique de la spécialisation fonctionnelle de la structure corticale. Cette remarque a conduit des chercheurs comme Mountcastle ou Burnod à considérer l'hypothèse que les colonnes corticales représentent les unités anatomiques discrètes du « calcul cortical » ou, autrement dit, elles sont les modules fonctionnels élémentaires du traitement de l'information au niveau cortical [87, 16]. De cette perspective, le cortex pourrait être vu comme un pavage bidimensionnel d'unités de « calcul neuronal élémentaire » ou, pour faire référence à

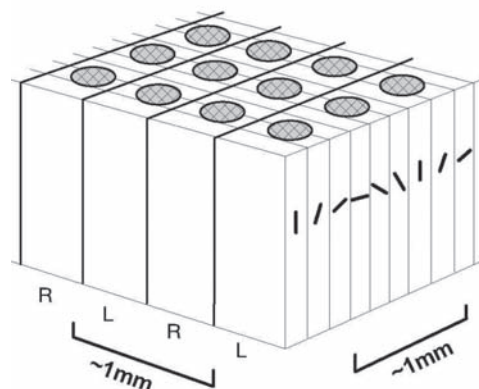


FIG. 2.4 – Illustration graphique du *modèle du cube de glace*, proposé par Hubel et Wiesel dans [64] comme modèle d’organisation du cortex visuel primaire. Ce modèle postule que chaque zone locale de V1 contient toutes les unités fonctionnelles élémentaires du traitement cortical, illustrées ici par les colonnes de préférence d’orientation (symbolisées sur la face droite du cube), les colonnes de dominance oculaire (symbolisées sur la face gauche du cube), et les colonnes de préférence de couleur (symbolisées par les petites ellipses sur la face supérieure du cube). Toutefois, ce modèle ne réussit pas à expliquer l’intégration d’autres caractéristiques corticales, et entre en contradiction avec l’influence des traitements récurrents mis en évidence dans la structure corticale. Image extraite et adaptée de [116].

Burnod, un hypercube en six dimensions, décrit par la topologie de connexion des unités des six couches corticales regroupées spatialement en colonnes corticales. Si cette idée séduisante par sa simplicité a influencé considérablement les approches neuroscientifiques ultérieures (qui visaient expliquer le fonctionnement cortical), elle s’est vue récemment remise en cause par plusieurs résultats expérimentaux [62]. Horton et Adams mettent en évidence le fait que certaines espèces ne développent jamais des cartes corticales d’orientation. Paradoxalement, chez d’autres espèces seulement un sous-ensemble d’individus développent ce type d’organisation corticale, ou encore plus étonnamment, seulement une partie de leur cortex exhibe cette propriété. De plus, à ce jour, les recherches expérimentales comparant ces espèces avec les autres (présentant des colonnes corticales) n’ont pas pu mettre en évidence une différence significative d’acuité perceptive ou d’autre capacité sensorimotrice. Ces aspects ont conduit Horton et Adams à prendre en compte la possibilité que l’organisation en colonnes du cortex puisse être un épiphénomène cortical (c’est-à-dire qui n’est pas conçu par un processus d’adaptation). Étayée par la vue que l’évolution est d’abord un processus permanent d’exploration du possible, et seulement ensuite un processus d’optimisation structurale [68], on pourrait considérer que l’évolution exploitera, au cours du temps, l’organisation en colonnes du cortex cérébral, pour lui attribuer une fonction spécifique. Malgré ces aspects fondamentaux, mis en cause par Horton et Adams, ces auteurs confirment néanmoins que, souvent, les neurones partageant les mêmes caractéristiques vis-à-vis de leurs champs récepteurs sont regroupés spatialement, et les connexions entre ces groupes neuronaux sont généralement périodiques. Ces aspects de généricité et périodicité, soulignés par cette vision, constituent pour nos recherches inspirées du traitement cortical, deux éléments fondamentaux pour la modélisation corticale.

Même si l’organisation en colonnes corticales et le développement de certaines cartes corticales commence dans la période prénatale et se déroule même en absence d’une expérience

perceptive directe entre l'environnement et l'individu [64], l'interaction directe entre ces deux entités joue un rôle déterminant dans ce développement cortical. Ce fait indique que l'organisation fonctionnelle corticale est le résultat d'un processus d'apprentissage neuronal.

Pour s'atteler à la problématique de la modélisation des processus neuronaux en général et corticaux en particulier, les chercheurs ont proposé au cours des années différents paradigmes et formalismes théoriques. Comme nous l'avons affirmé dans l'introduction de ce mémoire, notre démarche de modélisation partage une vision générale de modélisation des processus cognitifs, basée sur les approches computationnelle, connexionniste et dynamique. Dans la suite de ce chapitre nous décrivons d'abord les architectures générales de réseaux neuronaux, présentant ensuite comment le fonctionnement d'un modèle neuronal peut être considéré comme un système dynamique, capable d'engendrer un comportement dynamique émergent et d'implémenter un traitement à la fois computationnel et dynamique de l'information. Ces éléments théoriques fondateurs nous permettront d'introduire dans la section 2.4 une architecture neuronale capable d'offrir une possible modélisation du processus d'auto-organisation de nature corticale.

2.2 Paradigmes de modélisation neuronale

Le paradigme connexionniste considère que le fonctionnement du système nerveux peut être décrit par les processus émergents de l'interaction d'un ou plusieurs réseaux d'unités élémentaires modélisant l'activité des cellules nerveuses. Dans cette vue, un modèle neuronal est souvent défini par un triplet de type (X, u, w) , où X représente l'ensemble des unités élémentaires du réseau, et pour chaque unité $x \in X$ et chaque instant de temps t , $u(x, t)$ représente l'activité de l'unité x , et $w(x, y, t)$ représente la force d'interaction entre cette unité et une autre unité, y . Souvent, la partie (ou le total) des unités $\{x \in X\}$ qui peut recevoir le flot informationnel externe (perceptif) à traiter définit l'*entrée* du modèle. Dans le même contexte de modélisation, $w(x, y, t)$ porte le nom de *poids des connexions synaptiques*.

Les aspects qui différencient les modèles neuronaux entre eux concernent à la fois : le formalisme décrivant l'activité u des unités élémentaires (qui modélisent généralement l'activité des neurones ou groupes de neurones), la topologie des connexions établies entre ces unités, et les mécanismes qui ajustent dynamiquement les paramètres libres du modèle.

Bien qu'au niveau théorique toutes les architectures de réseaux neuronaux sont représentées par des structures de graphes, dans la plupart des cas, ces modèles neuronaux mettent en évidence une structure hiérarchique de traitement de l'information. Ainsi, pour distinguer le flot de traitement entre les différentes entités d'un modèle neuronal, les chercheurs ont proposé une taxonomie pour caractériser les connexions synaptiques entre les unités d'un réseau de neurones. Une unité neuronale peut recevoir des entrées provenant d'autres zones de traitement, hiérarchiquement inférieures, via des *connexions afférentes* ou *feedforward*. Une unité neuronale peut recevoir également des entrées provenant des zones hiérarchiquement supérieures de traitement (étages ultérieurs de traitement), via des *connexions de feedback*. Enfin, les connexions entre les unités neuronales appartenant à la même zone hiérarchique de traitement sont dénommées des *connexions latérales*. Dépendant de la distance (euclidienne) entre ces unités, les connexions latérales sont parfois différenciées entre *connexions latérales à courte distance* et *connexions latérales à longue distance*. Cette taxonomie des connexions synaptiques est employée également pour décrire la structure des connexions corticales (voir pour plus de détails [2]).

Cette classification des connexions synaptiques permet de distinguer différentes approches de modélisation neuronale. Dans la suite, nous exposons brièvement les architectures neuronales employées le plus souvent par ces approches.

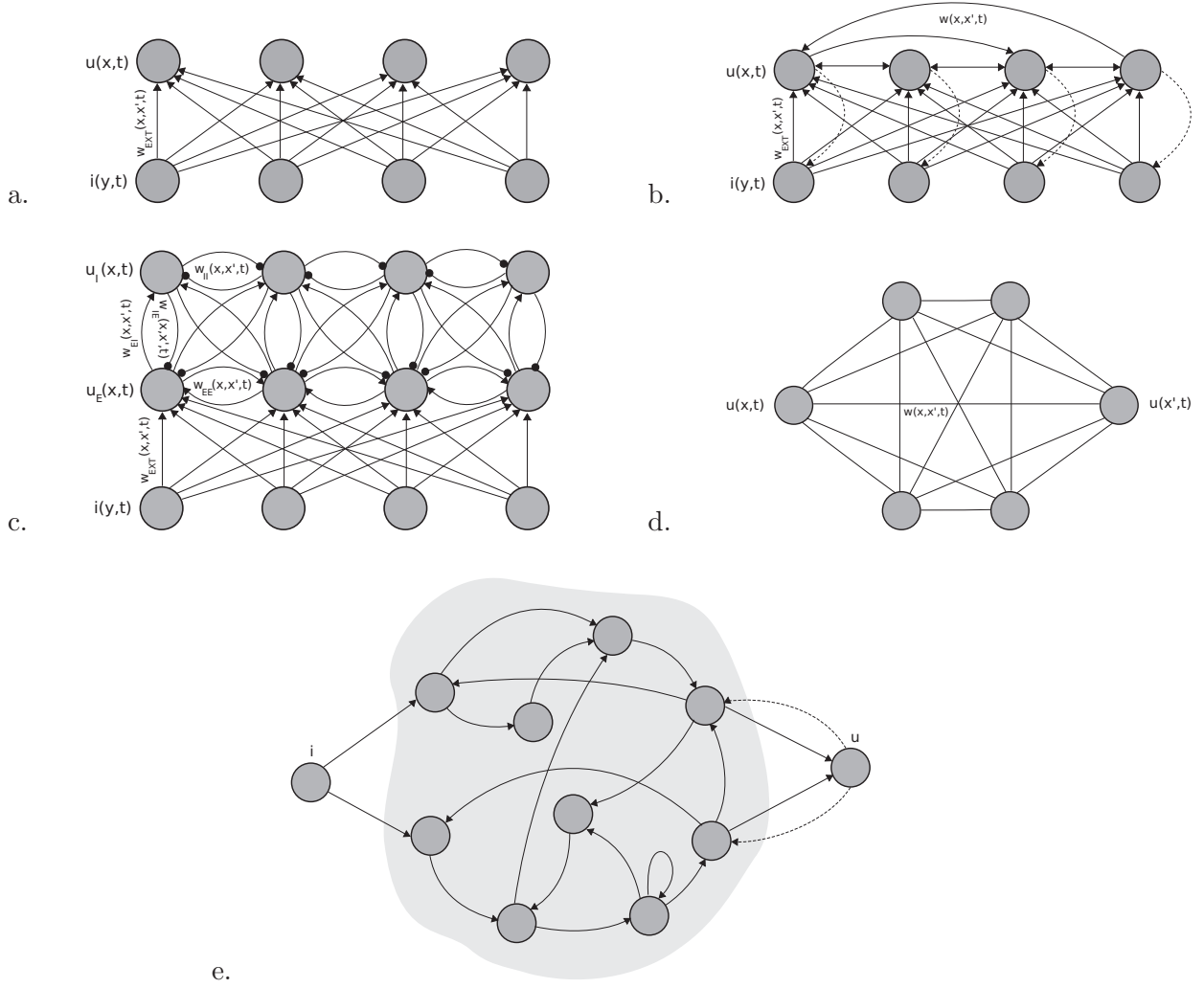


FIG. 2.5 – Différentes architectures de réseaux neuronaux. a. *Réseau feedforward*. Ce type de réseau a une structure strictement hiérarchique. b. *Réseau récurrent*. Ce type de réseau permet l'existence de connexions latérales à l'intérieur d'une couche neuronale (ici la couche supérieure), et éventuellement des connexions de feedback entre les unités des différentes couches. c. *Réseau récurrent avec des unités de type excitatrices/inhibitrices*. Ce type de réseau prend en compte, à part une couche d'entrée, deux couches neuronales distinctes (une exclusivement excitatrice et une exclusivement inhibitrice), interagissant via des connexions récurrentes. Pour raisons de clarté les connexions latérales à longue distance ne sont pas indiquées sur la figure. d. *Réseau de type mémoire associative de Hopfield*. Dans ce cas d'architecture, toutes les unités sont connectées entre elles, et leurs poids de connexion sont symétriques. e. Réseau de type *réservoir*. L'entrée du système est reflétée dans l'état global d'un réservoir, qui est un réseau récurrent d'unités neuronales avec des connexions synaptiques aléatoires (symbolisé sur la figure comme l'ensemble d'unités sur le fond gris). Dans certaines variantes de modélisation, la sortie du réservoir peut considérer des connexions de feedback avec le réservoir ou avec l'entrée du réseau.

2.2.1 Réseaux feedforward

Comme le montre la figure 2.5 (a), ce type de modèle neuronal repose sur une architecture neuronale complètement hiérarchisée, où les connexions entre les neurones sont exclusivement de nature afférente (feedforward). De manière générale, cette approche de modélisation considère une séparation des unités élémentaires en deux couches : une couche d'unités d'entrée, et une couche d'unités de sortie. La dynamique de l'activité de chaque unité de sortie x évolue selon une loi similaire à celle-ci :

$$\tau \frac{du(x, t)}{dt} = -u(x, t) + f\left(\sum_y w_{\text{EXT}}(x, y, t) i(y, t)\right) \quad (2.1)$$

où τ est une constante de temps, $i(y)$ représente l'activité de l'unité d'entrée y , et f est une fonction d'activation (souvent non-linéaire), qui modélise l'influence cumulative de la couche d'entrée ressentie par une unité de sortie. L'indice EXT du poids $w_{\text{EXT}}(x, y, t)$ vient souligner le fait que la connexion synaptique entre les unités x et y est de type feedforward.

Malgré la simplicité de ce type d'architecture, elle permet l'implantation de plusieurs modèles de traitement neuronal.

2.2.2 Réseaux récurrents

Par rapport aux réseaux feedforward, les modèles de réseaux récurrents reposent sur une architecture qui prend en compte des connexions latérales entre les unités appartenant à la même couche neuronale (voir la figure 2.5 (b)). Souvent, ces modèles intègrent également des connexions de feedback entre les unités des différentes couches. Par conséquent, la loi qui règle la dynamique des activités neuronales peut généralement s'exprimer ainsi :

$$\tau \frac{du(x, t)}{dt} = -u(x, t) + f\left(\sum_y w_{\text{EXT}}(x, y, t) i(y, t) + \sum_{x'} w(x, x', t) u(x', t)\right) \quad (2.2)$$

Bénéficiant des caractéristiques dynamiques plus riches qu'un réseau de type feedforward, ce modèle général de réseau neuronal constitue la fondation des architectures d'une partie importante des modèles neuronaux actuels.

2.2.3 Réseaux d'unités excitatrices/inhibitrices

Étayé expérimentalement par plusieurs études physiologiques, Dale avait conjecturé dans [27] un principe qui se réfère à la nature des cellules neuronales. Ce principe, connu aujourd'hui sous le nom de la *loi de Dale* [30], affirme que chaque cellule neuronale a une influence soit excitatrice, soit inhibitrice, et elle est la même vis-à-vis de toutes les cellules neuronales post-synaptiques avec lesquelles elle est connectée. De plus, la nature de cette influence reste la même pendant toute la durée de vie des cellules.

Un cas particulier de réseaux récurrents inspirés par ce principe considère une architecture neuronale qui, à part la couche d'entrée, prend en compte deux autres couches d'unités : une pour les unités excitatrices, et une pour les unités inhibitrices. La figure 2.5 (c) illustre graphiquement ce type de modèle.

Dans ce contexte, l'équation 2.2, qui décrit la dynamique générale des activités des unités

d'un réseau récurrent, se transforme en un système d'équations couplées :

$$\begin{aligned} \tau_E \frac{du_E(x, t)}{dt} = & - u_E(x, t) \\ & + f\left(\sum_y w_{\text{EXT}}(x, y, t) i(y, t)\right) \\ & + \sum_{x'} w_{\text{EE}}(x, x', t) u_E(x', t) - \sum_{x'} w_{\text{EI}}(x, x', t) u_I(x', t) \end{aligned} \quad (2.3)$$

$$\begin{aligned} \tau_E \frac{du_I(x, t)}{dt} = & - u_I(x, t) \\ & + f\left(\sum_{x'} w_{\text{IE}}(x, x', t) u_E(x', t) - \sum_{x'} w_{\text{II}}(x, x', t) u_I(x', t)\right) \end{aligned} \quad (2.4)$$

où il est considéré que seulement la couche excitatrice reçoit des connexions afférentes de la couche d'entrée, et $w_{\text{EE}, \text{EI}, \text{IE}, \text{II}}$ représentent les poids synaptiques modélisant l'efficacité des interactions entre les deux types d'unités, excitatrices et inhibitrices.

Cette architecture peut être employée pour décrire la structure corticale d'un point de vue mésoscopique, considérant le tissu cortical comme un ensemble d'unités neuronales excitatrices ou inhibitrices, et l'activité corticale comme le résultat de la dynamique émergente des interactions permanentes entre ces unités. Comme nous allons l'examiner dans le chapitre 3, cette architecture a conduit au cours des années au développement des modèles de champs neuronaux dynamiques, comme proposition de modéliser la topologie et l'activité de la structure corticale. Cette approche, qui intègre à la fois le paradigme de modélisation connexionniste et celle de type système dynamique, constitue le point de départ déterminant dans notre démarche de modélisation bio-inspirée.

2.2.4 Réseaux de type mémoire associative

Un autre modèle de réseau neuronal récurrent, entièrement non-hiérarchisé, a été proposé par Hopfield [60] (voir la figure 2.5 (d)), et est considéré aujourd'hui comme un modèle neuronal de référence pour toute une suite d'autres modèles neuronaux.

Dans ce modèle, à chaque nouvel instant de temps, l'activité de chaque unité du réseau change en fonction de son activité courante et de l'activité de toutes les autres unités du réseau avec lesquelles elle est connectée, selon la règle suivante :

$$u(x, t + 1) = \begin{cases} 1 & \text{si } \sum_{x'} w(x, x', t) u(x', t) + \theta(x) > 0 \\ 0 & \text{sinon} \end{cases} \quad (2.5)$$

où $\theta(x)$ est un seuil neuronal d'activation, propre à chaque unité x du réseau. Notons qu'une variante continue de cette règle peut être envisagée, similaire aux équations 2.1-2.4.

L'ordre de mise à jour des activités de l'ensemble des unités du réseau est entièrement aléatoire, et non pas imposé par un mécanisme centralisé d'évaluation. Cette évaluation asynchrone des activités des unités du réseau accorde à ce modèle une forte plausibilité biologique.

Le modèle proposé par Hopfield associe à cette règle de mise à jour une fonction d'énergie, définie comme suit :

$$E = -\frac{1}{2} \sum_{x, x'} w(x, x', t) u(x, t) u(x', t) - \sum_x u(x, t) \theta(x) \quad (2.6)$$

Considérons ensuite un tel réseau, où toutes les connexions synaptiques sont symétriques et les unités ne sont pas connectées directement à elles-mêmes. Dans ce contexte, Hopfield montre qu’une mise à jour asynchrone du réseau (suivant l’équation 2.5) garantit la convergence de tout état initial du réseau vers un état qui correspond à un minimum local de la fonction d’énergie E . Ces états sont des états stables du comportement dynamique du réseau.

Grâce à cette propriété, ce modèle neuronal peut être utilisé pour concevoir un mécanisme de *mémoire associative*. Soit $M = \{s_1, s_2, \dots, s_n\} \subset \{0, 1\}^{card(X)}$, une liste de n représentations symboliques, correspondant à n objets que nous souhaitons mémoriser. Soit w tel que M fasse partie de l’ensemble des états du réseau, pour lesquels la fonction E atteint un minimum local. Le calcul pour trouver les poids synaptiques nécessaires pour satisfaire à cette condition doit se réaliser selon une procédure d’apprentissage (voir la section 2.4). Selon la propriété ci-dessus énoncée, les états de M sont stables. Par conséquent, imposer au réseau un état quelconque $s \in M$ revient à « verrouiller » le réseau dans cet état s (tant qu’il est stable). Cet effet doit être interprété comme la reconnaissance du réseau que l’état (ou la représentation) s a été mémorisé.

Notons également que, grâce à la même propriété, il existe une forte probabilité que tout état s' , légèrement différent de s , converge vers s au cours du temps, s et s' appartenant souvent au bassin d’attraction du même minimum de la fonction d’énergie E . Cet aspect souligne la robustesse au bruit du modèle proposé par Hopfield, et affirme donc qu’il est possible de retrouver une information valide, mémorisée a priori, à partir d’une version incomplète ou altérée de celle-ci. Mentionnons aussi que la puissance de calcul de ce modèle neuronal est équivalente avec celle de la machine de Turing [104].

Bien que ce modèle de mémoire associative introduise plusieurs aspects novateurs, sa capacité de stockage est assez réduite (approximativement $0.15 \text{ card}(X)$), et parfois, ses états stables peuvent correspondre à des minima locaux ayant des valeurs de E assez élevées.

En dépit de ces limitations, il est généralement considéré que ce modèle neuronal a revitalisé l’intérêt de la communauté du domaine de l’intelligence artificielle pour les réseaux neuronaux. Suite à la parution de ce modèle, le paradigme des réseaux neuronaux est devenu un paradigme de modélisation viable, alternatif au paradigme de modélisation symbolique, largement dominant dans les approches théoriques qui ont précédé le modèle de Hopfield.

2.2.5 Réseaux stochastiques

Un modèle neuronal similaire au modèle de Hopfield ci-dessus analysé est l’ainsi nommée *machine de Boltzmann*. Dans le cas des deux modèles, l’activité d’une unité du réseau peut avoir seulement une des deux valeurs binaire, 0 ou 1, et la fonction d’énergie E pour chaque réseau est donnée par l’équation 2.6. De cette expression résulte que la quantité d’énergie apportée par l’activation d’une unité x équivaut à :

$$\Delta E(x, t) = \sum_{x'} w(x, x', t) u(x', t) + \theta(x) \quad (2.7)$$

L’aspect qui différencie ces deux modèles se réfère à la loi d’activation de leurs unités neuronales. Ainsi, l’activation de chaque unité x du réseau d’une machine de Boltzmann (c’est-à-dire $u(x, t) = 1$) peut se réaliser seulement avec une probabilité $p(x, t)$, où :

$$p(x, t) = \frac{1}{1 + e^{-\frac{\Delta E(x, t)}{T}}} \quad (2.8)$$

où T est une valeur scalaire positive, qui représente la « température » du réseau. Le terme de température provient d’une analogie avec un processus physique thermodynamique.

Comme nous pouvons le remarquer, la mise à jour des activités du réseau est imposée par une règle stochastique. Lorsque cette règle est appliquée de manière asynchrone pour les unités du réseau, pour un T fixé, la probabilité d'un état du réseau dépend de son énergie E . Sur le long terme, le réseau converge vers un « équilibre thermique », c'est-à-dire la distribution des probabilités des états du réseau décrit une distribution de Boltzmann (d'où le nom du modèle). Si T est abaissé graduellement, et si à chaque réduction, on laisse le réseau converger vers un équilibre thermique, le niveau d'énergie du réseau converge vers son minimum global. Ce processus de baisse progressive de T porte le nom de *recuit simulé* (en anglais *simulated annealing*). Notons que pour $T = 0$, la règle d'activation de la machine de Boltzmann devient entièrement déterministe, et le modèle neuronal devient par conséquent équivalent au modèle de Hopfield.

Si le modèle de Hopfield pouvait être utilisé pour construire un mécanisme de mémoire associative, la machine de Boltzmann peut être utilisée pour concevoir un mécanisme capable de générer une distribution statistique similaire à celle exhibée par un flot donné d'information. Autrement dit, ce modèle peut aider à générer une distribution de probabilités, à partir d'une série de motifs d'activation. Pour que les probabilités de voir un réseau converger vers un ensemble d'états soient dictées par une distribution externe donnée, il est nécessaire de trouver les poids synaptiques w tels que les états de convergence les plus probables aient les niveaux d'énergie les plus bas. Ce calcul peut être effectué par une procédure d'apprentissage (voir la section 2.4).

Par rapport au modèle de Hopfield, la machine Boltzmann, grâce à sa règle d'activation stochastique, ne rencontre pas le problème des minima locaux trop élevés de la fonction d'énergie E . Cependant, le temps nécessaire pour apprendre une distribution de probabilités augmente de manière exponentielle avec la taille de la machine.

2.2.6 Réservoirs

Un dernier type de modèle de réseaux récurrents que nous présentons ici, est paru assez récemment dans la littérature, et porte le nom de *réservoirs* (en anglais *reservoir computing*).

Ce modèle considère un réservoir, c'est-à-dire un ensemble d'unités avec des connexions aléatoires (et souvent fixes), interposé entre un ensemble d'unités d'entrée et un d'unités de sortie (voir la figure 2.5 (e)). Chaque nouvelle stimulation de l'entrée du réseau induit une dynamique (non-linéaire) spatio-temporelle de l'activité globale du réservoir. Ces réponses du réservoir (ses états) peuvent être ensuite combinées (linéairement), pour construire des réponses adéquates en rapport avec la séquence des entrées présentées.

Selon ce principe fonctionnel du réservoir, une entrée, et par la suite toute une séquence d'entrées du réseau se reflètent dans l'état du réservoir (ou, pour employer une métaphore, les entrées trouvent leurs « échos » dans la dynamique du réservoir). Basés sur cet aspect dynamique, plusieurs modèles de réservoir ont été proposés dans la littérature : les machines à état liquide ou LSM (de l'anglais *liquid state machines*) [75], les réseaux à état écho ou ESN (de l'anglais *echo state networks*) [69], ou les réseaux à état dépendant SDN (de l'anglais *state-dependent networks*) [15]. Notons que, par rapport à l'architecture décrite ci-dessus, ESN inclut une boucle entre les sorties et les entrées, et LSM intègre l'ensemble des unités d'entrée à l'intérieur du réservoir.

Différentes règles d'apprentissage peuvent être envisagées pour ajuster les poids synaptiques des unités connectées avec le réservoir, afin de conduire les sorties ou des séquences des sorties (nommées des trajectoires) du réseau vers des états souhaités.

Si d'autres modèles généraux de réseaux visent à traiter (extraire et classifier) des caractéristiques spatiales d'un espace d'entrée, le modèle de réservoir vise à traiter surtout les carac-

téristiques temporelles de cet espace. En particulier, les réservoirs peuvent être utilisés en tant que mécanismes de mémoire à court terme.

Le paradigme de calcul neuronal employé par les réservoirs montre plusieurs avantages. Par exemple, le même réservoir peut être utilisé pour effectuer des traitements différents, et il peut effectuer ces traitements à des échelles de temps différentes. Ces propriétés font de ce modèle un candidat important pour modéliser certains aspects génériques du traitement cortical (par exemple les activités des neurones d’une mini-colonne corticale).

Toutefois, il faut noter que la généralité des réservoirs limite le contrôle explicite du réservoir. De plus, comme dans certaines études neurophysiologiques, l’analyse des sous-parties d’un réservoir ne conduit pas à une meilleure compréhension de son fonctionnement global.

Nous considérons que ce paradigme de traitement neuronal, encore jeune, est potentiellement un outil général et puissant de modélisation neuronale.

2.3 De la dynamique au calcul

Dans les sections précédentes nous avons pu observer que chaque nouvelle stimulation (entrée) d’un réseau neuronal induit une certaine dynamique dans l’activité globale de ces unités. Ces dynamiques sont simples, comme celles de type stimulus-réponse, ou complexes, caractérisées par des périodes transitoires et des états stables. Chaque modèle neuronal vise à utiliser, modifier et faire valoir ses dynamiques, pour leur donner une certaine sémantique. Le plus souvent, les opérandes de cette sémantique sont les *états stables* d’un système dynamique. Dans la suite, nous traçons les lignes principales qui décrivent différentes méthodes pour réaliser le passage d’un paradigme dynamique à un paradigme de calcul. Pour plus de détails sur ce vaste sujet, le lecteur est invité à se reporter par exemple à [52, 51, 15, 98, 21, 58].

Pour commencer, remarquons que la dynamique exhibée par un réseau neuronal peut être qualifiée relativement à une échelle de temps. À l’échelle temporelle de l’activation d’une unité (qui correspond au niveau physiologique au temps de décharge d’un potentiel d’action d’un neurone), la dynamique est exprimée en tant que trains d’impulsions électriques (dans les modèles de *spiking neurons*), ou en tant que fréquence d’activation (dans les modèles de *firing rate neurons*). À une échelle de temps plus grande, le réseau est caractérisé par des états stables et des états transitoires. En particulier, la transition d’un état de stimulation du réseau à un état stable est connue sur le nom de *processus de relaxation*. Notons que, d’un point de vue macroscopique, l’activation d’une unité neuronale, ou la stabilité de l’état d’un réseau neuronal, sont souvent des événements « discrets » de la dynamique exhibée par ce réseau.

2.3.1 Quelques notions de la théorie des systèmes dynamiques

Du point de vue conceptuel, l’évolution d’un système dynamique continu (en particulier d’un réseau neuronal) peut être souvent décrite par une équation différentielle générale de type :

$$\frac{da(t)}{dt} = F(a(t)) \quad (2.9)$$

où $a(t)$ caractérise l’état du système à l’instant t , et F est une fonction modélisant la mise à jour de cet état. Notons que $a(t)$ est souvent une représentation (vectorielle) de plusieurs paramètres qui définissent l’état d’un système. Ainsi, $a(t)$ peut retenir non seulement les états courants d’activation des unités d’un réseau, mais aussi leurs historiques (dans une fenêtre de temps), ou bien d’autres variables caractéristiques.

Considérons A , l'ensemble de tous les états possibles $a(t)$, dans lesquels le système peut évoluer. Dans la terminologie des systèmes dynamiques, A est appelé *l'espace des phases* du système dynamique considéré. L'évolution (déterministe) à partir d'un état quelconque a_0 , au cours du temps, définit *l'orbite* de cet état a_0 .

Un sous-ensemble de l'espace des états, vers qui les orbites des états initiaux convergent dans le temps, s'appelle un *attracteur*. Plusieurs types d'attracteurs peuvent être différenciés : le *point fixe* ou le *point d'équilibre*, qui est un état dans lequel le système persiste, une fois cet état atteint par une des orbites du système (par conséquent, $a(t)$ est un point fixe si $F(a(t)) = 0$) ; le *cycle limite*, qui est une orbite fermée, que le système est forcé de la parcourir périodiquement, une fois qu'il y est capturé ; d'autres types, comme par exemple *l'attracteur étrange*, qui est aussi une orbite, mais d'une topologie impossible à être exprimée dans un espace de dimension entière. L'ensemble des orbites qui convergent vers un attracteur donné est connu sous le nom de *bassin d'attraction* de l'attracteur.

Si ces notions aident à caractériser l'évolution d'un système dynamique, un autre aspect visé par l'étude dynamique d'un système concerne sa robustesse. Il est ainsi important d'examiner si les états vers lesquels un système converge (ses attracteurs) sont stables, c'est-à-dire si, suite à des perturbations de l'état, le système est capable de revenir dans l'attracteur. Les outils mathématiques les plus usitées dans un tel contexte d'analyse font appel à la théorie des bifurcations. Par définition, un *point de bifurcation* d'un paramètre v d'un système dynamique est une valeur de ce paramètre qui, perturbée avec une quantité δ , entraîne l'évolution du système dynamique vers des orbites topologiquement différentes que celle obtenue en absence de cette perturbation. Un point de bifurcation entraîne alors un changement qualitatif par rapport à la dynamique non-perturbée du système. Ce changement peut engendrer par exemple la création ou la suppression de certains attracteurs.

2.3.2 Le calcul comme un effet émergent d'un système dynamique

Les concepts introduits dans la section précédente aident à mettre en évidence et à analyser différents aspects qualitatifs clé concernant le comportement d'un système complexe, c'est-à-dire différents effets émergents qui caractérisent la dynamique globale d'un tel système. Nous nous référons ici aux effets *émergents* dans la mesure où la dynamique du système n'est pas issue d'une décision centralisée basée sur une évaluation de l'état global du système, mais elle est uniquement la conséquence des activités et des interactions de ces activités des unités élémentaires qui composent le système dynamique considéré.

D'un point de vue général, nous pouvons affirmer que la dynamique émergente d'un système dynamique complexe repose sur ses attracteurs et sur les transitions entre ces attracteurs. L'évolution dynamique entraînée par la stimulation externe du système définit son mécanisme de traitement d'information. Comprendre comment l'interaction des unités neuronales individuelles peut se traduire dans un mécanisme de traitement d'information au niveau cérébral représente un des défis majeurs des neurosciences. Le paradigme connexionniste de modélisation neuronale vise à proposer des modèles de réseaux neuronaux capables de reproduire non seulement des résultats « finaux » du traitement neuronal, mais aussi des dynamiques similaires à celles observées expérimentalement.

Par exemple, le modèle des réseaux de type excitation/inhibition (revoir la section 2.2.3) exploite la dynamique de compétition entre les activités excitatrices et inhibitrices de ses unités élémentaires. Le résultat de cette compétition locale se traduit souvent par la formation de ce que l'on nomme *bulles d'activité* (en anglais *bumps*), c'est-à-dire de groupes compacts d'unités qui restent actives, pendant que les autres unités participant à la compétition cessent

leurs activités. Ce phénomène porte le nom de compétition de type *le-gagnant-prend-tout* (en anglais *winner-take-all* ou WTA) si une seule unité reste active suite à la compétition, ou *le-gagnant-prend-presque-tout* (en anglais *winner-take-most* ou WTM), si plusieurs unités gardent un niveau d'activation non nul. Si les activités des unités gagnantes persistent dans le temps, c'est-à-dire restent en équilibre (au moins pour une certaine période de temps), ces configurations dynamiques du réseau constituent des attracteurs pour sa dynamique. La compétition neuronale est alors un des paradigmes possibles pour faire converger l'évolution d'un réseau vers un état stable (attracteur). Cependant, un tel mécanisme serait assez peu utile s'il permettait seulement une transition (relaxation) d'un stimulus initial du réseau vers un de ses attracteurs. Ce qui empêche de concevoir ces transitions comme un flot de traitement de type feedforward sont les connexions récurrentes (de feedback), qui peuvent affecter les activités des unités du réseau (gagnantes ou non), et implicitement les états dans lesquels le système réside déjà. Par conséquent, les connexions récurrentes peuvent entraîner la « sortie » du système de ses attracteurs, et la poursuite d'autres orbites dynamiques. Par exemple, les connexions récurrentes peuvent induire des perturbations dans les états des attracteurs, qui peuvent ainsi créer des points de bifurcation et conduire ces états vers d'autres orbites dynamiques. Cet enchaînement dynamique (état initial, transition, attracteur, transition, etc.) peut définir, comme indiqué ci-dessus, un mécanisme de traitement d'information.

Un autre exemple de traitement neuronal à base d'attracteurs est illustré par les modèles de réseau de Hopfield ou de machine de Boltzmann (revoir les sections 2.2.4 et 2.2.5). Les états stables de la dynamique de ces réseaux sont « construits » (par une procédure d'apprentissage) tels qu'ils correspondent aux minima locaux d'une fonction d'énergie E , ce qui implique qu'ils sont des points fixes (et implicitement des attracteurs) d'un système dynamique. Utilisant cette propriété, ces modèles ont pu employer une sémantique de type mémoire associative pour leurs états d'entrée : chaque nouvelle stimulation du réseau converge (de manière déterministe) vers un attracteur. De plus, une légère modification de la stimulation initiale du réseau conduit le plus probablement vers le même attracteur, étant donné que la similarité entre deux entrées est corrélée avec leur probabilité de se trouver dans le même bassin d'attraction. Les capacités de traitement d'information de ce mécanisme neuronal générique de mémoire associative peuvent s'étendre si nous considérons, comme dans le cas des réseaux de type excitation/inhibition détaillé ci-dessus, des mécanismes dynamiques pour explorer plusieurs attracteurs.

Une technique de modélisation similaire est souvent implantée avec l'aide des réservoirs (revoir la section 2.2.6). Pour « sortir » d'un attracteur, on peut considérer des variables locales supplémentaires (des réservoirs), qui étendent l'espace des phases de sa dynamique, et qui correspondent aux attracteurs initiaux (voir la figure 2.6). Dans ce contexte, les états de ces attracteurs peuvent être rendus instables. Par conséquent, ces états peuvent devenir des *attracteurs transitoires*, appelés aussi des *attracteurs ruines* (en anglais *transient attractors* ou *attractor ruins*). Notons que le système peut toujours converger vers ces états, si les variables de type réservoir changent lentement. Cependant, le changement d'orbite dynamique est cette fois-ci possible, étant réglé par la dynamique de ces réservoirs. Une méthode alternative pour définir une orbite dynamique à l'aide d'un réservoir est de créer des trajectoires dynamiques dans l'espace de dimension $\text{card}(X)$, où X est l'ensemble d'unités du réservoir. Chaque état global $\{u(x)\}_{x \in X}$ du réservoir est défini par un point de cet espace multidimensionnel. Par conséquent, construire une orbite dynamique pour le réservoir considéré revient à imposer des transitions entre deux points distincts de cet espace d'états. Remarquons que cette méthode permet de définir ainsi une représentation spatiale du caractère temporel d'une dynamique.

Les détails présentés ici viennent étayer la vision décrite dans l'introduction de ce document, vision exprimée par van Gelder dans [120], qui postule la nature *intrinsèquement dynamique*

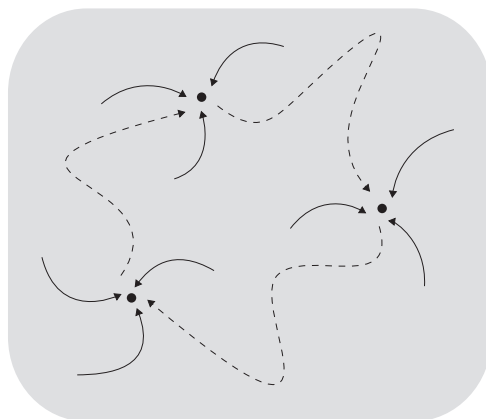


FIG. 2.6 – Illustration schématisée d'un nombre d'attracteurs (symbolisés par des points) d'un système dynamique, dans son espace de phases (symbolisé par le fond gris de la figure). Les lignes pleines représentent des orbites du système dynamique appartenant aux bassins d'attraction de ces attracteurs. Les lignes pointillées représentent des transitions qui permettent au système de « sortir » de ces attracteurs et d'évoluer vers d'autres états stables (voir attracteurs). Les flèches indiquent la convergence du système vers ses attracteurs. Le passage successif par plusieurs attracteurs définit une trajectoire dynamique, qui peut être utilisée pour l'implantation d'une sémantique de traitement d'information. Suivre le texte pour plus d'explications.

des processus cognitifs. Soulignons que cet auteur affirme que la nature du traitement neuronal n'opère pas avec des représentations symboliques, mais utilise un paradigme de *calcul* basé sur les caractéristiques dynamiques des processus neuronaux.

2.4 Modèles d'apprentissage cortical

Au niveau le plus général de définition, l'apprentissage comprend l'ensemble des processus qui permettent de traiter le flot de l'information sensorielle, en vue de mieux le prédire et le contrôler. Autrement dit, l'apprentissage vise à créer et à organiser en permanence des représentations, des modèles, des outils, des comportements, etc., capables de traiter de manière efficace un flot continu d'énergie et d'information, même en absence (totale ou partielle) des éléments de ce flot perceptif.

Au niveau d'un modèle neuronal, nous avons défini l'*apprentissage neuronal* comme étant l'ensemble des processus qui visent l'ajustement continu de ses paramètres libres, et qui conduisent à l'adaptation de la dynamique du modèle considéré, par rapport à son « expérience » perceptive. Dans certains cas, nous pouvons différencier plusieurs échelles de temps auxquelles ces changements ont lieu. Dans ce contexte, nous définissons le *méta apprentissage neuronal* comme étant l'ajustement des variables dites « lentes » du modèle (comme par exemple les taux d'apprentissage, les niveaux des seuils d'activation, etc.).

2.4.1 Paradigmes d'apprentissage dans les modèles neuronaux

Pour un modèle neuronal, le processus d'apprentissage se réfère souvent à l'étape d'entraînement, pendant laquelle le modèle traite un par un, des exemples d'entrées appartenant à un flot d'information (perceptive) et essaie d'ajuster ses paramètres libres, afin de mieux résoudre certaines tâches. Selon la nature de ces tâches, on distingue généralement trois principaux paradigmes d'apprentissage. L'*apprentissage supervisé* demande qu'après chaque nouvelle entrée présentée au modèle, son traitement conduit à un état stable, qui correspond à un certain état attendu. Autrement dit, après l'apprentissage, le modèle doit être capable de mettre en correspondance (de manière univoque) des entrées avec des états de sortie établis a priori. L'*apprentissage non-supervisé* demande une organisation des paramètres du modèle, afin que la dynamique du modèle puisse être corrélée avec certains aspects statistiques du flot d'information perçue. Enfin, l'*apprentissage par renforcement* est un paradigme qui implémente des stratégies de traitement d'information guidées par la maximisation d'une récompense (a priori définie par certains critères).

Notons par exemple que l'apprentissage cortical relatif aux niveaux élémentaires de traitement est essentiellement non-supervisé. Plus spécifiquement, ce processus d'apprentissage ne vise pas à construire une correspondance entre des états d'entrée et des états de sortie, pour la simple raison que les états de sortie ne peuvent pas être connus a priori. Aux niveaux de traitement plus élevés (impliquant par exemple plusieurs régions corticales, ainsi que des zones cérébrales extra-corticales), on peut mettre en évidence des processus d'apprentissage modélisés par le paradigme d'apprentissage par renforcement [45].

Dans notre contexte de recherche, nous sommes intéressé à la modélisation théorique des processus d'auto-organisation, mis en évidence dans les cartes corticales. Dans les paragraphes suivants, nous analysons les modèles neuronaux théoriques sur lesquels repose notre démarche de modélisation.

2.4.2 Apprentissage Hebbien

Il est généralement considéré que la base de l'apprentissage neuronal (et cortical en particulier) réside dans la plasticité synaptique [28], c'est-à-dire dans l'ajustement dynamique de la force des connexions neuronales. Une des découvertes expérimentales fondamentales étayant cette vue appartient à Hebb, qui avait énoncé dans [57] un principe devenu célèbre, connu aujourd'hui sous le nom de la *loi de Hebb*. En résumé, ce principe affirme que, si deux neurones, A et B, sont liés par une connexion synaptique, et si l'activation de A entraîne souvent une activation de B, alors l'efficacité de communication entre les deux neurones est renforcée, facilitant en cours du temps ce type d'activation. Cette affirmation a connu plusieurs interprétations au cours des années, mais elle est souvent énoncée ainsi : le poids de connexion synaptique entre deux neurones dépend proportionnellement de la corrélation ou de la covariance de l'activation des deux neurones. Ici, le poids de connexion synaptique modélise l'efficacité (ou la force) de communication entre les deux neurones. Ainsi, la forme la plus simple de la loi de Hebb peut s'exprimer par :

$$\tau \frac{dw(x, y, t)}{dt} = u(y, t)u(x, t) \quad (2.10)$$

où τ est une constante de temps, généralement appelé *taux d'apprentissage*.

Comme nous pouvons remarquer, telle quelle est définie, cette loi permet l'augmentation illimitée des poids synaptiques. Afin d'empêcher ceci, différentes variantes normalisées de cette

loi ont été proposées. Une de ces variantes, appelée la *loi d'Oja* [89], a l'expression suivante :

$$\tau \frac{dw(x, y, t)}{dt} = u(y, t)u(x, t) - \alpha u^2(y, t) w(x, y, t) \quad (2.11)$$

2.4.3 Cartes d'auto-organisation

L'auto-organisation est considérée comme étant l'émergence des formations (structures) spatiales, temporelles ou fonctionnelles exhibées par un système complexe, dues à l'interaction des parties composantes élémentaires de ce système, et influencées par un flot externe d'énergie ou d'information. Différents niveaux d'analyse (physique, chimique, biologique, sociale, planétaire, cosmique) ont mis en évidence un très grand nombre de processus d'auto-organisation pour les systèmes naturels. Au niveau biologique, de nombreuses études ont montré par exemple la formation des motifs géométriques répétitifs sur les exosquelettes ou sur la fourrure de certains animaux, la formation des fourmilières, des ruches ou des troupeaux, la formation des hiérarchies sociales dans des groupes d'animaux etc. Beaucoup de travaux théoriques ont abordé le problème de modélisation de ces phénomènes [119, 19, 111]. Malgré la difficulté de trouver et formaliser les principes généraux qui régissent les processus d'auto-organisation, les auteurs de ces études considèrent que le caractère non-linéaire des interactions des parties élémentaires du système, l'existence de la compétition au niveau de ces interactions, et les boucles de feedback (positif et négatif) sont des aspects essentiels dans ces processus.

Au niveau cortical, l'existence des cartes corticales, organisées selon les différentes modalités des stimuli perceptifs (revoir la section 2.1), représente un exemple éloquent d'auto-organisation. De nombreuses recherches ont été consacrées à l'étude de ce phénomène, et plusieurs formalismes généraux de calcul ont été proposés, visant à expliquer l'émergence de cette organisation corticale.

Différents mécanismes neuronaux sont supposés être employés par le cortex dans le traitement de l'information sensorielle. Un de ces mécanismes considère que la structure corticale essaye de réaliser une *quantification vectorielle* de l'espace multidimensionnel défini par le flot perceptif. De manière générale, la quantification vectorielle vise à décomposer un ensemble de données en plusieurs classes distinctes, et à associer à chaque classe un représentant (un vecteur) qui exhibe (au moins) une propriété commune ou similaire avec les éléments de la classe qu'il représente. Cette méthode de discrétisation et de synthèse d'information, appliquée à l'espace sensoriel perçu par le cortex, pourrait être à la base d'un encodage neuronal cortical qui permettrait l'optimisation du traitement cortical aux niveaux supérieurs. Notons que dans ce processus de quantification vectorielle, l'espace à quantifier est a priori inconnu : le mécanisme qui doit réaliser sa quantification ne connaît aucune description de cet espace. Dans la majorité des cas, l'espace est perçu exclusivement par une série (un flot) d'éléments extraits de cet espace, appelés *échantillons*. La problématique décrite ici pour la quantification vectorielle, peut être exprimée au niveau formel des cartes d'auto-organisation comme suit.

Considérons un *espace d'entrée* n -dimensionnel, $\Xi \in \mathbb{R}^n$, et $d : \Xi \times \Xi \rightarrow \mathbb{R}^+$ une fonction de distance sur cet espace. Soit $\{\xi_1, \xi_2, \dots\} \subset \Xi$ un ensemble d'échantillons tirés aléatoirement de l'espace Ξ , suivant une distribution inconnue. La quantification vectorielle vise à trouver un ensemble fini Ω de *vecteurs de référence* ou *prototypes* $\omega \in \Omega$, qui soient des éléments appartenant à l'espace Ξ estimant le mieux possible la distribution de ces échantillons. Autrement dit, le choix de l'ensemble des prototypes Ω vise à minimiser l'expression :

$$\sum_{\xi} d(\omega_{\xi}, \xi) \quad (2.12)$$

où la somme est effectuée selon les éléments $\xi \in \{\xi_1, \xi_2, \dots\} \subset \Xi$, et ω_ξ représente le prototype le plus proche de l'échantillon ξ , c'est-à-dire :

$$\omega_\xi = \operatorname{argmin}_{\omega \in \Omega} d(\omega, \xi) \quad (2.13)$$

Parmi les différentes approches qui ont été proposées comme solution au problème de la quantification vectorielle (voir par exemple l'algorithme *k-means* en ligne [76]), nous présentons dans la suite une méthode de modélisation neuronale inspirée par la structure corticale. Cette proposition est due à Kohonen [72], et porte le nom de *cartes d'auto-organisation*, ou en anglais *self-organizing maps* (SOM).

Considérons X un espace topologique d'un ensemble d'unités neuronales, et $r : X \times X \rightarrow \mathbb{R}^+$ la métrique sur cet espace (souvent la métrique euclidienne quand $X \subset \mathbb{R}$ ou $X \subset \mathbb{R}^2$)³. A chaque unité $x \in X$, on associe un prototype $\omega(x)$.

A chaque instant de temps t , un nouvel échantillon $\xi(t)$ est tiré aléatoirement de l'espace d'entrée Ξ . Chaque unité x calcule la ressemblance entre cet échantillon $\xi(t)$, et le prototype qui lui est associé, $\omega(x, t)$. Soit $x_\xi(t)$, l'unité dite « gagnante », qui ressemble le plus à cet échantillon (qui correspond donc au prototype ω_ξ), c'est-à-dire qui minimise la distance $d(\omega(x, t), \xi(t))$. L'algorithme de Kohonen impose alors que les prototypes du réseau ajustent leurs valeurs selon la règle d'apprentissage suivante :

$$\omega(x, t+1) = \omega(x, t) + \alpha v(r(x, x_\xi(t)), t)(\omega(x, t) - \xi(t)) \quad (2.14)$$

où α est une constante dans l'intervalle $(0, 1)$, appelée *taux d'apprentissage*, et v est une fonction de voisinage. Le caractère décroissant de la fonction v permet aux unités les plus proches (dans l'espace X) de l'unité $x_\xi(t)$ de modifier le plus leurs prototypes, afin de les approcher le plus (dans l'espace Ξ) de l'échantillon $\xi(t)$, et implicitement du prototype de cette unité $x_\xi(t)$. Les cas typiques pour la fonction de voisinage considèrent une fonction de type gaussienne (comme dans 2.15), avec une taille variable, qui diminue en cours du temps (comme dans 2.16). Notons toutefois que la taille variable de cette fonction de voisinage n'est pas une condition nécessaire pour atteindre la quantification vectorielle, elle peut être considérée aussi bien fixe. Pour cette raison, nos références suivantes à cet algorithme sous-entendent sa variante d'implantation avec une fonction de voisinage de taille fixe.

$$v(r(x, x_\xi(t))) = \exp\left(-\frac{r(x, x_\xi(t))^2}{2\sigma^2(t)}\right) \quad (2.15)$$

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\tau_\sigma}\right) \quad (2.16)$$

Considérant initialement des valeurs aléatoires pour l'ensemble des prototypes $\{\omega(x, 0)\}_{x \in X}$, et appliquant l'ajustement de leurs valeurs selon la procédure décrite ci-dessus, pour une séquence aléatoire d'échantillons $\{\xi \in \Xi\}$, l'algorithme de Kohonen converge vers la quantification vectorielle de l'espace d'entrée Ξ [24] (voir aussi, pour illustration, les exemples de la figure 2.7). Cependant, cette procédure, en plus de la quantification vectorielle, réussit à mettre en correspondance la similarité des prototypes avec la distance dans l'espace topologique X des unités qui les représentent, réalisant ainsi une projection topographique de l'espace Ξ (ayant la métrique d) sur l'espace X (ayant la métrique r). Comme souvent X est un ensemble de l'espace

³Soulignons ici que les espaces Ξ (l'espace d'entrée) et X (l'espace des positions des prototypes) ont souvent des dimensions différentes, et par conséquent, leurs métriques (respectivement d et r) peuvent être distinctes.

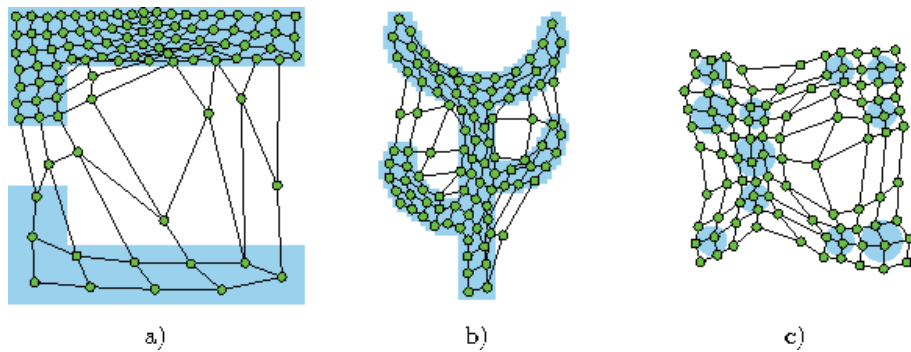


FIG. 2.7 – Cartes bidimensionnelles d’auto-organisation correspondant à différents espaces d’entrée Ξ , qui ont été obtenues par l’implantation de l’algorithme SOM. Les prototypes (symbolisés ici par les petits ronds) ajustent leurs valeurs au cours du temps, afin de converger vers une quantification vectorielle et une projection topographique de l’espace d’entrée considéré. Image extraite du site web <http://www.neuroinformatik.ruhr-uni-bochum.de>.

euclidien mono- ou bidimensionnel, et Ξ est un espace multidimensionnel, l’algorithme de Kohonen réussi à réaliser une procédure de réduction de dimensionalité d’un espace, tout en gardant des aspects topologiques de cet espace. Nous pouvons facilement déduire que ce résultat serait utile pour la compression ou la visualisation abstraite des données appartenant à un espace multidimensionnel, difficiles ou impossible à représenter visuellement dans l’espace 2D ou 3D.

L’algorithme SOM décrit une procédure d’apprentissage non-supervisé, étant donné que la sortie finale (l’ensemble des prototypes) n’est pas comparée avec un résultat étalon donné (ou avec un critère de performance). Notons également que SOM est capable de déduire un motif (une structure) du flot d’information, à cause du fait qu’il traite en cours du temps plusieurs échantillons *similaires* de l’espace d’entrée. Ce caractère aléatoire de découverte de l’espace d’entrée permet de saisir l’aspect statistique de la topologie de cet espace. Autrement dit, ce modèle d’apprentissage extrait la relation « invariante » entre les éléments redondants du flot d’information.

Un autre aspect que nous pouvons remarquer dans la procédure SOM se réfère au fait que cet algorithme implémente une compétition neuronale de type WTM : à chaque nouvelle présentation d’un échantillon $\xi(t)$, les unités neuronales entrent en compétition pour déterminer l’unité x_ξ pour laquelle le prototype qui lui est associé est le plus proche de cet échantillon (proximité du point de vue de la métrique d). De plus, même si cette procédure induit une architecture hiérarchique de traitement, elle est une architecture neuronale de type récurrent et non pas feedforward : l’unité qui gagne la compétition influence l’ajustement des valeurs des prototypes qui ont participé dans le processus de compétition neuronale.

En résumé, les aspects analysés jusqu’ici mettent en évidence trois principes de traitement d’information :

Compétition les unités entrent en compétition pour trouver le prototype le plus proche d’un nouvel échantillon ;

Coopération l’unité gagnante permet, grâce à la fonction de voisinage, de définir un sous-ensemble d’unités qui sont autorisées à effectuer une opération commune (ici, l’adaptation) ;

Adaptation les unités coopérantes ajustent les valeurs de leurs prototypes, en vue de les approcher de l’échantillon présenté, mais le changement s’effectue de manière inversement

proportionnelle avec la distance entre l'unité considérée et l'unité gagnante.

Selon Kohonen, l'algorithme SOM serait un pas important vers la modélisation des caractéristiques fondamentales du traitement d'information au niveau cortical (revoir aussi la section 2.1). Si nous partageons cette opinion, nous considérons toutefois que le modèle proposé par Kohonen montre certains inconvénients fonctionnels. Un premier aspect se réfère au fait que la compétition neuronale évoquée par SOM nécessite le calcul d'un minimum global (la distance $d(w(x, t), \xi(t))$), c'est-à-dire pour tout l'ensemble des prototypes. Si cette procédure est tout à fait immédiate du point de vue mathématique, elle est une procédure centralisée, dont la mise en œuvre par la nature neuronale du traitement cortical est peu probable. Un deuxième aspect, intrinsèquement lié au premier, concerne le fait que le mécanisme pour trouver l'unité gagnante de la compétition suppose une procédure séquentielle (voir synchrone) de recherche, employée pour tous les prototypes du modèle. Toutefois, ce traitement séquentiel de l'information est peu efficace du point de vue temps de réponse et donc, une fois de plus, peu probable au niveau biologique (cortical). Un dernier aspect que nous mentionnons ici se réfère au fait que l'algorithme SOM, tel qu'il est exprimé ici, ne peut pas s'étendre directement pour permettre l'existence de plusieurs unités gagnantes. Cependant, un modèle ayant cette propriété pourrait être utilisé pour modéliser le traitement multimodal d'information, remarqué au niveau cortical. Cet aspect met en évidence le fait que l'algorithme proposé par Kohonen ne peut pas être utilisé dans un contexte de traitement distribué de l'information, contexte essentiel non seulement au niveau neurophysiologique, mais également au niveau de modélisation appliquée (voir robotique), nécessaire pour le développement des systèmes situés complexes. Par conséquent, la motivation principale de notre démarche scientifique explorée dans la suite se réfère à la recherche d'un mécanisme de calcul neuronal qui pourrait à la fois utiliser les concepts introduits par SOM, et à la fois étendre son paradigme de traitement d'information à une large échelle. Pour poursuivre ce défi, dans la section suivante, nous analysons plus en détail, d'un point de vue qualitatif, les caractéristiques essentielles d'un traitement d'information de nature corticale, caractéristiques qui seraient également souhaitables pour un système neuronal autonome et adaptatif.

2.5 Principes de calcul cortical

La vue principalement passive d'observation et d'analyse du monde du réel, imposée par notre vision scientifique traditionnelle, a conduit, au cours des siècles, à une formalisation basée sur la logique classique du traitement de l'information. Cette vision a l'avantage de pouvoir soumettre chaque nouveau modèle de la réalité, décrit dans un langage logique formel, à une validation théorique et expérimentale. Si cette approche permet ainsi la réfutation logique des modèles incohérents, dans le monde du vivant, une stratégie de traitement d'information de type enchaînement logique (observation – conception – validation) peut s'avérer souvent extrêmement coûteuse ou même fatale, vues les contraintes énergétiques et temporelles imposées par l'environnement sur l'existence d'un organisme. L'incertitude et la complexité de la dynamique de l'environnement ont souvent forcé la nature à concevoir des solutions approximatives, qui réalisent un meilleur compromis entre plusieurs contraintes et contingences. La réfutation de ces solutions est souvent traitée à une échelle de temps différente de celle de l'action de l'individu, comme par exemple celle du mécanisme de la sélection naturelle.

Si la vision scientifique classique a réussi à créer, et crée encore un vaste et remarquable univers de concepts, théories et modèles du réel, elle n'est pas arrivée (jusqu'à présent du moins) à pouvoir proposer des solutions comparables aux solutions développées par la nature, en ce qui concerne les capacités d'adaptation et de robustesse. Ceci a amené récemment une partie

des scientifiques à chercher des alternatives aux solutions de traitement d'information basées explicitement sur les formalismes mathématiques de la logique classique, comme par exemple l'approche symbolique ou statistique (*Machine Learning*).

Comme anecdote, il est intéressant de remarquer que même le pionnier des modèles de traitement de l'information avec un mécanisme de manipulation de symboles (qui intègre la méthode de réfutation logique), Alan Turing, avait proposé, pour décider sur le caractère intelligent d'une entité, une méthode de réfutation entièrement empirique [118].

Dans la section 2.1, nous avons présenté différentes propriétés anatomiques et fonctionnelles du cortex cérébral, une des solutions les plus récentes que l'évolution a développé pour mieux satisfaire les contraintes physiques imposées par l'environnement, afin d'optimiser la « régulation du flot de l'énergie et de l'information » (Dan Siegel) perçu par un individu.

Sans doute une des contraintes les plus strictes imposées sur le fonctionnement métabolique d'un organisme vivant concerne sa consommation énergétique. La sélectivité fonctionnelle de la structure corticale, ainsi qu'une ségrégation spatiale et un traitement asynchrone du flot d'information reçu, implémentés par une topologie de connexion neuronale prédominante locale (avec beaucoup de connexions synaptiques à courte distance, et peu de connexions synaptiques à longue distance) pourraient constituer une solution efficace pour minimiser à la fois le « câblage dendritique » et le temps de propagation des signaux électriques, et par conséquent rentabiliser le taux de consommation énergétique du tissu cortical.

Dans [51], Gros propose de prendre en compte certains principes dans la conception bio-inspirée d'un système cognitif. Dans la suite, nous énumérons et nous détaillons chacun de ces principes.

(A) *L'absence des connaissances a priori de l'environnement.* Contrairement à la majorité des applications robotiques, un système cognitif artificiel ne doit pas inclure a priori des modèles de la structure ou de la dynamique de l'environnement. Bien évidemment, une interdiction totale des connaissances a priori est irréalisable, un minimum de modélisation étant toutefois nécessaire pour décrire la nature sensorielle du flot d'information traitée.

(B) *Traitement local de l'information.* Afin d'être capable à s'adapter aux changements structuraux ou dynamiques, ainsi que pour pouvoir être répliqué à des niveaux de complexité supérieurs, un système cognitif artificiel doit reposer sur un traitement local de l'information.

(C) *Architecture modulaire.* Afin de pouvoir résoudre différentes tâches, un système cognitif artificiel doit développer des mécanismes spécialisés de traitement d'information, basés sur des modules de traitement qui interagissent, mais fonctionnent de manière autonome.

(D) *Méta apprentissage via contrôle distribué.* Le méta apprentissage, c'est-à-dire l'ajustement des paramètres « lents » des mécanismes d'apprentissage du système cognitif, qui règlent l'adaptabilité et la sensibilité du système vis-à-vis les signaux qu'il perçoit ou qu'il traite, doit se faire par des mécanismes de contrôle distribués. Les signaux de contrôle doivent être générés par des unités d'analyse et de contrôle spécialisées. Ces unités doivent être capables d'évaluer certains aspects de l'état global du système, et doivent s'activer seulement quand certaines conditions sont remplies.

(E) *Optimisation de la dynamique interne du système.* Les transitions entre les états internes du système doivent être réglées suivant des principes homéostatiques, et la dynamique du système doit être optimisée afin que ces régimes transitoires n'influencent pas, d'une manière essentielle, les états stables du système.

Les aspects énumérés ici tracent les principes généraux à suivre dans la conception des systèmes cognitifs artificiels, et en particulier, des modèles neuronaux bio-inspirés. Une telle approche permet de développer des solutions adaptables, autonomes et robustes, qui soient capables de créer et d'opérer avec des concepts abstraits, et qui favorisent l'émergence des propriétés

dynamiques, allant jusqu'aux comportements.

Ces principes concernant la modélisation neuronale des systèmes cognitifs se retrouvent également dans d'autres travaux. Par exemple, Alexandre rappelle dans [2] que le traitement cortical est essentiellement distribué, se basant sur une compétition locale (mise en œuvre par des connexions latérales au sein d'une région corticale traitant la même modalité du flot perceptif) et sur une coordination cortico-corticale (mise en œuvre par des connexions à longue distance ou par des connexions de feedback entre différentes régions corticales), et étant régulé par des signaux de contrôle distribués. Grâce à ces aspects, un tissu principalement homogène au niveau anatomique est capable d'exhiber une organisation modulaire complexe. Selon Alexandre, les démarches futures de modélisation d'inspiration corticale des processus cognitifs comme l'attention, la communication, la planification, la conscience, etc., devraient prendre en compte ces principes constitutifs du traitement d'information exhibés par la structure corticale.

Sans imposer des directives qui réclament nécessairement la plausibilité biologique des modèles que nous pouvons proposer, ces principes visent à définir un cadre de développement théorique et appliqué qui permet d'exprimer et maintenir le caractère général et évolutif des solutions utilisant ce type de modèles. Pour ces raisons, notre approche de modélisation prend en compte et reste attachée aux principes énoncés ci-dessus.

2.6 Architecture d'apprentissage compétitif d'inspiration corticale

Inspirés d'une part par les principes de traitement d'information au niveau cortical, présentés dans la section 2.5, et d'une autre part par le modèle SOM de cartes d'auto-organisation, décrit dans la section 2.4, nous proposons ici une architecture neuronale générique d'apprentissage, d'inspiration corticale.

Une première abstraction inspirée du traitement cortical est illustrée sur la figure 2.8. Cette architecture décrit le processus d'auto-organisation en considérant deux modules distincts de traitement : un *module de détection*, et un *module de sélection*. Le module de détection est représenté par l'ensemble Ω de prototypes. Les deux modules reposent sur un ensemble d'unités de calcul, placées dans un espace topologique X . L'analogie avec la structure corticale nous amène à considérer X comme étant un espace bidimensionnel euclidien, ou mono-dimensionnel euclidien, pour les cas plus simples qui correspondent à une bande fine de tissu cortical.

Comme dans l'algorithme de SOM, à chaque instant t , un nouvel échantillon $\xi(t)$ est présenté à tous ces prototypes. Chaque unité du module de détection, correspondant à chacun de ces prototypes, calcule une valeur scalaire, notée $i(x, t)$, proportionnelle à la ressemblance entre la valeur du prototype et celle du nouvel échantillon. De manière générale :

$$i(x, t) = f(d(\omega(x, t), \xi(t))) \quad (2.17)$$

où f est typiquement une fonction décroissante, et a des valeurs dans l'intervalle $[0, 1]$ (pour des raisons de normalisation). Cette fonction, souvent appelée *fonction de reconnaissance* (en anglais *tuning curve*), quantifie la ressemblance entre un prototype et un échantillon : plus ceux-ci se ressemblent, plus la valeur de f s'approche de 1.

La distribution $\{i(x, t)\}_{x \in X}$, appelée aussi *distribution ou configuration d'entrée*, constitue l'entrée du module de sélection. Comme réponse à cette stimulation, le module de sélection calcule de manière distribuée une réponse $\{u(x, t)\}_{x \in X}$, appelée aussi *distribution ou configuration d'activités neuronales*. Ce calcul vise à favoriser (en leur attribuant une activité u élevée) un ou

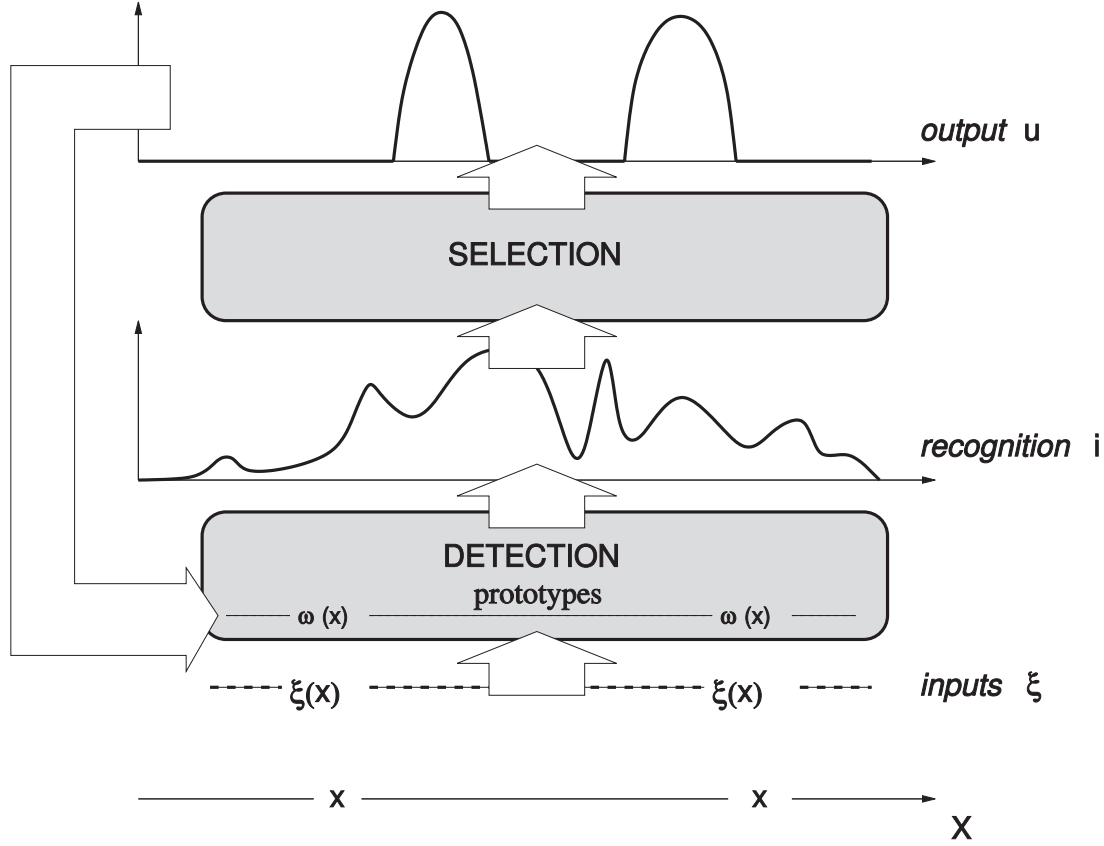


FIG. 2.8 – Illustration graphique d’une architecture neuronale d’inspiration corticale, décrivant le processus générique d’auto-organisation des cartes corticales. L’architecture se compose de deux modules de traitement : un module de détection, et un module de sélection. Chaque nouvel échantillon $\xi \in \Xi$ présenté au modèle, stimule chaque unité $x \in X$ du module de détection. Comme résultat, ce module calcule une distribution $\{i(x)\}_{x \in X}$, qui quantifie la ressemblance de chacun de ses prototypes à l’échantillon ξ . Cette distribution stimule ensuite le module de sélection, qui a comme but de favoriser l’activation des unités x correspondant à un niveau de ressemblance (mesuré par i) le plus élevé localement. Le résultat de ce calcul, représenté par la distribution $\{u(x)\}_{x \in X}$, module, selon une règle d’apprentissage, l’ajustement des prototypes $\omega(x)$.

plusieurs groupes compacts d'unités, pour lesquelles l'effet cumulatif de la mesure de ressemblance quantifiée par i , est le plus élevé. Le module de sélection a ainsi le but d'extraire (de manière distribuée) les groupes des prototypes les plus proches de l'échantillon ξ (du point de vue de la métrique de l'espace Ξ). Notons ici l'analogie avec la fonction de voisinage implantée par SOM, mais aussi la différence essentielle par rapport à ce concept : le traitement réalisé par le module de sélection doit être distribué, et non pas centralisé, comme dans le cas de l'algorithme de Kohonen.

L'activité neuronale $u(x, t)$, résultante du calcul du module de sélection, participe ensuite à l'ajustement des valeurs des prototypes, d'une manière similaire à l'apprentissage considéré par l'algorithme SOM (revoir l'équation 2.14). Ici, la règle d'apprentissage Hebbien est définie par :

$$\omega(x, t + 1) = \omega(x, t) + \alpha u(x, t)(\omega(x, t) - \xi(x, t)) \quad (2.18)$$

La modélisation du processus d'auto-organisation que nous proposons ici suit les principes de traitement cortical d'information énumérés dans la section précédente. Toutefois, pour réaliser le passage de ce schéma descriptif à une implémentation réelle, nous présentons dans le paragraphe suivant une architecture neuronale d'auto-organisation basée sur un modèle de réseau récurrent de type excitation/inhibition.

2.6.1 Architecture neuronale d'auto-organisation basée sur la dynamique des champs neuronaux

L'architecture présentée dans la figure 2.8, et décrite ci-dessus, repose sur un module de sélection, au lieu de la procédure implémentée par l'algorithme SOM pour trouver le prototype le plus ressemblant à un nouvel échantillon.

Examinant les différents paradigmes de calcul neuronal dynamique analysés dans les sections 2.2 et 2.3, nous avons choisi une approche basée sur le modèle des *champs neuronaux dynamiques* pour implémenter le module de sélection de cette architecture d'auto-organisation.

Le modèle des champs neuronaux dynamiques est un modèle de réseau récurrent de type excitation/inhibition, qui repose sur la conjecture (étayée expérimentalement) qu'un nombre important de phénomènes et de mécanismes fonctionnels corticaux émergent d'un processus de compétition neuronale locale et distribuée au niveau cortical. Les champs neuronaux dynamiques modélisent la structure corticale en considérant un ensemble bidimensionnel (ou monodimensionnel dans les cas simplifiés) d'unités de calcul neuronal, connectées entre elles d'une manière locale, et exerçant l'une vers l'autre soit une influence excitatrice, soit une influence inhibitrice. Dans la suite de notre recherche, nous allons examiner en détail la manière dont ce modèle peut en effet rendre possible une implémentation du module de sélection de l'architecture d'auto-organisation, en accord avec les principes de calcul cortical que nous avons affirmé vouloir suivre dans nos démarches de modélisation neuronale.

La variante discrétisée du schéma générale présentée dans la figure 2.8, illustrant cette fois-ci l'implantation de l'architecture d'auto-organisation avec le modèle des champs neuronaux dynamiques en tant que module de sélection, est présentée dans la figure 2.9.

Un premier aspect qui découle de cette implantation concerne le traitement local et distribué. Comme nous pouvons l'observer à partir de ce graphique, les activités des unités élémentaires de cette architecture (c'est-à-dire $\{i(x, t)\}_{x \in X}$ pour le module de détection, et $\{u(x, t)\}_{x \in X}$ pour le module de sélection) sont entièrement déterminées par des calculs et des interactions locales. Notons que dans ce contexte, une unité neuronale de l'architecture décrite ici ne correspond pas à une cellule neuronale individuelle, mais plutôt à une abstraction fonctionnelle d'un groupe de

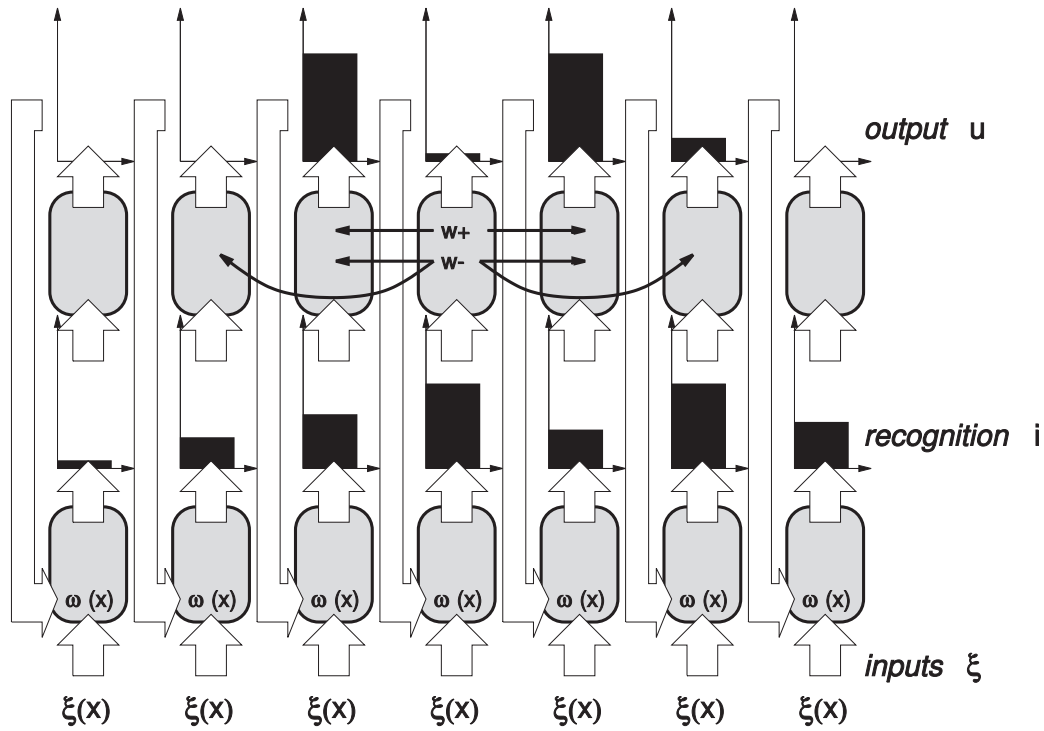


FIG. 2.9 – Schéma discrétisé de l'architecture d'auto-organisation présentée dans la figure 2.8. Le graphique présente en particulier l'implantation du module de sélection, réalisée avec le modèle des champs neuronaux dynamiques. Les unités de ce module sont localement couplées entre elles, via les poids synaptiques $w-$ et $w+$, et implémentent un modèle de réseau récurrent de type excitation/inhibition.

neurones (cellule de calcul), comme par exemple l'unité fonctionnelle représentée par la minicolonne corticale. De ce point de vue, le modèle de traitement que nous proposons repose sur la vision architecturale du cortex décrite par Burnod dans [16].

Dans un deuxième temps, notons que les deux modules de cette architecture doivent être considérés comme deux processus dynamiques distincts, caractérisés par des paramètres variables et aux constantes de temps très différentes. L'influence de ces paramètres dans le couplage des deux modules de l'architecture neuronale décrite ici, mais aussi l'éventuel couplage avec d'autres modules de traitement neuronal, doit être prise en compte dans la complexité de la dynamique de ce système. Dans le contexte que nous décrivons ici, la réponse du module de sélection dépend à chaque instant de la distribution d'entrée $\{i(x, t)\}_{x \in X}$, calculée par le module de détection. Comme nous allons le présenter et l'analyser dans les chapitres suivants, un champ neuronal dynamique répond à une telle stimulation avec une distribution d'activités neuronales $\{u(x, t)\}_{x \in X}$ caractérisée par une ou plusieurs *bulles d'activité neuronale*, qui se constituent à l'endroit des zones les plus stimulées localement par l'entrée i . Dans certains régimes et conditions de paramétrage, les bulles convergent vers un état stable (devenant ainsi des attracteurs pour le système dynamique du champ). Comme d'une part les bulles d'activité neuronale du champ, qui sont le pendant de la fonction de voisinage dans SOM, deviennent fortes et stables seulement après un certain temps de relaxation (une période transitoire), et comme d'une autre part la distribution $\{u(x, t)\}_{x \in X}$ module l'ajustement des valeurs des prototypes (revoir la règle d'apprentissage de l'équation 2.18), il est nécessaire que le module de détection ait une échelle de temps d'évolution dynamique inférieure à celle du module de sélection. Afin d'implémenter cette contrainte dynamique, dans nos simulations, nous considérons chaque nouvel échantillon $\xi(t)$ constant pour 100 pas de simulation (ce qui implique $i(x, t)$ constant pour la même période de temps). De plus, pour permettre un apprentissage (l'ajustement des valeurs des prototypes) guidé principalement par les bulles stables du champ neuronal, nous imposons un taux d'apprentissage assez faible, $\alpha = 2 \cdot 10^{-4}$. Ainsi, les prototypes s'ajustent essentiellement pour les unités ayant gagné la compétition neuronale (c'est-à-dire les unités correspondant aux bulles d'activité neuronale).

Un autre aspect relatif à l'architecture illustrée sur la figure 2.9 est le schéma discret de calcul. Notons que, si d'autres modèles préfèrent un formalisme mathématique continu pour exprimer leurs propriétés fonctionnelles, ces modèles sont en général discrétisés lors de leur implantation sur des machines (architectures) de calcul conventionnelles. Comme le but de nos recherches est le développement de modèles neuronaux distribués et évolutifs, capables d'être utilisés dans des architectures à large échelle de calcul parallèle intensif comme [42], nous avons choisi de décrire notre modèle d'auto-organisation directement par un schéma discret. En outre, cette expression discrète est en accord avec la structure discrète du tissu cortical.

Enfin, nous notons que, de manière analogue au réseau de Hopfield (voir la section 2.2.4), l'architecture d'auto-organisation que nous proposons permet l'évaluation asynchrone des activités de ses unités élémentaires. Cette propriété vient confirmer l'attachement aux principes de calcul cortical exprimés précédemment.

L'implantation logicielle de cette architecture d'auto-organisation basée sur la dynamique des champs neuronaux est réalisée en utilisant la bibliothèque logicielle `bi_jama` [41], une plate-forme de programmation parallèle multi-couches d'automates cellulaires.

2.6.2 Comparaison avec d'autres modèles de compétition neuronale

Analysant l'architecture neuronale d'auto-organisation décrite dans les paragraphes précédents, nous pouvons observer qu'elle réussit à décrire plusieurs paradigmes de compétition, et par extension, différentes approches de quantification vectorielle.

Si par exemple le résultat du module de sélection est toujours une unique impulsion de Dirac, qui correspond à l'unité x ayant la valeur $i(x, t)$ maximale parmi toute la distribution d'entrée $\{i(x, t)\}_{x \in X}$, alors l'architecture implémente une version en ligne de l'algorithme k-means [76].

Dans un deuxième cas, si nous considérons le résultat du module de sélection comme étant défini par une fonction de voisinage, généralement une gaussienne, centrée à l'endroit qui maximise la valeur $i(x, t)$ de la distribution i , nous obtiendrons l'algorithme SOM proposé par Kohonen [72]. Notons que les deux algorithmes, k-means et SOM, reposent sur le calcul d'un extremum global (pour trouver l'unité gagnante de la compétition). Dans ses implantations classiques, cette procédure de calcul est centralisée, ce qui contrevient aux principes de calcul cortical qui constituent nos fondations de modélisation neuronale.

Une variante plus proche de notre démarche, qui s'intègre aussi dans le schéma général de l'architecture d'auto-organisation que nous proposons ici, considère comme module de sélection un réseau neuronal avec des poids synaptiques latéraux (les $w+$ et $w-$ sur la figure 2.9) ajustables au cours du temps. Toutefois, comme nous allons l'analyser dans la suite du document, ce modèle neuronal, appelé RF-LISSOM [83, 84] nécessite une remise à zéro de toutes ses activités neuronales, avant la présentation d'un nouvel échantillon. Nous considérons que cette procédure, utilisant une opération centralisée sur toutes les activités du réseau neuronal du module de sélection, contrevient également aux principes de calcul cortical évoqués dans ce chapitre.

2.7 Résumé

Motivés par un contexte de recherche qui vise le développement informatique des modèles de calcul neuronal inspirés du fonctionnement cérébral, nous avons présenté dans le début de ce chapitre plusieurs caractéristiques anatomiques et fonctionnelles du cortex cérébral qui pourraient constituer une forte source d'inspiration pour la conception future des systèmes cognitifs de traitement d'information.

Nous avons ensuite passé en revue plusieurs paradigmes de modélisation neuronale. Dans la section 2.3.2, nous avons présenté le traitement neuronal sous une perspective dynamique. Dans ce contexte, nous avons souligné l'intérêt d'étudier différents aspects dynamiques concernant l'évolution d'un réseau neuronal, et notamment la stabilité de ses états. Toutefois, d'une part la complexité de ces systèmes dynamiques, et d'une autre part leur non-linéarité intrinsèque, rendent l'étude même des modèles neuronaux proposés assez difficile avec une approche mathématique analytique.

Dans la même section 2.3.2 nous avons décrit plusieurs méthodes générales pour développer des mécanismes de traitement d'information basés sur la dynamique d'un système neuronal. Bien que ces variantes n'incluent pas, par exemple, des moyens de calcul neuronal à base d'oscillations permanentes (voir les ondes de propagation d'activité neuronale, mis en évidence par des études expérimentales comme [109, 50, 129]), nous considérons l'étude approfondie de ces approches essentielle, dans le but de rapprocher les paradigmes courants de modélisation neuronale de la manière dont le traitement neuronal a lieu au niveau cérébral. Dans cette vision, le calcul neuronal peut être vu comme un effet émergent de la dynamique complexe issue de l'interaction des unités fonctionnelles élémentaires. Guidées par cette vision, nos recherches visent à continuer l'exploration des approches de modélisation neuronale dynamique, qui permettent d'étendre les capacités des modèles de calcul neuronal d'inspiration corticale.

Afin de fonder une base solide de développement informatique, qui permettrait l'extension et l'adaptabilité à large échelle des modèles neuronaux, dans la section 2.5, nous avons énuméré et analysé un certain nombre de principes de modélisation du calcul cortical, principes qui ont

été également évoqués par plusieurs auteurs du domaine.

En accord avec ces principes, dans la section 2.6, nous avons proposé une architecture générique d'auto-organisation, et nous avons décrit une implantation possible de celle-ci avec un modèle de champs neuronaux dynamiques.

Dans le chapitre suivant nous allons présenter en détail le paradigme de calcul des champs neuronaux dynamiques, analysant leurs caractéristiques dynamiques et fonctionnelles qui visent à nous aider à implémenter l'architecture d'auto-organisation que nous avons proposée ici.

Nota : les travaux sur la conception de l'architecture d'auto-organisation d'inspiration corticale introduite dans la section 2.6 de ce chapitre ont été décrits en détail dans le cadre de l'article 6, mentionné dans l'annexe A.

Chapitre 3

Champs neuronaux dynamiques

Dans les chapitres précédents, nous avons conclu que l'essence du comportement intelligent ne demeure ni dans la structure de l'information interprétée, ni dans l'anatomie des outils de traitement d'information appartenant à l'interprète, ni même dans les représentations que celui-ci conçoit parfois pour faciliter ce traitement. Cette essence réside dans la *dynamique* de l'interaction continue entre l'interprète et le flot d'information perçue.

Du fait des propriétés remarquables exhibées par la structure corticale et présentées dans le chapitre 2, de nombreuses études se sont consacrées à analyser la dynamique de traitement d'information de nature corticale. L'homogénéité anatomique mésoscopique du cortex a permis l'utilisation d'un cadre théorique et expérimental générique pour analyser les activités corticales sous différentes conditions de stimulation.

Néanmoins, du fait qu'un vaste nombre de neurones composent le cortex, les outils expérimentaux se sont vus vite limités dans leur ambition de mettre en évidence les aspects dynamiques de l'activité corticale d'un grand ensemble de neurones. Pour aborder ce nouveau problème, au lieu d'étudier en détail l'activité des neurones individuels, les chercheurs ont proposé de modéliser l'*interaction* entre les activités neuronales des populations de neurones. Ces modèles ont été appelés génériquement des modèles de *champs neuronaux*. Ayant comme but la modélisation de type large échelle des processus de compétition et coopération des activités neuronales dans le tissu cortical, plusieurs modèles théoriques ont été proposés et raffinés au cours du temps. Des outils d'analyse provenant de la théorie des systèmes dynamiques ont permis la mise en évidence de multiples phénomènes dynamiques de l'activité neuronale. Grâce aux avancées du domaine, certains processus caractérisant la macro-dynamique corticale ont pu être expliqués, mais aussi d'autres applications robotiques ont pu intégrer avec succès des solutions basées sur la dynamique des champs neuronaux.

Le but de ce chapitre est de tracer l'évolution du domaine des champs neuronaux dynamiques et présenter les modèles de référence du domaine, ainsi que les principaux résultats théoriques existant dans la littérature.

Nous commençons par présenter les modèles fondateurs de champs neuronaux dynamiques. Nous examinons ensuite les motifs dynamiques issus du comportement d'un champ et les aspects concernant l'existence ou la stabilité de ces phénomènes. Le chapitre se poursuit avec les extensions notables de ces modèles afin d'enrichir leur dynamique. Au final, nous analysons les méthodes de simulation numérique des champs neuronaux dynamiques.

Dans ce chapitre, nous analysons tous ces résultats de manière critique, afin d'offrir une perspective de l'état de l'art et surtout une vision de l'avenir du domaine des champs neuronaux dynamiques.

3.1 Concepts généraux

3.1.1 Nécessité d'une modélisation mésoscopique de l'activité corticale

Étant inspirés par les capacités remarquables du cerveau, les chercheurs se sont employés à analyser au cours du temps divers aspects du fonctionnement cérébral, afin de comprendre et révéler la nature intrinsèque du comportement intelligent. Différentes méthodes et niveaux d'analyse ont vu le jour pour étudier la dynamique du traitement d'information de nature cérébrale.

Motivés par ces résultats, les scientifiques ont longtemps cherché à décortiquer et mettre en évidence l'ensemble des mécanismes cérébraux qui seraient responsables de la naissance et du développement des comportements adaptatifs des organismes vivants [21, 95].

Plusieurs arguments fondés sur les propriétés remarquables du cortex, présentées dans le chapitre 2, amènent à considérer que la structure corticale a un rôle décisif dans la formation et le fonctionnement des processus cognitifs. Un des défis majeurs des chercheurs est de découvrir les mécanismes intrinsèques de la plasticité corticale, c'est-à-dire d'expliquer comment à partir d'une homogénéité anatomique le cortex réussit à se spécialiser et se réorganiser pour traiter un flot perceptif multimodal, ainsi que pour résoudre des tâches cognitives très variées. Les résultats expérimentaux obtenus grâce aux travaux de pionniers des neurosciences, comme Santiago Ramón y Cajal [17, 18], Mountcastle [86] ou Hubel et Wiesel [66], ont ouvert le chemin pour une abondance de recherches concernant diverses caractéristiques de la structure corticale. Toutefois, la méthode expérimentale rencontre certaines limites, quand il s'agit de mettre en évidence des aspects clé de la physiologie ou du fonctionnement cortical (par exemple l'intégralité de la topologie de connexion ou l'interaction entre les activités de larges masses de neurones). Par conséquent, les chercheurs ont essayé de concevoir des modèles théoriques du cortex, afin de construire des hypothèses et d'inférer de nouvelles connaissances à propos des phénomènes corticaux. Les champs neuronaux dynamiques s'inscrivent dans cette démarche, constituant une tentative de décrire de manière formelle la compétition entre les activités neuronales corticales.

3.1.2 Approches de modélisation mésoscopique de l'activité corticale

Les travaux mentionnés ci-dessus ont pu mettre en évidence une organisation horizontale en six couches, et une organisation verticale en colonnes du tissu cortical. D'après les éléments présentés dans le chapitre 2, les colonnes corticales peuvent être considérées comme étant les modules fonctionnels élémentaires du traitement neuronal cortical. Inspirés par cette perspective, les modèles de champs neuronaux dynamiques considèrent le tissu cortical comme un maillage ou un *champ* d'unités fonctionnelles. Cependant, les champs neuronaux ne se limitent pas à modéliser l'activité des colonnes corticales, leur but étant de modéliser l'activité globale de groupes (ou populations) de neurones, rassemblés ici sous le nom d'*unités fonctionnelles*. Cette propriété de généralité fonctionnelle représente une des caractéristiques fondamentales de ce paradigme neuronal.

Une deuxième abstraction de ce modèle porte sur la modélisation de la densité spatiale de ces unités. Le rapport très élevé entre la superficie du tissu cortical (environ 2600 cm^2 chez l'homme) et le diamètre d'une colonne corticale (environ $300 - 600 \mu\text{m}$) permet de considérer qu'au niveau mésoscopique, ce maillage d'unités fonctionnelles est un continuum. Notons X l'ensemble support de ce maillage (les positions de ces unités fonctionnelles dans la couche corticale). Ainsi, nous considérons dans la suite $X = \mathbb{R}$ pour le cas unidimensionnel, et $X = \mathbb{R}^2$ pour le cas bidimensionnel.

Le troisième aspect structurel pris en compte par le modèle des champs neuronaux considère une topologie de connexion fixe entre les unités du champ. Celle-ci, généralement définie par un *noyau de poids des connexions latérales* noté w , est une fonction relative à la distance entre les unités du champ dans l'espace X . Dans la littérature, les fonctions w les plus souvent utilisées sont : la fonction unitaire (voir l'équation 3.1), la fonction normale ou gaussienne (voir l'équation 3.2), le chapeau mexicain, appelée aussi différence de gaussiennes (voir l'équation 3.3), ou le chapeau de magicien (voir l'équation 3.4).

$$w(|x - x'|) = \begin{cases} 1 & \text{si } |x - x'| \leq R_0 \\ 0 & \text{sinon} \end{cases} \quad (3.1)$$

$$w(|x - x'|) = e^{-\alpha|x-x'|^2} \quad (3.2)$$

$$w(|x - x'|) = A^+ e^{-a|x-x'|^2} - A^- e^{-b|x-x'|^2} \quad (3.3)$$

$$w(|x - x'|) = -w_0(1 - |x|)e^{-|x-x'|} \quad (3.4)$$

Afin de pouvoir modéliser la dynamique corticale, il est nécessaire de prendre en compte la façon dont son activité est influencée par l'activité neuronale d'autres sources extra-corticales. Ainsi, le modèle général du champ considère que chaque unité fonctionnelle $x \in X$ du champ reçoit un stimulus $i(x, t)$, qui définit l'*entrée corticale* à chaque instant du temps t . La réponse $u(x, t)$ de l'unité x définit son *activité neuronale*. Cette réponse dépend à la fois de la stimulation externe $i(x, t)$, mais aussi des activités neuronales des unités voisines, avec lesquelles l'unité x est connectée (conformément au noyau des poids des connexions latérales w). Savoir comment intégrer l'influence de ces facteurs dans l'activité globale $u(x, t)$ de chaque unité x du champ, à chaque instant de temps t , constitue le défi des modèles de champs neuronaux et donne la particularité de la dynamique exhibée par chacun d'eux.

Une fois définies les « composantes élémentaires » du modèle de champ, il nous reste à mettre en correspondance ces termes avec une interprétation physiologique de l'activité corticale, ou autrement dit (d'un point de vue modélisation théorique et simulation informatique) il est nécessaire de définir ce que chaque unité fonctionnelle $x \in X$ du champ « calcule ».

Si nous interprétons i comme étant la somme des courants ioniques pré-synaptiques, u peut être vu comme la fluctuation des potentiels électriques déclenchés par la stimulation du champ avec i . Dans ce cas, nous pouvons affirmer que le champ modélise la génération et la dynamique de propagation des *potentiels d'action* (en anglais *spikes*) sur la surface corticale. Cette interprétation peut être modélisée en utilisant un codage neuronal de type *d'intégration et décharge* (en anglais *integrate and fire*), pour chaque unité x du champ. Dans ce modèle d'activité neuronale, les impulsions neuronales sont générées seulement si la contribution des courants pré-synaptiques reçus par l'unité devient supérieure à un certain seuil prédéfini. Dans ce cas, u a la forme de trains d'impulsions électriques (en anglais *spike train*), déchargées au cours du temps dans la couche corticale.

Toutefois, bien qu'elles soient conçues pour modéliser fidèlement les processus physiologiques neuronaux, ces approches ont en pratique un coût de modélisation très élevé. Afin de contourner ces faiblesses, une nouvelle abstraction théorique, inspirée de la mécanique statistique, a été proposée. Cette approche, appelée la *théorie des champs moyens* (en anglais *mean field theory*), propose de modéliser seulement le nombre moyen d'impulsions générés dans l'unité de temps, c'est-à-dire le *taux de décharge* des impulsions (en anglais *firing rate*). Pour cette raison, ces modèles sont connus sous le nom de *modèles fréquentiels* (en anglais *firing rate models*). Montrant une grande robustesse au bruit spatial ou temporel, ces modèles sont très souvent utilisés en pratique, grâce à leur relative simplicité de simulation et à la possibilité de les étudier avec des

outils théoriques. Toutefois, nous devons retenir que ces modèles ne sont pas capables de mettre en évidence certains aspects temporels résultant de la dynamique sous-jacente des impulsions neuronales (par exemple, ils ignorent l'importance des moments du temps de la génération des potentiels d'action, ou encore, ils ignorent leur corrélation temporelle). En conclusion, cette approche revient à considérer $u(x, t)$ comme étant le taux instantané de l'activation neuronale de l'unité x du champ. Elle soutient la vision assumée par les auteurs du domaine de modéliser l'activité de groupes ou populations de neurones, et pas nécessairement l'activité individuelle des cellules corticales.

Une approche de modélisation étendue, plus fidèle que cette dernière, prend en compte le fait qu'à chaque instant du temps, une certaine fraction des unités actives du champ se trouvent dans la période réfractaire (elles ne peuvent pas contribuer aux activités des unités voisines avec lesquelles celles-ci sont connectées). Cet aspect est souvent modélisé par une *fonction de seuil* (appelée également *fonction de transfert*), notée σ . Dans la pratique, ces fonctions sont exprimées par une fonction de Heaviside (cf. 3.5) ou par des fonctions sigmoïdes (cf. 3.6, 3.7).

$$\sigma_H(u) = \begin{cases} 0 & \text{si } u \leq 0 \\ 1 & \text{sinon} \end{cases} \quad (3.5)$$

$$\sigma(u) = 1/(1 + e^{-u}) \quad (3.6)$$

$$\sigma(u) = \tanh(u) \quad (3.7)$$

3.1.3 Description générale de la dynamique d'un champ neuronal

Les éléments présentés dans les paragraphes précédents permettent de décrire de manière formelle un modèle de champ neuronal. L'expression générale de ce type de formalisme peut être écrite sous la forme suivante :

$$\tau \frac{du(x, t)}{dt} = aA + bB + cC + D \quad (3.8)$$

$$A = -u(x, t) \quad (3.9)$$

$$B = \int_{x'} w(|x - x'|) \sigma(u(x', t)) dx' \quad (3.10)$$

$$C = i(x, t) \quad (3.11)$$

où la réponse u du champ évolue à chaque pas de temps t en fonction de ces quatre termes : A , l'activité courante de l'unité ; B , la contribution des activités neuronales des unités voisines avec lesquelles l'unité x est connectée ; C , la stimulation externe ; D , une quantité qui tient place pour d'autres termes qui pourraient influencer l'activité d'une unité fonctionnelle (par exemple un potentiel de repos, un niveau de bruit etc.). a, b et c sont des scalaires de pondération des quantités exprimées par les termes A, B et C , et τ est une constante de temps. La figure 3.1 montre graphiquement certains de ces éléments.

Une fois intégrés sous la forme des équations de champ tous les principes de modélisation corticale énumérés précédemment, nous pouvons nous demander quelles sont les valeurs numériques des paramètres de ces équations. En fait, cette problématique ouvre toute une série de questions sous-jacentes. Par exemple, une première question demanderait s'il existe des valeurs physiologiques à attribuer à ces paramètres. Certaines études, comme par exemple [77], ont pu mettre en correspondance différents paramètres d'un champ avec des valeurs physiologiques mesurées expérimentalement. Toutefois, ces valeurs caractérisent souvent une classe assez limitée de régimes dynamiques de l'activité corticale. Une deuxième question pourrait donc demander quels

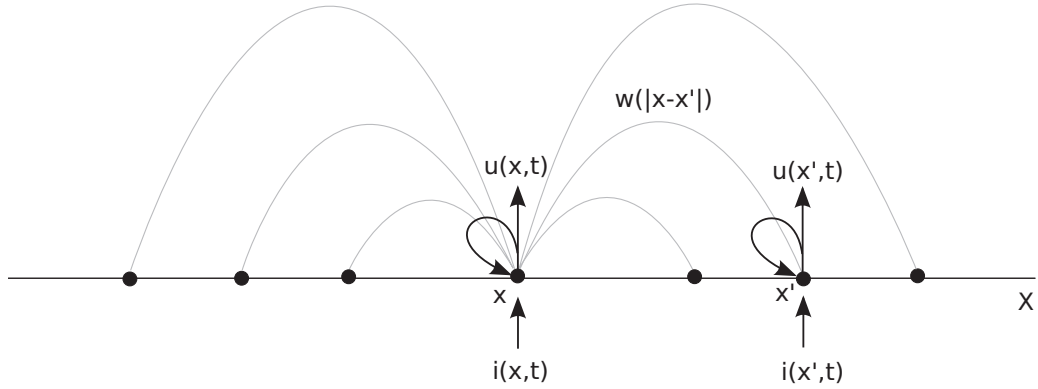


FIG. 3.1 – Illustration graphique du schéma général, décrit par les équations 3.8-3.11, utilisé pour la modélisation de l'activité dynamique d'un champ neuronal unidimensionnel.

sont les critères pour choisir ces valeurs. Comme nous allons constater dans la suite, de l'analyse de plusieurs modèles de champ neuronal, différentes valeurs de leurs paramètres peuvent engendrer un nombre assez varié de comportements dynamiques. Toutefois, pour trouver les intervalles de valeurs qui garantissent ces comportements dynamiques, nous avons souvent besoin de faire appel à une analyse mathématique approfondie de ces équations. Ces études analytiques visent le plus souvent la recherche des conditions d'existence et de stabilité des solutions $u(x, t)$ des équations du champ. D'autres aspects analysés concernent l'étude dynamique des périodes d'oscillation, des états d'équilibre, des temps de relaxation (le temps nécessaire pour u pour revenir dans un état d'équilibre, quand le système dynamique subit une certaine perturbation), ou bien d'autres temps de transitions considérés significatifs.

Nous constatons ainsi que le problème de choix de paramètres pour un champ neuronal est un problème de recherche en soi. Dans le même contexte d'étude théorique, ce problème entraîne l'étude de la génération, la propagation et la stabilité de divers motifs dynamiques, dans différents cas de l'évolution dynamique du champ. Une telle analyse pourrait faciliter l'évaluation de l'influence des échelles temporelles ou spatiales de la modélisation, ou la mise en évidence de certains méta-paramètres du modèle de champ.

3.1.4 Critiques du paradigme de modélisation des champs neuronaux dynamiques

L'équation générique de champ de type 3.8 met en évidence les éléments considérés essentiels qui gouvernent la dynamique de l'activité corticale, responsables de la compétition entre les populations de neurones. Néanmoins, ces aspects de modélisation pourraient être parfois considérés insuffisants pour capturer certaines dynamiques de l'activité corticale.

Une première remarque concerne le fait que le modèle des champs neuronaux dynamiques ne fournit pas une explication pour l'ontogenèse de la topologie neuronale corticale. Le modèle se base sur une topologie de connexion neuronale de type 1D (\mathbb{R}) ou 2D (\mathbb{R}^2) a priori définie, les poids des connexions latérales entre les neurones restant constants en temps. D'autres modèles, notamment RF-LISSOM [83, 84], prennent en compte la plasticité de la connectivité corticale et intègrent dans le modèle la dynamique d'adaptation des poids latéraux. Cependant, nous rappelons que le but de ces modèles est loin d'offrir une explication pour tous les processus corticaux, leur point fort étant plutôt de simplifier (abstraire) ces processus afin de faciliter l'étude à la

fois théorique et de simulation de ces phénomènes. D'ailleurs, les champs neuronaux dynamiques seraient destinés à mettre en évidence la compétition entre les activités de populations de neurones, une fois les poids synaptiques des connexions latérales stabilisés, selon d'autres processus neuronaux (d'apprentissage), intrinsèques à la structure corticale. De ce point de vue, nous pouvons affirmer que les champs neuronaux sont caractérisés par une échelle de temps supérieure à la génération des trains d'impulsions neuronaux, mais inférieure à la durée de la réorganisation des poids synaptiques corticaux.

Une autre critique des modèles actuels de champs neuronaux dynamiques concerne la modélisation des unités neuronales avec des méthodes de type *champ moyen* et notamment des *modèles fréquents*, considérés trop « simplistes » pour mettre en évidence des processus inhérents à la dynamique corticale (comme par exemple mentionnés précédemment, les phénomènes de synchronisation des activités corticales). L'argument opposé à cette critique affirme que les simplifications de ces modèles sont bien une *vision statistique* de la probabilité de décharge neuronale, une abstraction qui n'est pas sensible aux phénomènes perturbateurs de bruits, d'instabilités temporelles ou de changements locaux spontanés de topologie de connexion. En résumé, ce type de modélisation serait bien plus adéquat pour caractériser les phénomènes macroscopiques globaux de l'activité corticale. Au détriment d'une grande fidélité physiologique, ces approches aident à gagner une clarté et une efficacité dans la simulation de ces modèles. Cet aspect est étayé par un intérêt constant de la communauté scientifique vis-à-vis de ces modèles fréquents [92, 22]. En conclusion, la théorie de type *champ moyen* reste appropriée pour appréhender qualitativement les processus concernant la macro-dynamique de la structure corticale [125, 29].

Enfin, nous mentionnons la critique de l'homogénéité des unités neuronales d'un champ (toutes ayant les mêmes paramètres structurels), étant donné le fait que même si le cortex est un tissu quasi homogène, nous pouvons bien distinguer différents types de cellules, et différents types de connexions (la symétrie des poids de connexions n'étant pas biologiquement plausible). La réponse vis-à-vis de cette remarque fait appel, comme avant, au fait que ces approches moyennent les valeurs des paramètres caractérisant la couche corticale, également homogénéisent ces valeurs. Cependant, même si ce n'est pas fréquent, nous trouvons dans la littérature des tentatives d'inclure dans la modélisation des champs des poids de connexion asymétriques [61], ou bien de prendre en compte l'interaction de plusieurs (couches de) champs ayant différentes caractéristiques, qui introduisent ainsi une certaine hétérogénéité dans la modélisation [70].

Résumant ces remarques présentées ici, nous pouvons affirmer que les hypothèses de travail considérées par les modèles de champs neuronaux dynamiques sont issues d'un besoin de faire un compromis entre la difficulté d'une modélisation fidèle du point de vue biologique et, la nécessité pratique d'étudier et simuler ces phénomènes dynamiques de nature corticale. La possibilité d'employer des études théoriques complexes, ainsi que de simuler et d'utiliser leur dynamique dans la conception de solutions bio-inspirées, constituent des avantages qui font de ce paradigme un modèle neuronal digne d'être considéré dans les démarches de modélisation de la structure corticale.

Les champs neuronaux dynamiques sont un domaine très actif, proposant souvent de nouvelles extensions, afin d'enrichir la panoplie des phénomènes dynamiques d'inspiration corticale. Même dans ces conditions d'hypothèses simplificatrices, il reste encore beaucoup d'aspects théoriques et pratiques à étudier, en ce qui concerne la dynamique des champs neuronaux. L'argument ultime pour motiver l'intérêt vers les modèles de champs neuronaux dynamiques est que nous sommes intéressés à découvrir ce que ces hypothèses nous permettent de modéliser, ou autrement dit, comment nous pouvons exploiter ces hypothèses, quelle est leur capacité de calcul et quelles sont leurs limites dans une démarche informatique. Ainsi, plutôt que de souligner les déficiences de ces approches, nous nous attacherons d'abord à bien établir quelles sont leurs

qualités. Notre but général de recherche est d'étudier les phénomènes émergents qui dérivent de ces descriptions formelles de type champs neuronaux, afin d'ouvrir des pistes de recherche face aux problématiques énumérées ci-dessus.

Dans la suite de ce chapitre, nous allons tracer l'itinéraire de l'évolution des modèles de référence de champs neuronaux dynamiques, indiquant les nouveautés introduites par chaque modèle, afin de discerner les particularités et capacités de ces approches.

3.2 Modèles théoriques et comportements dynamiques

3.2.1 Modèle de Wilson et Cowan

On doit la première formalisation historique notable de la compétition entre les activités corticales aux auteurs Wilson et Cowan [125], au début des années 1970.

S'appuyant sur les résultats des travaux de Mountcastle, Hubel et Wiesel, qui étayaient l'existence des unités fonctionnelles corticales, Wilson et Cowan proposent une nouvelle vision de modélisation corticale. Faisant l'analogie avec la manière d'étudier certains phénomènes physiques, ils considèrent que pour mettre en évidence les mécanismes des processus corticaux de haut niveau, il faut modéliser l'activité du tissu cortical à une échelle spatiale plus grande que celle du neurone.

Deuxièmement, ce modèle cortical prend en compte le principe neuro-physiologique de Dale [27, 30], qui affirme que (généralement) chaque cellule neuronale se manifeste soit comme excitatrice, soit comme inhibitrice, vis-à-vis toutes les autres cellules post-synaptiques avec lesquelles elle est connectée. De plus, le caractère (excitateur ou inhibiteur) des cellules reste inchangé pendant toute leurs vies. En outre, Wilson et Cowan soulignent que multiples résultats expérimentaux indiquent que l'activité corticale est principalement la conséquence des interactions entre ces cellules excitatrices et inhibitrices.

Ces considérations ont permis à Wilson et Cowan d'élaborer le premier formalisme de référence de champs neuronaux dynamiques, qui modélise la compétition entre les populations corticales excitatrices et inhibitrices. Ainsi, ce modèle comporte deux ensembles d'unités (de type E - excitatrices et I - inhibitrices), réparties suivant une topologie bidimensionnelle. Toutefois, basé sur des arguments d'homogénéité spatiale, le formalisme est décrit et analysé pour le cas 1D, ce qui pourrait correspondre en réalité à une bande très fine (de l'épaisseur d'une seule colonne) de tissu cortical. Pour encore faciliter l'usage pratique et l'étude mathématique du modèle, les auteurs ont opté pour la description du comportement du champ avec des équations temporelles « lentes » (en anglais *time coarse-grained equations*), qui moyennent les variations temporelles des variables dans une certaine fenêtre de temps.

Utilisant les notations introduites dans la section 3.1, le formalisme proposé par Wilson et Cowan peut être décrit ainsi :

$$\begin{aligned} \tau \frac{du_E(x, t)}{dt} &= -u_E(x, t) \\ &\quad + (1 - r_E u_E(x, t)) \\ &\quad \times \sigma_E(+ (u_E \otimes w_{EE})(x, t) - (u_I \otimes w_{IE})(x, t) + i_E(x, t)) \end{aligned} \quad (3.12)$$

$$\begin{aligned} \tau \frac{du_I(x, t)}{dt} &= -u_I(x, t) \\ &\quad + (1 - r_I u_I(x, t)) \\ &\quad \times \sigma_I(- (u_I \otimes w_{II})(x, t) + (u_E \otimes w_{EI})(x, t) + i_I(x, t)) \end{aligned} \quad (3.13)$$

où \otimes représente une opération de convolution spatiale, généralement définie comme :

$$(f \otimes w)(x, t) = (w \otimes f)(x, t) = \int_{-\infty}^{\infty} w(y) f(x - y, t) dy \quad (3.14)$$

Dans ces équations, $u_E(x, t)$ et $u_I(x, t)$ représentent respectivement le nombre d'unités excitatrices et inhibitrices, actives à l'instant t et à l'endroit x du champ. $i_E(x, t)$ et $i_I(x, t)$ désignent les stimuli afférents instantanés des deux unités corticales, placées à l'endroit x du champ. w_{EE}, w_{II}, w_{EI} et w_{IE} sont les noyaux des poids des connexions latérales entre les deux types d'unités du champ (excitatrices et inhibitrices). σ_E et σ_I sont les fonctions de transfert pour les deux ensembles d'unités. Pour finir, r_E et r_I sont les périodes réfractaires pour les deux types de populations, et τ est la constante de temps membranaire.

S'intéressant aux phénomènes dynamiques globaux de l'activité du champ décrite par les équations 3.12 et 3.13, Wilson et Cowan développent une analyse mathématique détaillée, afin de pouvoir étudier l'influence des paramètres dans le comportement du champ. Ainsi, grâce à cette analyse, les auteurs mettent en évidence trois modes de fonctionnement dynamique. Les outils mathématiques qui leur ont permis d'obtenir les résultats résumés ci-dessous sont présentés en détail dans la section 3.3.

Dans le cas du premier comportement dynamique observé (voir la figure 3.2 (a)), une brève stimulation du champ, suffisamment élevée, induit une réponse forte du champ. Même après l'arrêt de la stimulation, cette réponse continue à augmenter, atteint un niveau maximal, puis descend vers son état de repos. Pour cette raison, ce comportement du champ est appelé *activation transitoire*. L'activation et la durée de ce mode dynamique du champ dépend à la fois de l'intensité et de la durée du stimulus appliqué au champ.

Le deuxième comportement dynamique peut être mis en évidence en cas de stimulation constante et continue d'une zone quelconque du champ. Dans certaines conditions de paramétrage, la réponse du champ est caractérisée par l'oscillation continue des activités excitatrices et inhibitrices (voir la figure 3.2 (b)). Par conséquent, ce mode dynamique du champ est défini par des *cycles limites*. La fréquence d'oscillation est proportionnelle à l'intensité de la stimulation, mais elle reste indépendante de la largeur de celle-ci. Dans d'autres conditions de paramétrage, le même type de stimulation de champ peut provoquer la *propagation des ondes d'activités* excitatrices et inhibitrices dans le champ (voir la figure 3.2 (c)).

Le dernier type de comportement dynamique, évoqué par Wilson et Cowan, est caractérisé par des *états spatiaux stables non homogènes*. Dans ce cas, la stimulation du champ génère une réponse forte du champ, cette réponse persistant au même endroit et au même niveau d'activation, même après que la source de stimulation a complètement disparu (voir la figure 3.2 (d)).

Notons que l'existence de chacun de ces régimes dynamiques est conditionnée par le dépassement de certains seuils temporels ou d'intensité concernant la stimulation du champ.

Les deux auteurs considèrent que ces trois comportements (mis en évidence pour trois ensembles différents de paramètres du modèle) peuvent être vus comme modélisant trois spécialisations anatomiques différentes du tissu cortical. La plausibilité biologique de cette affirmation s'étaie sur la similarité observée entre les résultats numériques générés par la simulation du modèle de champ, et les différents résultats expérimentaux sur le fonctionnement cortical.

Dans leur analyse, Wilson et Cowan proposent aussi des variantes plus élaborées du modèle, qui prennent en compte par exemple, la vitesse de propagation latérale de l'activité neuronale. Toutefois, c'est seulement récemment que les approches de modélisation ont reconsidéré l'intégration de ces aspects dans l'expression analytique du champ, ainsi que leur impact dans le comportement dynamique du champ. Souvent, les études faisant référence au modèle de Wilson

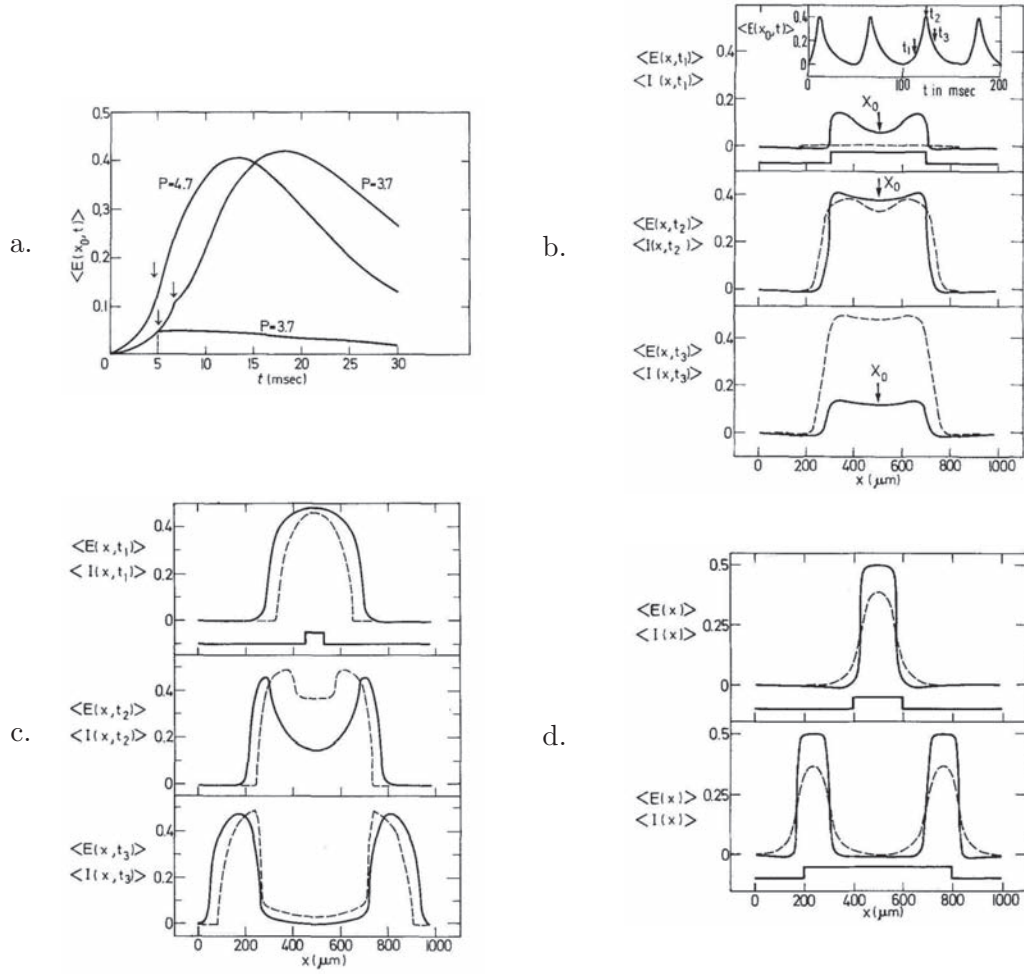


FIG. 3.2 – Illustrations graphiques, extraites de [125], présentant divers comportements dynamiques du modèle de Wilson et Cowan. Les paramètres affichés ici, c'est-à-dire $\langle E(x, t) \rangle$ (ligne continue), $\langle I(x, t) \rangle$ (ligne pointillée) et P (ligne continue, figurée tout en bas des figures (b,c,d)), sont les correspondants des paramètres $u_E(x, t)$, $u_I(x, t)$ et $i_E(x, t)$, présents dans les équations 3.12 et 3.13. Les flèches marquent le moment d'arrêt de la stimulation. (a) régime d'activation transitoire, pour différentes amplitudes P d'un stimulus (centré à l'endroit x_0 du champ), qui a été présenté pour différentes périodes de stimulation ; (b) trois phases successives de l'évolution du champ, pendant une seule période d'un régime oscillateur ; dans ce cas, le stimulus (centré à l'endroit x_0 du champ) est maintenu en permanence ; (c) génération d'une paire d'ondes de propagation, due à la brève présentation d'un stimulus externe ; (d) deux états spatiaux stables non homogènes, générés par la présentation de deux stimulations différentes ; une fois générées, les activités excitatrices et inhibitrices du champ persistent en absence des stimuli initiaux.

et Cowan préfèrent simplifier les équations, considérant les périodes de réfraction r_E et r_I nulles, et les propagations latérales instantanées.

Intégrant plusieurs aspects essentiels de la modélisation corticale, le formalisme proposé par Wilson et Cowan est devenu en cours du temps un modèle de référence des champs neuronaux dynamiques, et comme nous allons le constater dans les prochains paragraphes, une source d'inspiration fondamentale pour la suite des modèles proposés dans la littérature.

En résumé, *les contributions de Wilson et Cowan abordent trois aspects principaux de modélisation : (a) la description de type large échelle de l'activité neuronale du tissu cortical par un champ continu d'un ensemble d'unités fonctionnelles génériques ; (b) la séparation des influences neuronales des populations excitatrices et inhibitrices ; (c) la description de l'activité de chaque unité du champ par des équations temporelles « lentes », désignant des contributions neuronales moyennes.*

3.2.2 Modèle d'Amari

Inspiré par les travaux de plus en plus nombreux visant la modélisation du tissu cortical, Amari propose en 1977 un nouveau modèle de champ neuronal [4], basé sur les mêmes principes de compétition et coopération entre les activités excitatrices et inhibitrices des neurones que le modèle de Wilson et Cowan.

Généralisant la vision de modélisation fonctionnelle utilisée par Wilson et Cowan, Amari propose un cadre formel étendu pour décrire l'activité neuronale d'un champ multi-couches, c'est-à-dire une couche pour chaque ensemble de neurones corticaux du même type. Les couches d'un tel champ ne correspondent pas nécessairement aux six couches anatomiques du tissu cortical, la séparation proposée par Amari étant plutôt une méthode simplifiée pour formaliser et étudier de manière unifiée les interactions de nature corticale.

Afin de faciliter encore plus l'étude analytique, Amari considère le cas particulier d'un champ unidimensionnel mono-couche (intégrant dans la même couche les contributions excitatrices et inhibitrices), où les connexions latérales à longue distance entre les unités fonctionnelles du champ sont explicitement de nature inhibitrice. Cette condition revient alors à un noyau de poids de connexion de type chapeau mexicain (voir la fonction 3.3). Considérant aussi les hypothèses d'homogénéité spatiale du champ et la propagation instantanée des activités neuronales, Amari décrit son modèle de base par l'équation suivante :

$$\tau \frac{du(x,t)}{dt} = -u(x,t) + \int_{-\infty}^{\infty} w(|x-x'|)\sigma(u(x',t))dx' + i(x,t) + h \quad (3.15)$$

où la seule nouvelle notation introduite ici est h , désignant le niveau de repos de l'activité du champ.

Alors que Wilson et Cowan basent leur approches sur les aspects structuraux de la modélisation corticale, ou la plausibilité biologique de leur modèle, Amari vise avec ce nouveau formalisme une analyse théorique détaillée des aspects dynamiques du comportement du champ, notamment l'existence et la stabilité des solutions de l'équation du champ.

Grâce à cette étude analytique, Amari met en évidence cinq principaux comportements dynamiques, c'est-à-dire cinq types de solutions d'équilibre pour l'équation du champ. Une première solution, notée ϕ – *solution*, est caractérisée par une activité nulle pour toutes les unités du champ. Au contraire, la deuxième solution, notée ∞ – *solution*, présente un niveau d'excitation maximal pour chaque unité du champ. Ces deux types de solutions sont monostables. Pour le troisième type de solution, une forte stimulation du champ déclenche une propagation illimitée

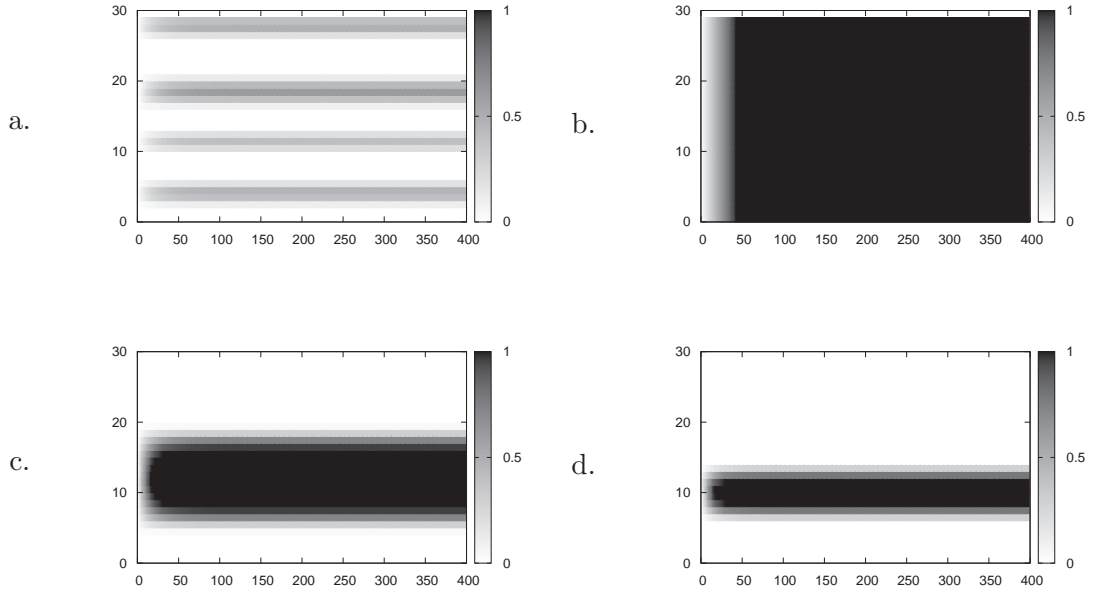


FIG. 3.3 – Différentes solutions $u(x, t)$ de l'équation 3.15, obtenues dans différentes conditions de paramétrage. Les solutions (a-c) sont obtenues en absence de stimulation externe, alors que la solution (d) est le résultat d'une stimulation externe fixe et permanente du champ. Chaque graphique montre en niveau de gris l'intensité de la réponse $u(x, t)$ du champ. L'abscisse représente le temps t en nombre de pas de simulation, et l'ordonnée représente l'ensemble des positions $\{x \in X\}$ du champ. a. instabilités de Turing en domaine spatial ; b. ∞ – solution ; c. a – solution en absence d'entrée ; d. a – solution en présence d'une stimulation de type gaussienne, placée à l'endroit $x_0 = 10$ du champ.

de l'excitation tout au long du champ. Si la stimulation initiale a un niveau plus faible, elle déclenche une réponse qui s'évanouit en cours du temps (si cette stimulation est maintenue toujours à un niveau faible). Une quatrième solution possible est caractérisée par une zone finie d'excitation locale, appelée souvent bulle d'excitation. Cette solution, qui peut aussi s'évanouir en cours de temps, est notée a – solution si elle s'étend sur une zone de longueur a . Ces deux dernières solutions possibles sont donc bistables. Finalement, le cinquième type de solution maintient une excitation périodique en domaine spatial (des zones d'excitations locales apparaissent partout dans le champ). Ce type de comportement est appelé instabilité de Turing. La figure 3.3 montre graphiquement certaines de ces solutions possibles de l'équation 3.15.

Parmi ces différents comportements que l'équation 3.15 peut exhiber, nous nous focalisons dans les paragraphes suivants sur l'analyse de seulement deux de ces phénomènes dynamiques : la formation des bulles stables d'excitation locale et l'émergence des instabilités de Turing. Comme nous allons présenter ensuite dans le chapitre 4, de multiples travaux concernant la simulation ou l'interprétation des processus physiologiques ou cognitifs font appel à une modélisation qui exhibe ces deux comportements.

Formations stables d'excitation locale

Un premier comportement remarquable du modèle d'Amari concerne l'émergence des formations stables d'excitation locale dans la réponse $\{u(x, t)\}_{x \in X}$ du champ. Ces formations, notées *a* – *solutions* dans les travaux d'Amari, sont souvent connues dans la littérature sous le nom de *bulles* d'activité (en anglais *bumps* ou *bubbles*). Ces formations spatiales sont des zones locales d'unités du champ qui augmentent progressivement leur niveau d'activité neuronale, jusqu'à attendre un seuil de saturation à l'équilibre.

Pour étudier les propriétés des formations stables d'excitation locale, Amari considère deux cas séparés d'analyse : la réponse du champ respectivement en absence totale et, en présence locale d'une stimulation externe stationnaire *i*. Pour des raisons de simplicité mathématique, son analyse est réalisée pour une fonction de transfert $\sigma = \sigma_H$ (voir l'équation 3.5), argumentant toutefois qu'en considérant n'importe quelle autre fonction σ de type sigmoïde, les résultats analytiques seront équivalents du point de vue qualitatif. Résumons ici ces résultats.

Comme mentionné, dans le premier cas d'analyse, Amari considère $i(x, t) = 0, \forall x \in X, \forall t \geq 0$. Notons :

$$R[u] = \{x \in X \mid u(x) > 0\} \quad (3.16)$$

la région excitée du champ. Une solution locale de type bulle d'activité de longueur *a* doit respecter à l'équilibre :

$$u(x, t) = \int_{R[u]} w(|x'|) dx' + h \quad (3.17)$$

Sans restreindre la généralité, nous pouvons considérer que la zone excitée occupée par la bulle est exactement l'intervalle $R[u] = (0, a) \subset \mathbb{R}$. Notons ensuite :

$$\begin{aligned} W(x) &= \int_0^x w(|x'|) dx' \\ W_{max} &= \max_{x>0} W(x) \\ W_{min} &= \min_{x>0} W(x) \\ W_\infty &= \lim_{x \rightarrow 0} W(x) \end{aligned}$$

Suite à son analyse, Amari établit les conditions d'existence des solutions que nous avons décrites précédemment :

$$\begin{aligned} \phi - \text{solution} &\Leftrightarrow h < 0 \\ \infty - \text{solution} &\Leftrightarrow 2W_\infty > -h \\ a - \text{solution} &\Leftrightarrow h < 0, a > 0, W(a) + h = 0 \end{aligned}$$

Plusieurs bulles peuvent émerger et coexister à l'équilibre dans le champ, à condition que l'inhibition à longue distance soit suffisamment faible (pour que les contributions excitatrices locales puissent dominer d'autres régions du champ). Ainsi, la densité des bulles d'activation dans le champ est dictée par le profil du noyau des poids des connexions latérales, c'est-à-dire le rapport entre l'influence cumulée des poids des connexions excitatrices et celle des poids des connexions inhibitrices.

Ensuite, Amari étudie la stabilité de ces formations d'excitation locale. Dans son approche, il calcule le déplacement des bords d'une région excitée, déplacement induit par un petit changement de l'activité neuronale des unités du champ. Pour que la bulle formée reste stable, ce

déplacement doit converger vers zéro. Ceci se traduit dans la condition suivante :

$$\frac{dW(a)}{da} < 0$$

Dans le deuxième cas d'analyse, Amari s'intéresse aux phénomènes transitoires qui précèdent l'atteinte d'un état d'équilibre d'une bulle d'activité, dans un contexte caractérisé par une stimulation locale stationnaire du champ ($i(x, t) = i(x)$). Dans ces conditions, une bulle qui commence à se former dans le champ est entraînée dans un déplacement spatial local, afin de trouver un état (y compris une position) d'équilibre. Elle se dirige vers un endroit où l'influence de la stimulation i est la plus grande. En conséquence, si une bulle quelconque se forme au début sur l'intervalle $(x_1(t), x_2(t))$, elle va se déplacer vers la gauche si $i(x_1(t)) > i(x_2(t))$, ou vers la droite dans le cas contraire. Ce mouvement cesse quand les influences des stimuli externes ressentis par la bulle sont équilibrées.

En dehors des stimulations externes, par définition, la réponse d'un champ neuronal dépend des interactions entre les activités de toutes ses unités. Ainsi, dès leur apparition, les bulles d'activation interagissent. Ainsi, si deux bulles sont suffisamment proches, elles vont se rapprocher et fusionner en une seule, afin d'atteindre un état d'équilibre. Si elles se trouvent plus loin une de l'autre, sans que la distance entre elles ne dépasse un certain seuil, les deux bulles vont se repousser pour rétablir l'équilibre dynamique du champ. Finalement, si les deux bulles ont émergé suffisamment loin une de l'autre, elles ne vont pas du tout interagir. Comme dans le cas d'analyse précédent (d'absence totale de stimulation), le profil du noyau des poids des connexions latérales impose des contraintes pour la densité des bulles d'activité qui peuvent coexister dans le champ.

Instabilités de Turing

Un deuxième phénomène dynamique révélé par le comportement des champs neuronaux d'Amari se réfère aux *instabilités de Turing*. Dans certaines conditions de paramétrage et dans un contexte d'absence totale de stimulation externe, le champ entier devient « pavé » avec des motifs d'activation stationnaires périodiques dans le domaine spatial (voir la figure 3.3 (a)).

Les mêmes phénomènes dynamiques avaient été mis en évidence par Turing [119] dans sa proposition de modéliser certains processus de morphogenèse par un système d'équations de type *réaction-diffusion*. En résumé, Turing modélise l'interaction entre deux principaux processus chimiques qui interviennent dans l'étape de développement de la forme d'un organisme : la production locale des particules de substance (due aux réactions chimiques), et la propagation de ces particules dans l'environnement interne de l'organisme (diffusion spatiale). Les résultats de cette modélisation montrent la formation d'une distribution de concentration de substance qui est caractérisée par des zones de forte concentration (bandes ou bulles) qui « sillonnent » de manière périodique tout le champ. Ce phénomène porte, depuis les travaux de Turing, le nom d'instabilités de Turing, et montre comment des états stables non-uniformes peuvent être générés à partir des états homogènes.

Cette étude montre également comment l'agrégation à large échelle de comportements locaux, décrits par des instructions « simples », peut engendrer des comportements complexes, tout à fait nouveaux. Notons que cet intérêt de Turing pour le concept de l'émergence est présent dans plusieurs de ses travaux. Par exemple, dans un autre ouvrage de référence pour le domaine du traitement de l'information [117], Turing propose un mécanisme général de calcul, connu actuellement sous le nom de *machine de Turing*. Ce mécanisme montre comment des procédures de calcul très complexes peuvent être implémentées utilisant un ensemble minimal d'opérations élémentaires.

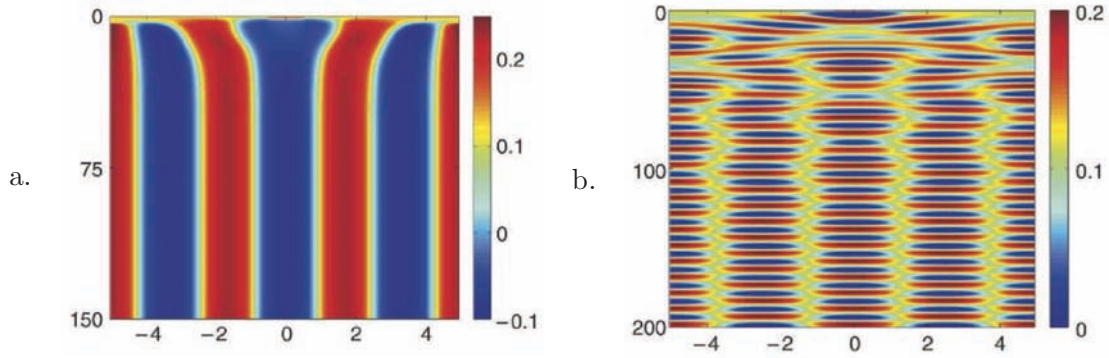


FIG. 3.4 – Illustrations graphiques, extraites de [128], présentant deux types d’instabilités de Turing : a. spatiale ; b. temporelle. Chaque graphique montre l’amplitude de la réponse $u_E(x, t)$ de la couche excitatrice du champ neuronal. Cette fois-ci, l’abscisse représente l’ensemble des positions $\{x \in X\}$ du champ, et l’ordonnée représente le temps t , en nombre de pas de simulation.

De manière intuitive, les processus de compétition de type excitation/inhibition de nature corticale sont analogues aux processus chimiques de type réaction-diffusion décrits ci-dessus. Avec cette analogie, il est possible de montrer que le comportement d’un champ neuronal en absence d’une stimulation externe est capable d’exhiber aussi des phénomènes d’instabilités de Turing. Amari est un des premiers chercheurs à étudier analytiquement les conditions de manifestation de ce phénomène pour un modèle de champ neuronal. Maintenant il est bien connu que d’autres modèles exhibent aussi des instabilités de Turing. Par exemple, la même propriété a été mise en évidence pour un modèle légèrement différent que celui proposé par Wilson et Cowan [128]. En outre, les graphiques de la figure 3.4, extraits de [128], mettent en évidence l’existence des instabilités de Turing non seulement dans le domaine spatial (voir le graphique (a)), mais aussi dans le domaine temporel (voir le graphique (b)).

Pour une étude approfondie sur le sujet des instabilités de Turing, le lecteur est invité à consulter en détail [128]. Nous en retiendrons ici que l’émergence de ce phénomène dynamique dépend de deux principaux facteurs : le rapport entre les contributions synaptiques excitatrices et inhibitrices (rapport dicté par le profil du noyau des poids des connexions latérales), et la pente des fonctions de transfert sigmoïdes qui modélisent le taux de décharge neuronal.

Comme décrit ci-dessus, dans les cas d’instabilité de Turing, les motifs d’activité neuronale qui émergent dans un champ se manifestent avec une grande régularité spatiale ou temporelle. Cet aspect peut être interprété comme un processus d’auto-organisation de l’activité corticale. En effet, ce comportement dynamique a été utilisé pour expliquer certains phénomènes physiologiques corticaux. Ceci, ainsi que d’autres usages appliqués des phénomènes d’instabilités de Turing sont présentés en détail dans le chapitre 4.

En résumé, Amari propose un modèle unifié des contributions excitatrices et inhibitrices d’un champ neuronal unidimensionnel, utilisant un noyau des poids des connexions latérales de type chapeau mexicain. Grâce à cette manière d’exprimer le formalisme des interactions neuronales, Amari réussit à étudier dans un cadre théorique étendu plusieurs aspects clé concernant la dynamique des solutions des équations du champ. En particulier, dans ses travaux de référence, il met en évidence des phénomènes dynamiques comme par exemple la formation des bulles locales d’activité neuronale ou l’apparition des instabilités de Turing, donnant aussi les conditions

d'existence et de stabilité de ces comportements dynamiques.

Les deux formalismes neuronaux décrits jusqu'ici, c'est-à-dire celui proposé par Wilson et Cowan et celui d'Amari, sont considérés comme étant les modèles fondateurs de la théorie des champs neuronaux dynamiques. Dans la suite, nous allons présenter d'autres modèles théoriques de référence qui leur ont succédé dans la littérature du domaine, proposés afin d'améliorer, d'enrichir, d'intégrer ou de faciliter l'étude de certaines caractéristiques dynamiques de modélisation ou d'analyse théorique.

3.2.3 Extension bidimensionnelle des modèles de champs neuronaux

Analyse de Taylor

Si les formalismes de champ proposés par Wilson et Cowan ou Amari sont décrits pour une topologie générale du champ, ces auteurs analysent dans leurs travaux exclusivement le cas des champs unidimensionnels. Alors que ce contexte facilite l'étude analytique de la dynamique des modèles proposés, les champs neuronaux sont censés modéliser la structure corticale. Assez récemment, cet aspect a motivé les chercheurs à poursuivre une analyse plus détaillée du comportement des champs bidimensionnels.

Une première démarche dans ce sens est due à Taylor [114], qui étend le modèle d'Amari pour le cas $X = \mathbb{R}^2$, reprenant aussi l'analyse menée par Amari à propos de l'existence et de la stabilité des solutions du champ. Résumons dans ce qui suit ces résultats.

Partant des mêmes hypothèses de modélisation qu'Amari (c'est-à-dire l'homogénéité de la topologie du champ et une connectivité d'inhibition longue distance de type chapeau mexicain entre les unités de champ), Taylor montre que le champ doit respecter la même solution d'équilibre que celle décrite par Amari dans 3.17 pour le cas unidimensionnel.

Considérons maintenant, sans restreindre la généralité, que $R[u]$ (la zone d'excitation du champ, définie par 3.16) correspond à la zone $D_R = \{x \in \mathbb{R}^2 \mid |x| < R\}$. Notons :

$$\begin{aligned} W_R(x) &= \int_{D_R} w(|x - x'|) dx' \\ W_\infty &= \lim_{R \rightarrow \infty} W_R(x) \end{aligned}$$

Suite à son analyse, Taylor met en évidence les conditions d'existence des solutions du champ, conditions analogues à celles trouvées par Amari :

$$\begin{aligned} \phi - \text{solution} &\Leftrightarrow h < 0 \\ \infty - \text{solution} &\Leftrightarrow W_\infty > -h \\ D_R - \text{solution} &\Leftrightarrow h < 0, R > 0, W_R(R) + h = 0 \end{aligned}$$

De manière similaire, Taylor déduit aussi les conditions de stabilité des $D_R - \text{solutions}$:

$$\frac{dW_R(x)}{dR} < 0$$

Analyse de Werner et Richter

Malgré l'analyse de Taylor, qui montre une analogie parfaite entre ses résultats et ceux d'Amari, Werner et Richter mettent en évidence dans [123] le fait que ces solutions ne sont pas les seules possibles pour un champ bidimensionnel décrit par l'équation 3.15. En particulier, Werner et Richter abordent l'analyse dynamique du champ en cherchant des solutions symétriques radiales.

Reprenant les arguments d'Amari, la solution $u(x)$ est une solution stationnaire de type D_R – solution si :

$$u(x) = \int_{D_R} w(|x - x'|) dx' + h \quad (3.18)$$

ce qui devient vrai si et seulement si toutes les trois conditions suivantes sont remplies :

$$W_R(x) > -h, \text{ pour } |x| < R \quad (3.19)$$

$$W_R(R) = -h \quad (3.20)$$

$$W_R(x) \leq -h, \text{ pour } |x| > R \quad (3.21)$$

Notons que, pour w de type chapeau mexicain et pour tout x suffisamment grand, $W_R(x)$ s'approche de zéro, restant toujours négatif. Par conséquent, pour satisfaire la condition 3.21, il est nécessaire que h soit négatif.

Amari, dans son analyse unidimensionnelle du champ, montre que pour $h < 0$ et la condition 3.20 satisfaite, les conditions 3.19 et 3.21 sont implicitement remplies. Toutefois, cette affirmation peut devenir fausse dans le cas bidimensionnel. Si l'inégalité $W_R(0) > W_R(R)$ est toujours valable dans le cas unidimensionnel, elle n'est pas nécessairement vraie aussi dans le cas bidimensionnel. Si $W_R(0) \leq W_R(R)$ et $W_R(R) = -h$, la condition 3.19 ne peut pas être satisfaite. Dans ce cas, $u(x)$ est bien nul aux frontières de D_R et négatif dehors D_R , mais pas positif dans tout l'intérieur de D_R . Cette conclusion met en évidence l'existence des solutions de type *anneau d'excitation* pour un champ bidimensionnel de type Amari.

Concernant la stabilité des D_R – solutions, la conclusion d'Amari change dans la mesure où, pour le cas bidimensionnel, la condition suivante est seulement nécessaire, mais pas aussi suffisante :

$$\frac{dW_R(x)}{dR} < 0$$

Une analyse plus détaillée de Werner et Richter montre qu'une nouvelle dynamique peut émerger dans le comportement du champ, dynamique caractérisée par une triple stabilité : dépendant des conditions d'entrée, l'excitation émergée dans le champ peut s'évanouir, s'étendre sur tout le champ ou se stabiliser sous une forme locale d'excitation.

En conclusion, nous pouvons affirmer que les travaux de Werner et Richter mettent en évidence de nouveaux comportements dynamiques, dont la manifestation est intrinsèque à la topologie du champ.

3.2.4 Modèle d'Erlhagen

Nous résumons ici l'approche de modélisation des champs neuronaux dynamiques employée dans les travaux d'Erlhagen et de ses collaborateurs. Dans ses recherches [80, 35, 36, 61], Erlhagen est principalement intéressé par l'utilisation des paradigmes dynamiques de traitement neuronal, et notamment des champs neuronaux, pour développer des comportements robotiques autonomes et adaptatifs, avec des capacités cognitives.

S'inspirant des modèles de ses prédécesseurs, Erlhagen propose dans [35] un modèle de champ où les activités des unités excitatrices et inhibitrices sont exprimées selon les deux équations suivantes :

$$\begin{aligned} \tau \frac{du(x, t)}{dt} = & -u(x, t) + i(x, t) + h \\ & + g(u(x, t)) \times \left[\int_{-\infty}^{\infty} w_u(|x - x'|) \sigma(u(x', t)) dx' - v(x, t) \right] \end{aligned} \quad (3.22)$$

$$\tau \frac{dv(x, t)}{dt} = -v(x, t) + \int_{-\infty}^{\infty} w_v(|x - x'|) \sigma(u(x', t)) dx' \quad (3.23)$$

La particularité de ce formalisme réside dans la modulation non-linéaire g (une fonction sigmoïde) de l'influence cumulée des unités voisines du champ dans l'activité de chaque unité excitatrice. D'après Erlhagen, g pourrait être interprété comme un mécanisme neuronal de feedback latéral qui aurait le rôle d'intégrer le contexte spatial du traitement neuronal, contexte partagé par les champs récepteurs des unités voisines du champ.

Selon le seuil de la sigmoïde g , le champ décrit par les équations 3.22 et 3.23 exhibe deux régimes fonctionnels. Si ce seuil est assez élevé, comme dans le cas du modèle de Wilson et Cowan, le champ manifeste un comportement de type *activation transitoire* si stimulé brièvement avec un signal externe local suffisamment fort. Dans les mêmes conditions de paramétrage, la réponse du champ est capable de suivre un stimulus local et fort qui se déplace de manière continue le long du champ. Au contraire, dans un deuxième régime de fonctionnement, si le seuil de la fonction g est suffisamment petit, le champ produit des ondes d'activation qui s'auto-entretiennent et continuent à se déplacer suivant la direction indiquée par la stimulation externe, même quand celle-ci n'est plus active.

Le deuxième mode dynamique manifesté par ce modèle de champ pourrait être interprété comme un mécanisme d'extrapolation vis-à-vis du processus de stimulation externe. Avec ce modèle, Erlhagen offre une base de modélisation neuronale pour développer des mécanismes génériques d'anticipation et d'extrapolation, ce qui faciliterait ensuite l'implantation des boucles de prédiction dans des architectures neuronales.

L'approche présentée succinctement ici fait partie intégrante d'une vision plus large que Erlhagen formule dans [36], et qui expose une approche générique du développement des solutions robotiques, basée sur le paradigme des champs neuronaux dynamiques. Nous résumons cette vision, ainsi que les résultats expérimentaux résultant de la mise en œuvre de cette approche dans le chapitre 4.

Pour résumer, retenons ici que la démarche d'Erlhagen met en évidence l'importance des modulations non-linéaires qui peuvent influencer essentiellement l'intégration des différentes contributions neuronales dans l'activité globale d'un champ.

3.2.5 Modèle de Pinto et Ermentrout

Pour des raisons de simplicité de formalisation ou d'analyse théorique, tous les modèles de champ présentés jusqu'ici prennent en compte un nombre relativement restreint d'hypothèses de modélisation, ce qui limite la diversité de comportements dynamiques possibles de ces modèles. Même si plusieurs aspects dynamiques ont été mis en évidence et étudiés dans le cadre de ces hypothèses, nous constatons un intérêt récent croissant vers l'intégration des aspects plus raffinés dans la modélisation des activités neuronales.

Une approche de référence qui suit cette tendance appartient à Pinto et Ermentrout. Dans [92, 93], les auteurs analysent, un par un, différents aspects qui pourraient être inclus dans la modélisation de l'activité de champ, examinant ensuite les conditions nécessaires ou suffisantes pour générer certains comportements dynamiques.

Une première variante d'extension proposée par Pinto et Ermentrout se réfère à la vitesse de propagation de l'activité neuronale au long de l'axone. L'expression analytique de l'activité du champ change dans ce cas comme suit :

$$\tau \frac{du(x, t)}{dt} = -u(x, t) + i(x, t) + h + \int_{-\infty}^{\infty} w(|x - x'|) \sigma(u(x', t - \frac{|x - x'|}{v})) dx' \quad (3.24)$$

Les auteurs s'intéressent dans ce cas à l'existence des solutions u de type front d'onde progressive, c'est-à-dire de forme $u(x - ct)$, où $u(-\infty) = 1$, $u(\infty) = 0$ et $u(0) = \theta$ (ici, c désigne la vitesse de propagation du front de l'onde, et θ le seuil de l'activité neuronale du front). Analysant le comportement dynamique du champ dans un contexte simplifié (la sigmoïde $\sigma(u) = \sigma_H(u - \theta)$ de type Heaviside avec le seuil θ , et le niveau de repos d'activité neuronale h , nul) et en absence de stimulation externe, les auteurs constatent que la vitesse de propagation par l'axone, v , peut avoir une influence sur la vitesse c de l'onde d'activité seulement si ces deux variables (v et c) ont des valeurs numériques très proches. Toutefois, les enregistrements physiologiques expérimentaux montrent un rapport d'environ deux ordres de magnitude entre les vitesses v et c . Par conséquent, Pinto et Ermentrout concluent que, pour la majorité des cas, la vitesse v pourrait être considérée infinie.

Une deuxième approche concerne la prise en compte de l'influence des courants ioniques et synaptiques dans la génération de l'activité neuronale :

$$\begin{aligned} \tau \frac{du(x, t)}{dt} = & g_L(u_L - u(x, t)) + i(x, t) + h \\ & + g_{syn} \left(\int_{-\infty}^{\infty} w(|x - x'|) \sigma(u(x', t - \frac{|x - x'|}{v})) dx' \right) (u_{syn} - u(x, t)) \end{aligned} \quad (3.25)$$

où g_L et u_L représentent la conductance et le potentiel de fuite (en anglais *leak*), et g_{syn} et u_{syn} représentent la conductance et le potentiel synaptiques.

Dans ce cas, une analyse similaire à celle du premier cas conclut sur le fait que varier le potentiel synaptique est équivalent à la modification du seuil de la fonction non-linéaire σ .

Dans un troisième temps, les auteurs examinent les effets de l'introduction d'une influence inhibitrice de feedback sur l'activité du champ :

$$\tau \frac{du(x, t)}{dt} = -u(x, t) + i(x, t) + h + \int_{-\infty}^{\infty} w(|x - x'|) \sigma(u(x', t)) dx' - v(x, t) \quad (3.26)$$

$$\frac{1}{\eta} \frac{dv(x, t)}{dt} = -v(x, t) + u(x, t) \quad (3.27)$$

D'après les auteurs, v pourrait représenter l'adaptation de la fréquence de décharge neuronale, la dépression synaptique, ou un autre processus lent qui limite l'excitation dans le réseau.

Cette fois-ci, les auteurs s'intéressent à l'existence des impulsions progressives dans le champ. Étudiant le comportement du champ dans les mêmes hypothèses simplifiées que celles mentionnées précédemment, les auteurs déduisent une relation numérique entre la vitesse de l'impulsion, v , la largeur a de cette impulsion, et le seuil de décharge de la fonction sigmoïde σ . Quand l'influence de feedback est lente (η petit), il existe deux solutions de type impulsion : une étroite et lente, et une autre large et rapide. Avec l'augmentation de η , les deux solutions convergent et disparaissent.

Dans la suite de ce document, quand nous allons faire référence au modèle de Pinto et Ermentrout, nous allons considérer implicitement le modèle de champ défini par les équations 3.26

et 3.27. Ce dernier formalisme a fait l'objet de nombreuses études, plusieurs chercheurs manifestant un intérêt croissant vers ce nouveau modèle de champ. En particulier, comme nous allons présenter dans le prochain chapitre, nos travaux ont été également beaucoup inspirés par ces approches de modélisation. Nous notons aussi que, si la plupart des études récentes de ce modèle ont porté sur les aspects analytiques de sa dynamique (voir la section 3.3), notre recherche a été motivée par l'utilisation de ses caractéristiques dynamiques dans des contextes pratiques, appliqués, et notamment ceux qui visent la conception des mécanismes neuronaux d'auto-organisation.

En conclusion, soulignons que Pinto et Ermentrout proposent des alternatives de modélisation neuronale, capables d'intégrer le traitement des phénomènes axonaux, dendritiques ou de feedback, et d'étendre les caractéristiques dynamiques des modèles classiques des champs neuronaux.

3.2.6 Autres modèles de champ

Poussés par la nécessité d'une modélisation biologique plus détaillée, ou encore par un besoin pratique, qui demande aux modèles de champs neuronaux d'expliquer ou d'exhiber un comportement dynamique complexe, plusieurs auteurs ont proposé récemment différentes alternatives de modélisation aux modèles de champ déjà existants, afin d'étendre ou raffiner l'ensemble de leurs caractéristiques. Dans leur vaste majorité, ces propositions concernent souvent un des quatre aspects de modélisation énumérés ci-dessous :

- la dynamique des phénomènes de propagation des impulsions électriques au long de l'axone ;
- la dynamique des activités électriques provenant de l'arbre dendritique ;
- l'adaptation dynamique du niveau de repos de l'activité des cellules neuronales ;
- l'ajustement dynamique des poids synaptiques des connexions latérales entre les unités neuronales du champ.

Modélisation de la dynamique axonale

Si Pinto et Ermentrout ont accordé une importance réduite à la dynamique axonale dans leurs travaux, comme nous l'avons déjà présenté dans la section 3.2.5, d'autres auteurs reconsidèrent la pertinence de la prise en compte de cet aspect dans le formalisme d'un champ neuronal [67].

Par exemple, se basant sur le fait (montré expérimentalement) que les fibres axonales des connexions intracorticales ne sont pas myélinisées, Hutt, Bestehorn et Wennekers [67] proposent la modélisation des interactions neuronales dans la couche corticale par une équation de champ qui prend en considération la vitesse de propagation des impulsions électriques le long des axones de ses unités neuronales.

Hutt et al. décrivent le taux de décharge instantanée $P_{E,I}(x, t)$ au niveau des synapses excitatrices et inhibitrices des cellules par l'expression suivante :

$$P_{E,I}(x, t) = \int_{-\infty}^{\infty} \beta_{E,I} f_{E,I}(x, x') P(x', \frac{|x - x'|}{v}) dx' + \mu_{E,I} P^{ext}(x, t) \quad (3.28)$$

où $\beta_{E,I}$ sont des facteurs de pondération pour les fonctions $f_{E,I}$ qui définissent la probabilité des connexions synaptiques latérales, v est la vitesse de propagation des impulsions à travers l'axone, P^{ext} est une activité excitatrice externe, modulée par les constantes de couplage $\mu_{E,I}$.

Les potentiels électriques créés par ces impulsions dans les synapses peuvent ensuite être exprimés ainsi :

$$u_{E,I}(x, t) = g_{E,I} \int_{-\infty}^t h_{E,I}(t - \tau) P_{E,I}(x, \tau) d\tau + \mu_{E,I} P^{ext}(x, t) \quad (3.29)$$

où $h_{E,I}$ modélisent les fonctions de réponse impulsionnelle et $g_{E,I}$ représentent les facteurs de gain synaptique.

Rassemblant les deux équations 3.28 et 3.29, et fixant des valeurs pour certains paramètres de ces équations (voir [67]), le potentiel électrique au niveau des membranes des cellules neuronales peut être calculé suivant l'équation unifiée :

$$\begin{aligned} u(x, t) &= u_E(x, t) - u_I(x, t) \\ &= \int_{-\infty}^t dt \int_{-\infty}^{\infty} \{h(t - \tau)(a_E f_E(x, x') - a_I f_I(x, x')) \\ &\quad \times \sigma(u(x', \tau - \frac{|x - x'|}{v})) + \mu P^{ext}(x, \tau)\} dx' \end{aligned} \quad (3.30)$$

Analysant la stabilité des solutions de cette équation intégrale, Hutt et al. mettent en évidence des régimes dynamiques caractérisés par des ondes progressives, et en particulier par des phénomènes d'instabilité de Turing ou des « défauts topologiques » (pour plus de détails, voir [67]). Rappelons au final que dans ce modèle de champ la valeur $u(x, t)$ représente un potentiel électrique, et non plus un taux de décharge instantanée des impulsions nerveuses, comme dans les cas des modèles traditionnels présentés jusqu'ici.

Modélisation de la dynamique dendritique

Cette deuxième alternative de modélisation concerne les modèles de champ qui intègrent dans leur formalisme la propagation retardée des impulsions électriques le long des dendrites.

Roxin, Brunel et Hansel proposent dans [100] un modèle de champ considérant un délai temporel fixe δ nécessaire pour la propagation des activités neuronales entre chaque paire d'unités connectées du champ :

$$\tau \frac{du(x, t)}{dt} = -u(x, t) + \sigma(i(x, t) + \int_{-\infty}^{\infty} w(|x - x'|) u(x', t - \delta) dx') \quad (3.31)$$

où $\sigma(u) = u \sigma_H(u)$.

Utilisant des méthodes analytiques ou numériques, les auteurs mettent en évidence une large gamme de régimes dynamiques pour cette équation : état uniforme stationnaire, bulle stationnaire, bulle oscillante, état uniforme oscillant, ondes stationnaire, ondes de propagation, ondes vacillantes, et état dynamique aperiodique (voir la figure 3.5).

Dans une analyse complémentaire à celle-ci, Roxin et al. examinent une modélisation plus réaliste de l'équation 3.31, avec des neurones impulsionnels (dont les activités sont décrites par un modèle de type Hodgkin-Huxley). Les résultats de cette deuxième analyse montrent des comportements dynamiques généralement analogues à ceux observés dans le premier cas. Cette conclusion conduit ainsi à considérer suffisante la modélisation avec des unités neuronales fréquentielles, afin d'assurer un comportement dynamique qualitativement pertinent pour le modèle de champ proposé. L'introduction d'une modélisation simple des phénomènes dendritiques rend le modèle plus riche et plus réaliste au niveau dynamique.

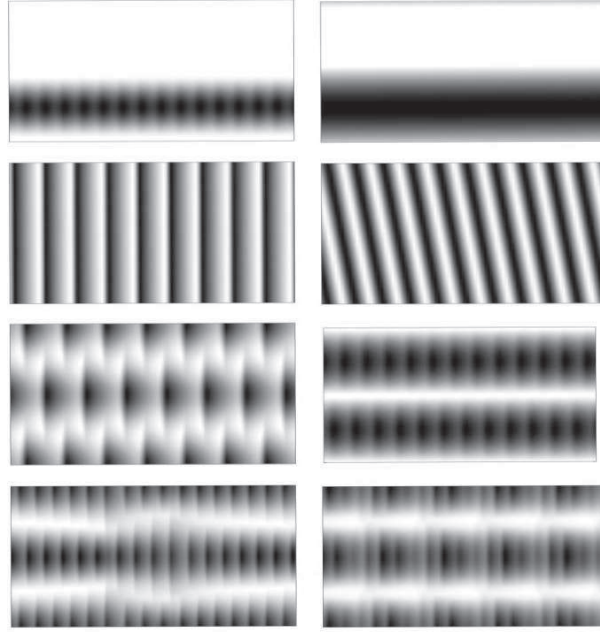


FIG. 3.5 – Illustrations graphiques, extraites de [100], présentant divers comportements dynamiques du modèle de Roxin et al. Pour chacun de ces graphiques, l’abscisse représente le temps t , et l’ordonnée représente l’espace x . L’amplitude de l’activité neuronale $u(x, t)$ est représentée par niveaux de gris. Du haut vers le bas, les graphiques montrent, pour la colonne gauche : bulle oscillante, état uniforme oscillant, ondes stationnaires et état dynamique aperiodique ; et pour la colonne droite : bulle stationnaire, ondes de propagation, ondes stationnaires et état dynamique aperiodique.

Dans le même contexte de modélisation, nous notons également ici les résultats de Bressloff et Folias [13, 39, 40], qui examinent en détail le comportement dynamique du modèle de Pinto et Ermentrout, pour lequel ils interprètent la contribution $v(x, t)$ comme étant une influence inhibitrice ressentie par chaque neurone du champ, due aux phénomènes de propagation dendritique (voir les équations 3.26 et 3.27). Dans ces travaux, Bressloff et Folias analysent l’influence des entrées non-homogènes sur la stabilité des activités impulsives stationnaires du champ. Suite à cette analyse, ils montrent que (a) une non-homogénéité suffisamment grande de la stimulation externe peut induire une stabilisation des réponses impulsives stationnaires et (b) une réduction graduelle de la non-homogénéité de la stimulation externe peut induire la formation des fronts d’ondes oscillants, appelés en anglais *breathers* (voir la figure 3.6).

Modélisation de la dynamique du niveau de repos neuronal

Comme établi dans le paragraphe 3.2.2, Amari montre que la condition d’existence et de stabilité d’une bulle d’excitation, pour l’activité d’un champ en absence de stimulation externe, se traduit par une relation directe entre le niveau de repos général du champ et la forme du noyau des poids des connexions latérales. Ce fait pourrait sous-entendre que le niveau de repos neuronal h a une valeur prédéterminée. Toutefois, certains auteurs ont considéré approprié d’inclure et d’examiner dans leurs approches de modélisation l’ajustement dynamique du niveau h de repos

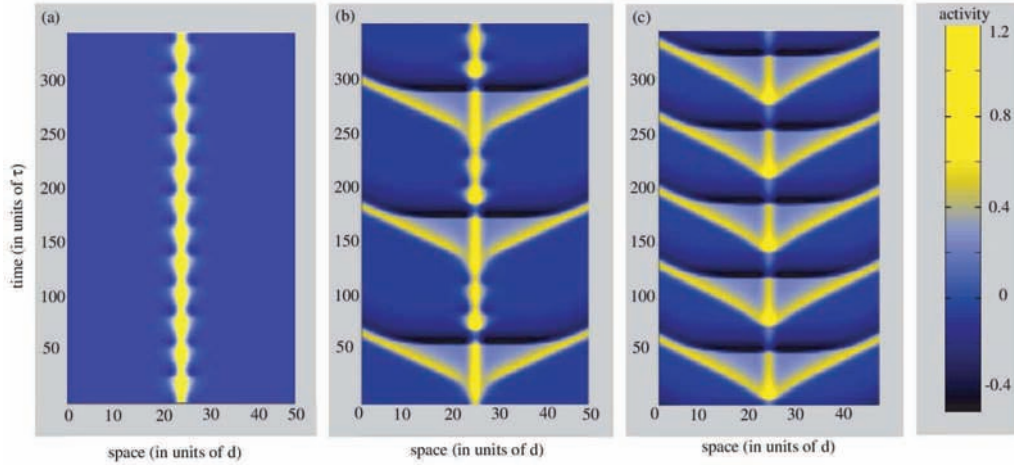


FIG. 3.6 – Illustrations graphiques, extraites de [39], présentant l'activité d'un champ de type Pinto et Ermentrout, en condition de stimulation externe non-homogène. En fonction du rapport entre le nombre de paires d'impulsions émises pendant une période d'oscillation (0 : 1 pour le graphique gauche, 1 : 4 pour le graphique du milieu, et 1 : 2 pour le graphique droit), et aussi de l'amplitude de la stimulation externe, le champ peut exhiber des fronts d'ondes oscillants (gauche), ou peut émettre des impulsions supplémentaires (graphique du milieu et graphique à droite).

de l'activité neuronale.

C'est le cas du modèle proposé par Erhlagen dans [36], où l'ajustement dynamique de ce paramètre suit la règle suivante :

$$\frac{dh(t)}{dt} = -r_{h,min}c_h(t)(h(t) - h_{min}) - r_{h,max}(1 - c_h(t))(h(t) - h_{max}) \quad (3.32)$$

où $r_{h,max}$ et $r_{h,min}$ sont des taux d'ajustement, $|h_{min}| > W_{min}$ et $|h_{max}| < W_{max}$ sont les niveaux de repos minimal respectivement maximal du champ, et $c_h(t)$ est un paramètre qui varie selon l'expression :

$$c_h(t) = (N_u(t) - N_i(t))N_u(t) \quad (3.33)$$

N_u et N_i indiquent à chaque instant du temps t , l'existence d'une activité neuronale du champ ou d'une stimulation externe :

$$N_u(t) = \sigma_H\left(\int_{x \in X} u(x, t) dx\right) \quad (3.34)$$

$$N_i(t) = \sigma_H\left(\int_{x \in X} i(x, t) dx\right) \quad (3.35)$$

D'après ces relations, le paramètre $c_h(t)$ est un paramètre indicateur, qui devient actif seulement quand l'activité du champ n'est plus « soutenue » par une source de stimulation externe. Son activation permet, dans ce contexte, de déclencher le réajustement général du niveau de repos de l'activité du champ.

Par conséquent, un modèle de champ de type Amari, qui modifie son niveau de repos selon ce critère, serait capable d'exhiber, en cas d'une forte stimulation externe, des bulles d'excitation

d'une durée finie, qui s'évanouissent graduellement une fois que la stimulation qui les a générées n'est plus active. Comme nous allons présenter dans la section 4.2.1 du chapitre suivant, ce comportement dynamique permet d'implémenter un mécanisme de mémoire facilitant la restitution de la continuité temporelle de l'activité, un aspect essentiel pour concevoir ensuite différents mécanismes cognitifs neuronaux.

Une autre approche notable a été proposée par Coombes et Owen dans [23]. Reprenant les conclusions des recherches de Hill (datant de 1936), Coombes et Owen conçoivent un modèle de champ où le niveau de repos h varie selon l'évolution de l'activité u du champ : si u se maintient élevé pour suffisamment longtemps, le seuil d'activation du champ augmente aussi, progressivement ; réciproquement, le seuil d'activation redescend graduellement, une fois que l'activité u diminue vers ce niveau de repos. Ce comportement est traduit par Coombes et Owen dans l'équation suivante :

$$\frac{dh(x,t)}{dt} = -(h(x,t) - h_0) + kg(u(x,t) - \theta) \quad (3.36)$$

où h_0 est le seuil minimal du niveau de repos, g modélise l'adaptation de h selon le niveau courant d'excitation dans le champ (typiquement par une fonction sigmoïde), et k est un facteur qui module cette adaptation. Remarquons que, par comparaison au modèle d'ajustement dynamique proposé par Erlhagen, cette deuxième approche considère que le niveau de repos de chaque unité de champ évolue indépendamment des autres unités du champ, alors que dans 3.32, $h(t)$ est un paramètre de l'ensemble du champ. Nous déduisons d'ici que seulement l'approche de Coombes et Owen respecte les principes de calcul cortical énoncés dans la section 2.5, notamment la localité du traitement neuronal dans la structure corticale.

Analysant le comportement dynamique du champ dans des hypothèses simplifiées (noyau des poids des connexions latérales de type chapeau mexicain et g une fonction de type Heaviside), les auteurs mettent en évidence différentes conditions nécessaires pour que le nouveau modèle de champ soit capable d'exhiber des motifs (des bulles ou des ondes de propagation) d'excitation stables.

Modélisation de la dynamique des poids synaptiques. Plasticité et paradigmes d'apprentissage

Parmi les différents processus dynamiques qui caractérisent le fonctionnement d'une structure neuronale, le phénomène du changement adaptatif des poids des connexions entre les unités neuronales (nommé généralement *l'apprentissage des poids synaptiques*) est, par excellence, ce qui confère à cette structure neuronale ses capacités d'adaptation.

Pour une structure de type champ neuronal, cet apprentissage concerne l'ajustement dynamique (et autonome) du noyau des poids des connexions latérales. Par conséquent, nous limitons notre discours dans ce paragraphe aux approches concernant exclusivement cet aspect. L'apprentissage d'autres poids synaptiques dans des architectures neuronales plus évoluées (y compris celles se basant sur la dynamique des champs neuronaux et visant à implémenter des mécanismes d'apprentissage non-supervisé, ou par renforcement) est abordé dans le chapitre 4, qui présente en détail les applications notables des modèles théoriques de champ.

Dans une série de travaux de référence du domaine [106, 83, 84], Miikkulainen, Bednar, Choe et Sirosh abordent le problème de l'ontogenèse et du développement synergique des connexions latérales entre les colonnes corticales. De leurs approches, nous résumons ici celle concernant le modèle de développement neuronal RF-LISSOM. L'architecture de ce modèle comprend une couche de stimulation extra-corticale et une couche corticale (voir la figure 3.7). Chaque unité

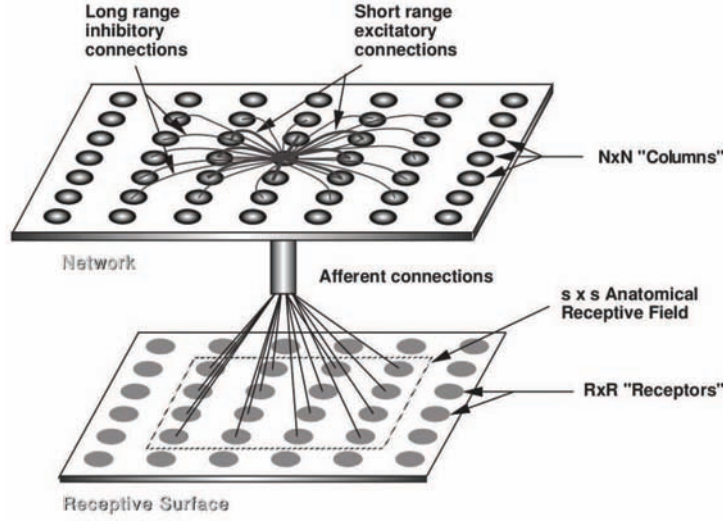


FIG. 3.7 – Illustration graphique, extraite de [83], qui décrit l'architecture générale du modèle RF-LISSOM. Le modèle est composé d'une couche extra-corticale (symbolisée ici par la couche inférieure) et d'une couche corticale, implémentée par un champ neuronal (symbolisé ici par la couche supérieure). Chaque unité du champ est connectée avec un ensemble d'unités extra-corticales (via son champ récepteur), et avec différentes unités voisines du champ (qui exercent une influence excitatrice à courte distance ou une influence inhibitrice à longue distance).

corticale x reçoit la contribution cumulée $i(x, t)$ des activités neuronales d'un ensemble compact $RF(x)$ (le champ récepteur de l'unité x) de neurones afférents (appartenant à la couche extra-corticale). Notons que, dans cette architecture, les champs récepteurs de deux unités corticales distinctes peuvent se chevaucher partiellement. Ainsi, la stimulation externe reçue par une unité quelconque x du champ est décrite par l'expression :

$$i(x, t) = \sum_{y \in RF(x)} w_{EXT}(x, y, t) s(y, t) \quad (3.37)$$

où $w_{EXT}(x, y, t)$ représente le poids synaptique de la connexion afférente entre l'unité corticale x et l'unité extra-corticale y (qui appartient au champ récepteur $RF(x)$ de l'unité x), et $s(y, t)$ représente le niveau d'activation de l'unité extra-corticale y .

Dans la couche corticale, les unités sont liées par des connexions latérales (excitatrices à courte distance et inhibitrices à longue distance), qui permettent la mise en œuvre des processus de coopération locale et de compétition globale. Stimulée par $i(x, t)$, l'activité de chaque unité corticale est définie comme suit :

$$u(x, 0) = \sigma(i(x, t)) \quad (3.38)$$

$$u(x, t) = \sigma(i(x, t) + \gamma_E \sum_{y \in X} w_E(x, y, t) u(y, t - 1) - \gamma_I \sum_{y \in X} w_I(x, y, t) u(y, t - 1)) \quad (3.39)$$

où γ_E et γ_I sont deux paramètres scalaires pondérant respectivement les contributions latérales excitatrices et inhibitrices, et σ une fonction de transfert sigmoïde qui peut ajuster ses seuils en cours du temps, afin de permettre une adaptation de la sélectivité locale du champ.

Les poids synaptiques des connexions latérales et des connexions afférentes s'ajustent suivant une règle de Hebb normalisée :

$$w(x, y, t + 1) = \frac{w(x, y, t) + \alpha u(x, t)p(y, t)}{\sum_{y \in X} w(x, y, t) + \alpha u(x, t)p(y, t)}, x \in X \quad (3.40)$$

où w représente à la fois w_{EXT} , w_E ou w_I , α représente le taux d'apprentissage (ayant des valeurs différentes pour les trois types différents de connexions), et $p(y, t)$ représente l'activité pré-synaptique, c'est-à-dire à la fois $s(y, t)$ pour les connexions afférentes, ou $u(y, t)$ pour les connexions latérales.

Pour illustrer l'efficacité de leur modèle, les auteurs abordent plusieurs problèmes concernant l'émergence des cartes corticales auto-organisées (par exemple la carte de dominance oculaire ou la carte de la sensibilité d'orientation), le codage épars (la minimisation du nombre de neurones actifs impliqués dans un certain traitement), le problème de segmentation et de regroupement visuel. Étant donné que l'ajustement des poids synaptiques est réalisé en permanence (ne cesse pas après un certain nombre de pas), le modèle vise aussi à expliquer des phénomènes de réorganisation corticale en cas de lésion cérébrale ou sensorielle.

D'après les auteurs, au début de chaque simulation, l'activité du champ forme un profil avec une seule bulle d'excitation mais, avec le temps, la réponse du champ se distribue dans plusieurs bulles locales. Par conséquent, la plupart des connexions latérales affaiblissent leurs poids synaptiques, pour arriver finalement à être complètement élaguées. Cette approche réussit donc à expliquer l'émergence des interactions de type chapeau mexicain entre les unités de champ, forme généralement utilisée dans les modèles classiques des champs neuronaux.

Malgré ces résultats, le modèle présente trois inconvénients notables. Au regard des considérations informatiques présentées dans la section 2.5, l'ajustement des poids synaptiques est effectué seulement après que toutes les unités du champ ont atteint un état d'équilibre. Deuxièmement, la normalisation de ces poids utilise une procédure de calcul global. Troisièmement, à chaque pas de temps, les unités du réseau remettent à zéro leur niveau d'activité. Tous ces trois aspects mettent en évidence l'existence de processus centralisés, ce qui pourrait être interprété comme une contradiction aux principes de traitement local de nature corticale, principes évoqués dans la section mentionnée avant.

D'autres travaux abordent la problématique de l'ontogenèse et du développement structural des champs neuronaux dynamiques. C'est le cas de Gläser, Joubelin et Goerick [49]. L'approche de ces auteurs consiste à intégrer dans la modélisation des activités neuronales du champ des principes d'homéostasie, réglant de manière dynamique et autonome le rapport entre les contributions excitatrices et inhibitrices des unités du champ, afin de maintenir la stabilité globale du champ à travers le temps. L'architecture de leur modèle est similaire à celle du modèle RF-LISSOM, avec la différence que la couche corticale est séparée en deux couches d'unités : excitatrices et inhibitrices. La dynamique de l'activité corticale est modélisée par un champ

neuronal de type Wilson et Cowan :

$$\begin{aligned} \tau_E \frac{du_E(x, t)}{dt} = & -u_E(x, t) + \sum_{y \in X} g(|x - y|) w_{EE}(x, y, t) \sigma(u_E(y, t)) \\ & - \sum_{y \in X} w_{EI}(x, y, t) \sigma(u_I(y, t)) \\ & + \sum_{y \in X} w_{EXT}(x, y, t) \sigma(i(y, t)) + h_E(x, t) \end{aligned} \quad (3.41)$$

$$\tau_I \frac{du_I(x, t)}{dt} = -u_I(x, t) + \sum_{y \in X} g(|x - y|) w_{IE}(x, y, t) \sigma(u_E(y, t)) + h_I \quad (3.42)$$

où σ est une fonction de transfert sigmoïde, et g est une fonction de distance qui décroît de manière exponentielle, et qui module l'efficacité des connexions latérales. Cependant, les poids de ces connexions ne sont pas imposés a priori, et s'ajustent en cours du temps selon une règle d'apprentissage de type Oja :

$$w(x, y, t + 1) = w(x, y, t) + p(x, t)u(y, t) - w(x, y, t)p^2(x, t) \quad (3.43)$$

où w et p ont la même signification que dans l'équation 3.40.

Les auteurs affirment que cette modélisation du processus d'apprentissage ne suffit pas pour maintenir en permanence l'équilibre dynamique entre les contributions excitatrices et inhibitrices du champ. Par conséquent, Gläser et al. proposent une modélisation plus détaillée de ce processus, basée sur un modèle de plasticité synaptique homéostasique. Inspirés par les conclusions des travaux expérimentaux comme [101, 46], les auteurs introduisent un nouveau paramètre dynamique, BDNF (de l'anglais *brain-derived neurotrophic factor*), qui intervient dans la régulation de l'activité de certaines cellules neuronales :

$$BDNF_K(x, t) = 1 + \beta_H \frac{\bar{a}(x, t - 1) - \hat{a}}{\hat{a}} \quad (3.44)$$

où β_H est le taux d'apprentissage homéostasique, $a_K(x, t)$ représente le taux de décharge (*firing rate*) instantané d'une activité neuronale a quelconque, \bar{a} représente sa moyenne temporelle, \hat{a} représente un taux de décharge visé par l'état homéostasique, et $K \in EXT, E, I$ indique la provenance de l'activité a (c'est-à-dire d'une couche extra-corticale, d'une couche corticale excitatrice ou d'une couche corticale inhibitrice).

Se basant sur le fait qu'un niveau élevé de BDNF induit une réduction des poids synaptiques des connexions excitatrices, et un niveau bas de BDNF induit une augmentation des poids synaptiques des connexions inhibitrices et vice-versa [46], Gläser et al. proposent de modifier la règle d'apprentissage définie par l'équation 3.43 comme suit :

$$w_{EXT}^*(x, y, t) = \frac{w_{EXT}^*(x, y, t - 1) + \alpha \Delta w_{EXT}(x, y, t)}{BDNF_E(x, t) BDNF_{EXT}(y, t)} \quad (3.45)$$

$$w_{EE}^*(x, y, t) = \frac{w_{EE}^*(x, y, t - 1) + \alpha \Delta w_{EE}(x, y, t)}{BDNF_E(x, t) BDNF_E(y, t)} \quad (3.46)$$

$$w_{EI, IE}^*(x, y, t) = (w_{EI, IE}^*(x, y, t - 1) + \alpha \Delta w_{EI, IE}(x, y, t)) BDNF_E(y, t) \quad (3.47)$$

où $w_{EXT, EE, EI, IE}^*$ sont les nouveaux poids des connexions synaptiques.

Comme dernier point dans leur approche de modélisation, les auteurs prennent en compte l'ajustement dynamique des potentiels de repos de l'activité neuronale des unités excitatrices :

$$h_E(x, t) = h_E(x, t - 1) + \beta_T \frac{\hat{u} - \bar{u}_E(x, t - 1)}{\hat{u}} \quad (3.48)$$

où β_T est le taux d'ajustement homéostasique du niveau de repos de l'activité neuronale.

Le modèle neuronal décrit ici est ensuite appliqué par Gläser et al. à la résolution d'une tâche de coordination sensorimotrice œil-main. L'espace d'entrée, c'est-à-dire l'ensemble des stimuli perçus par la couche extra-corticale, est défini par les coordonnées spatiales qui encodent les positions du regard et de la main (relatives au corps ou à l'œil). Les expériences montrent que les profils des champs récepteurs des unités corticales s'ajustent en cours du temps. Par conséquent, avec le temps, les entrées similaires de stimulation induisent des sorties (l'activité neuronale du champ) similaires. Cet aspect confirme une quantification vectorielle de l'espace d'entrée par le modèle neuronal proposé. Toutefois, ces résultats ne montrent pas une organisation topographique de la couche corticale (les unités voisines du champ neuronal n'ont pas nécessairement des champs récepteurs similaires). Afin de rendre possible cet aspect, il est nécessaire d'imposer certaines contraintes au niveau des connexions latérales entre les unités du champ.

Le problème de l'émergence des représentations topographiques est abordé dans [48]. Gläser et al. regardent les SOM comme étant des réseaux élastiques d'unités connectées, leur but étant de respecter des contraintes de distance, afin de minimiser les défauts de représentation topographique. Dans cette interprétation, les auteurs proposent comme contrainte le critère de minimisation de la longueur totale du circuit des connexions latérales. Sous l'aspect de la modélisation, ils proposent un mécanisme local qui modélise les interactions entre les unités corticales par des interactions similaires aux ressorts physiques. Ainsi, chaque unité x du champ ressent deux types d'influences (forces élastiques), une force attractive $F(x, y)^+$ et une force répulsive $F(x, y)^-$ comme décrit par les équations :

$$F^+(x, y) \propto w(x, y, t) |x - y| \quad (3.49)$$

$$F^-(x, y) \propto |x - y|^{-1} \quad (3.50)$$

Par conséquent, les distances entre les unités du champ s'ajustent pour maintenir l'équilibre dynamique entre ces interactions :

$$\Delta x_E \propto \sum_{y \in X} (F_{EE}^+(x, y) + F_{IE}^+(x, y) + F_{EI}^+(x, y) + F_{EE}^-(x, y)) \quad (3.51)$$

$$\Delta x_I \propto \sum_{y \in X} (F_{IE}^+(x, y) + F_{EI}^+(x, y) + F_{II}^-(x, y)) \quad (3.52)$$

Suite à cette modélisation, les relations de voisinage entre les unités du champ changent au cours du temps, ce qui se traduit par un processus de réorganisation topographique.

Les résultats de Gläser et al. montrent comment la quantification vectorielle et la représentation topographique, obtenues grâce aux traitements corticaux d'information, sont deux propriétés distinctes et (relativement) indépendantes, fait mis en évidence également par des études expérimentales récentes [59].

Nous notons que les méthodes de modélisation présentées par Gläser et al. permettent de minimiser le nombre initial de paramètres du modèle nécessaires à fixer (étant donné l'ajustement automatiquement des autres paramètres). Les auteurs soulignent également que, par rapport au modèle RF-LISSOM, les poids synaptiques sont ajustés à chaque pas de temps, alors que

dans l'approche de Miikkulainen et al., l'ajustement se produit seulement après la stabilisation globale de l'activité du champ. De ce point de vue, le modèle proposé par Gläser et al. respecte pleinement les principes de calcul cortical énoncés dans la section 2.5. Cependant, la complexité de ce modèle peut nous inciter à chercher des alternatives de modélisation pour des processus spécifiques de traitement cortical, modèles qui seraient plus appropriés pour une implémentation informatique à large échelle.

3.3 Méthodes d'analyse dynamique des modèles de champ

Si les différentes propositions de formalisation de l'activité corticale décrivent souvent intuitivement certains processus neuronaux concernant la dynamique corticale au niveau macroscopique, le caractère non-linéaire des équations intégral-différentielles de ces formalismes rend souvent impossible la résolution exacte de ces équations et, par conséquent, la description exhaustive de leurs régimes dynamiques.

Cette difficulté a conduit nombreux auteurs à concevoir différents outils analytiques pour étudier le comportement dynamique des modèles théoriques proposés. Ces études visent généralement à trouver les conditions nécessaires ou suffisantes pour l'existence et la stabilité des solutions des équations du champ.

Toutefois, étant donnée la complexité de ce vaste sujet, et le rôle plutôt périphérique qu'il joue dans nos recherches, nous nous limitons ici à énumérer seulement quelques aspects qualitatifs concernant cette littérature. Pour une synthèse de ces travaux, nous recommandons au lecteur de s'adresser à l'excellente thèse de Venkov [121] (voir notamment le chapitre 7), qui examine en détail la dynamique de plusieurs modèles de champs neuronaux, réalisant également une comparaison étendue entre les différentes méthodes d'étude analytique de ces modèles théoriques.

Une des méthodes les plus souvent employées pour établir la stabilité d'une solution de champ est celle de l'analyse des perturbations. En résumé, une solution $u(x, t)$ est considérée stable si en la perturbant avec une petite quantité $\psi(x, t)$, elle converge toujours vers sa forme stationnaire $\bar{u}(x)$:

$$u(x, t) = \bar{u}(x) + \phi(x, t) \quad (3.53)$$

Amari, dans un des premiers travaux dédiés à ce sujet [4], utilise cette méthode d'analyse pour examiner la dynamique des *régions de frontière* qui caractérisent une solution $u(x, t)$. Ainsi, la condition de stabilité pour une solution de type bulle d'excitation $\{u(x, t) | u(x, t) > 0 \text{ pour } \forall x \in (x_1, x_2), \text{ et } u(x, t) \leq 0 \text{ pour } \forall x \in X \setminus (x_1, x_2)\}$ revient à imposer :

$$\frac{\partial u(x, t)}{\partial x}(x_i(t), t) \frac{dx_i}{dt} + \frac{\partial u(x, t)}{\partial t}(x_i(t), t) = 0, \quad i = 1, 2.$$

Toutefois, si cette méthode est appropriée pour analyser la stabilité des solutions caractérisées par des régions contiguës stationnaires (comme celles définies par une unique bulle d'excitation), elle ne peut pas être utilisée pour analyser d'autres types de solutions plus complexes (par exemple avec un comportement oscillatoire), ou encore pour les solutions non-homogènes (dépendant de l'entrée de stimulation externe $i(x, t)$).

Une alternative à cette approche, très souvent utilisée dans la littérature, considère la dérivée de la condition 3.53 pour la variante linéarisée de la solution homogène $u(x, t)$ du champ (obtenue notamment en linéarisant les fonctions sigmoïdes intervenant dans les équations). Ce système d'équations est ensuite converti dans un système d'équations différentielles ordinaires, grâce à un développement en série Fourier de la fonction ϕ :

$$\hat{\phi}(k, t) = \int_{-\infty}^{\infty} \phi(x, t) e^{ikx} dx \quad (3.54)$$

Avec ces notations, le nouveau système d'équations peut s'exprimer sous une forme algébrique de type $\partial_t Z = \mathbf{A} \cdot Z$, où $Z = [\hat{\phi}]^T$. Ainsi, les invariants $\det(A)$ et $\text{tr}(A)$ de la matrice \mathbf{A} peuvent être calculés, ce qui permet au final d'obtenir les valeurs propres λ_{\pm} du nouveau système linéaire. En fonction de ces valeurs, plusieurs types d'instabilité dynamique peuvent être mis en évidence (Turing, Hopf, ou de Turing-Hopf).

Une troisième alternative que nous ne faisons qu'évoquer ici est proposée par Coombes dans [22], et se base sur la fonction d'Evans. Bien que cette dernière approche se prouve élégante pour étudier la stabilité d'ondes non-linéaires dans un domaine sans frontières, Venkov montre dans [121] (chapitre 7) que l'approche d'Amari et celle d'Evans garantissent l'obtention des mêmes résultats qualitatifs pour les solutions de type bulles stationnaires d'activité. Toutefois, la méthode développée par Coombes facilite l'étude d'autres modèles théoriques de champ (modélisant les délais axonaux ou synaptiques) dans le cadre unifié de la même plate-forme analytique.

Grâce à ces études, nous connaissons aujourd'hui certains intervalles pour les valeurs des paramètres qui garantissent certains régimes dynamiques. Malgré une contribution fortement importante vis-à-vis de l'avancement du domaine des champs neuronaux dynamiques, cette analyse théorique a été souvent réalisée dans des conditions assez restrictives, en absence de stimulations externes, ou dans des régimes qui ne sont pas suffisamment pertinents ou utiles du point de vue biologique ou robotique. La difficulté d'analyser de manière exhaustive les comportements dynamiques de ces modèles a mené les chercheurs à se consacrer de plus en plus à des études théoriques approfondies, afin de trouver certaines conditions nécessaires ou suffisantes pour caractériser divers aspects dynamiques. Nous considérons néanmoins que cette réorientation vers un « problème technique » a détourné en quelque sorte l'attention de l'intérêt pratique à utiliser ces modèles dans des applications de simulation des processus corticaux, pourtant une des motivations principales initiales de la conception de tels modèles.

3.4 Méthodes de simulation des modèles théoriques

Si le but initial de la formalisation des activités corticales était l'étude analytique, avec des outils mathématiques, du fonctionnement cortical, les recherches récentes s'orientent de plus en plus vers l'utilisation de ce paradigme dynamique comme outil pour le traitement de l'information. Afin de répondre aux défis de ce nouveau but, les chercheurs ont dû développer des méthodes de calcul robustes et efficaces, qui soient capables de simuler les formalismes mathématiques décrivant les modèles de champs proposés. Ainsi, simuler les équations d'un champ revient à discrétiser un système continu d'équations intégral-différentielles. Toutefois, notons que le système dynamique modélisé par ces modèles, c'est-à-dire le cortex, est un système discret en soi : les neurones (ou les colonnes corticales) sont des entités discrètes dans l'espace, et les impulsions électriques qu'ils génèrent sont des événements discrets dans le temps. Si la formalisation mathématique en domaine continu facilite l'étude analytique des régimes dynamiques d'un modèle neuronal, elle doit être « reconvertie » en une expression discrète, afin de pouvoir simuler les comportements dynamiques du modèle.

Ici nous résumons les principales méthodes de simulation utilisées dans la littérature, mettant aussi l'accent sur leurs avantages ou leurs limites.

Une première (et la plus souvent utilisée) approche de simulation des équations de champ fait appel à la méthode de discrétisation d'Euler. En résumé, celle-ci affirme que nous pouvons obtenir une approximation de la solution $u(t)$ de l'équation $\frac{\partial u(t)}{\partial t} = f(t, u(t))$ si nous utilisons la

formule discrète récursive :

$$u(t+1) = u(t) + \Delta t f(t, u(t)) \quad (3.55)$$

où $u(0) = u_0$ représente une condition initiale du système et Δt le pas de discrétisation.

Si cette méthode a en général un faible coût de calcul, elle peut toutefois devenir inefficace si le pas de discrétisation Δt est trop petit, car trop de pas sont nécessaires pour atteindre l'instant où u devient stationnaire. $u(t)$ peut également devenir inexact si, au contraire, ce pas est trop grand. Pour améliorer ces aspects, une généralisation de cette méthode est parfois utilisée pour assurer une meilleure précision pour les résultats de simulation. Celle-ci, appelée la méthode de Runge-Kutta, utilise plusieurs points dans l'intervalle $[t_n, t_{n+1}]$ (typiquement 4 points) pour estimer la valeur approximative de $u(t)$. Toutefois, cette méthode est plus coûteuse du point de vue calculatoire.

Certains outils logiciels implémentant ces méthodes, comme par exemple XPPAUT [37] (développé par B. Ermentrout), ou GENESIS [11] (développé au laboratoire de J. Bower), offrent la possibilité de simulation des équations de champ non-linéaires, donnant aussi la possibilité de créer et de visualiser des diagrammes des trajectoires dynamiques de l'univers des phases du modèle, des bifurcations de Hopf, des courbes null-cline etc.

Une autre méthode pour simuler numériquement la solution d'une équation de champ utilise les réseaux de convolutions. Il est immédiat d'observer que par exemple, la forme discrète de l'équation du modèle d'Amari 3.15, peut être écrite ainsi :

$$\Delta u(x, t) = -u(x, t-1) + (u \otimes w)(x, t-1) + i(x, t) + h(x, t) \quad (3.56)$$

où \otimes est l'opérateur de convolution spatiale, comme défini par l'équation 3.14. Ainsi, l'activité du champ à l'instant t , $\{u(x, t)\}_{x \in X}$ peut être exprimée comme la somme de l'activité de plusieurs champs : le champ de l'activité neuronale au pas précédent, $\{u(x, t-1)\}_{x \in X}$, le champ résultant de la convolution entre $\{u(x, t-1)\}_{x \in X}$ et le noyau des poids des connexions latérales $\{w(|x-y|)\}_{y \in X}$, le champ de la stimulation externe courante $\{i(x, t)\}_{x \in X}$ et le champ représentant le niveau de repos de l'activité neuronale $\{h(x, t-1)\}_{x \in X}$. Si cette méthode permet un calcul numérique assez rapide de l'activité neuronale u d'un champ, elle ne peut pas être utilisée pour simuler des équations de champ couplées. Dans ce cas, pour calculer (au même pas de temps) les valeurs de deux variables (exprimées par des équations de champ couplées), il faudrait établir une priorité de calcul pour ces valeurs. Toutefois, le fait d'imposer cette priorité de calcul signifierait d'imposer une synchronisation de l'activité des deux variables. Comme le montre aussi l'analyse de l'automate cellulaire du *Jeu de la Vie* de Conway [47, 88], une procédure d'évaluation synchrone peut engendrer des comportements oscillants incorrectes, ce qui affecterait complètement le régime dynamique de la solution continue du champ. Par conséquent, cette méthode est appropriée exclusivement pour exprimer l'activité d'un champ d'une manière *feedforward*.

En outre, nous notons que la modélisation numérique des traitements de type multimodal, ou des traitements de champs ayant différentes échelles de temps, peut entraîner également des comportements qui ne correspondent pas aux résultats établis par les études analytiques des équations continues. Cet aspect, évoqué aussi par Coombes dans [22] ou Hutt et al. dans [67], montre les limites de ces méthodes de discrétisation.

A cause des raisons énumérées ici, nous avons choisi de simuler numériquement les modèles de champs (dans nos analyses empiriques) par un schéma de mise à jour asynchrone, employée à chaque pas de simulation. Ceci implique le fait que l'activité de chaque unité du champ est mise à jour utilisant la forme discrète de l'expression qui la décrit, mais l'ordre de mise à jour

pour les unités du champ est aléatoire, pour garantir une évaluation asynchrone de l'ensemble de ces unités.

A cause des raisons énumérées ici, nous avons choisi de simuler numériquement les modèles de champs (dans nos analyses empiriques) par un schéma de mise à jour asynchrone, employée à chaque pas de simulation. Ainsi, les équations du modèle de champ sont discrétisées selon la méthode d'Euler décrite par l'équation 3.55, considérant un pas de discrétisation temporelle $\Delta t = 1$. En outre, afin de contourner les effets indésirables des évaluations synchrones et garantir une évaluation asynchrone pour l'ensemble des unités du champ, l'ordre de mise à jour pour ces unités est aléatoire. De ce point de vue, le schéma d'évaluation que nous employons dans nos simulations est similaire à la règle d'activation stochastique utilisée par le modèle de réseau récurrent de Hopfield (revoir la section 2.2.4).

3.5 Discussion

Dans ce chapitre, nous avons examiné en détail les caractéristiques structurales et les propriétés dynamiques, ainsi que les approches générales d'étude pour plusieurs modèles de champs neuronaux dynamiques.

Issu d'un besoin pratique de modélisation fonctionnelle de la structure corticale, le paradigme des champs neuronaux dynamiques utilise une approche de modélisation connexionniste, computationnelle et dynamique pour modéliser l'activité électrique au niveau mésoscopique sur la surface corticale. Tous les formalismes théoriques de ce paradigme reposent sur le principe général conformément auquel la dynamique des activités corticales peut s'exprimer par le résultat des interactions entre différentes entités fonctionnelles élémentaires (neurones ou ensembles de neurones). Ce principe, compatible avec les principes de calcul neuronal que nous avons présentés dans la section 2.5, confère aux modèles de champs neuronaux un avantage supplémentaire dans l'implantation logicielle des architectures de calcul à large échelle qui soient discrètes, parallèles et distribuées.

Grâce à ses caractéristiques structurelles, le modèle de champs neuronaux présente quelques propriétés fonctionnelles qui le distinguent d'autres formalismes neuronaux. Premièrement, ce paradigme implémente un modèle simple et général, local et distribué de compétition entre plusieurs sources de stimulation. De plus, son activité est souvent stable et robuste par rapport aux sources de bruit, ce qui peut conduire à considérer la dynamique des champs neuronaux comme le résultat émergent d'un processus de construction de cohérence. Enfin, l'activité de chaque unité du champ maintient une trace temporelle du processus de compétition, elle s'ajustant graduellement en fonction d'elle-même, des sources externes de stimulation, mais aussi des activités d'autres unités du champ. Cette caractéristique permet au champ de générer un comportement dynamique fonctionnant à une échelle de temps différente de l'échelle dynamique des signaux électriques qui le stimule, ce qui peut être vu comme une hiérarchisation temporelle du traitement de l'information.

Bien qu'au niveau élémentaire les activités des unités du champ ont une expression assez simple, la nature (excitatrice ou inhibitrice) et l'ampleur de leurs interactions (modélisée par les noyaux des poids de connexions latérales) peuvent engendrer au niveau global du champ une grande variété de comportements dynamiques. Afin d'étudier certains aspects (comme les conditions d'existence et de stabilité) de ces phénomènes, plusieurs auteurs ont proposé différentes méthodes et outils analytiques. Toutefois, comme nous l'avons exprimé dans la section 3.3, le caractère non-linéaire de ce paradigme rend difficile ou incomplète l'étude analytique pour tout le spectre dynamique de ces comportements exhibés par les modèles de champs neuronaux.

Malgré ces difficultés, ou les hypothèses simplificatrices employées par les modèles de champs neuronaux dynamiques, ce paradigme s'est imposé comme une alternative viable aux autres approches (symboliques ou connexionnistes) pour modéliser différents aspects de la dynamique de l'activité corticale. En outre, ce paradigme offre une base théorique formelle pour la conception et le développement des modèles de calcul neuronal, capables d'implémenter la vision de calcul comme effet émergent d'un système dynamique, que nous avons décrite dans la section 2.3.2. Pour étayer ce point de vue, dans le chapitre suivant, nous examinons en détail comment certains comportements des champs neuronaux dynamiques ont pu être mis en évidence expérimentalement, mais aussi comment les propriétés dynamiques des champs neuronaux peuvent être exploitées dans un contexte pratique de modélisation des processus cognitifs et des systèmes situés. Cependant, comme nous l'avons déjà fait dans ce chapitre, et nous allons le faire également dans le chapitre suivant, l'usage appliqué des modèles des champs neuronaux fait surgir parfois un certain nombre de problèmes pratiques concernant leur comportement dynamique. Pour traiter une partie de cet ensemble de problèmes, nous proposons dans le chapitre 5 une approche alternative d'étude de ce paradigme de calcul neuronal.

Chapitre 4

Applications des champs neuronaux dynamiques

Dans le chapitre 2, nous avons mis en évidence le rôle du traitement cortical dans une multitude de processus neuronaux, et nous avons ensuite énuméré différentes approches de modélisation neuronale, inspirées par les caractéristiques anatomiques et fonctionnelles de la structure corticale. Inspirés nous aussi par ces résultats expérimentaux et théoriques, nous avons proposé dans la section 2.6.1 une architecture générique d’auto-organisation, qui vise à modéliser le traitement cortical de l’information, lié à l’émergence des cartes corticales d’auto-organisation.

Afin d’implémenter cette architecture, nous avons promu dans le paragraphe 2.6.1 l’usage des champs neuronaux dynamiques, un paradigme de calcul neuronal conçu pour modéliser la compétition neuronale au niveau de la structure corticale. Dans ce contexte, nous avons décrit le principe fonctionnel selon lequel un champ neuronal dynamique peut être utilisé pour implémenter le module de sélection de l’architecture d’auto-organisation introduite avant.

Pour fonder les bases théoriques de notre démarche, nous avons présenté dans le chapitre précédent les fondements de modélisation des champs neuronaux dynamiques, et nous avons ensuite présenté plusieurs alternatives de formalisation mathématique de ce modèle de calcul neuronal. Si ces formalismes montrent une variété des approches de modélisation de l’activité corticale, l’étude phénoménologique de leur dynamique est un sujet amplement exploré par la communauté scientifique. Afin de mieux comprendre les caractéristiques et les capacités de ces formalismes, il est nécessaire de placer ces différents modèles de champs neuronaux dans un contexte fonctionnel pratique. Par conséquent, dans ce chapitre nous présentons plusieurs résultats de référence qui concernent l’utilisation des champs neuronaux dynamiques dans ce contexte appliqué, résultats qui ont constitué pour nous également des sources d’inspiration pour nos propres approches de modélisation.

Les démarches d’utilisation pratique des champs neuronaux dynamiques, liées à la conception d’applications bio-inspirées, abordent généralement une des trois problématiques de modélisation suivantes : (a) la modélisation au niveau neurophysiologique, qui vise à corréler les données enregistrées expérimentalement avec celles obtenues par la simulation des champs neuronaux ; (b) la modélisation des processus cognitifs, qui vise à proposer des architectures neuronales capables de modéliser les mécanismes sous-jacents aux processus cognitifs ; (c) la modélisation des systèmes situés (et en particulier des plates-formes robotiques), qui vise à doter un système physique d’un comportement autonome et adaptatif, similaire au comportement des systèmes biologiques évolués. Dans ce chapitre, nous analysons au fur et à mesure ces approches de modélisation, indiquant aussi la portée de leur influence vis-à-vis de notre démarche de modélisation.

4.1 Modélisation de l'activité corticale au niveau physiologique

Les approches décrites dans les paragraphes suivants visent à corrélérer différents comportements dynamiques d'un champ neuronal (bulles d'activité, ondes de propagation, instabilités de Turing) avec différents phénomènes neuro-physiologiques mis en évidence par les résultats expérimentaux.

4.1.1 Modélisation de l'activité du cortex visuel primaire, induite par une stimulation visuelle simple

Dans [81], Meyer examine plusieurs phénomènes neurophysiologiques observés sur la surface corticale, et enregistrés utilisant la technologie VSDI (imagerie cérébrale par teinture voltosensible). Offrant une haute résolution spatiale et temporelle, la technique VSDI permet d'enregistrer l'activité électrique de larges masses de neurones, ce qui rend possible la comparaison entre des données expérimentales caractérisant la dynamique mésoscopique de l'activité neuronale, et des données résultant d'une modélisation théorique de ces phénomènes.

Afin d'expliquer la dynamique de certaines activités corticales, Meyer considère une modélisation des réponses du cortex visuel primaire avec des champs neuronaux dynamiques. Dans ses travaux, Meyer examine la réponse spatio-temporelle exhibée par l'aire V1 du cortex d'un chat (voir la figure 4.1 (B), colonne droite), induite par la présentation de différents stimuli visuels (voir la figure 4.1 (B), colonne gauche).

Meyer propose une architecture neuronale de type feedforward, qui modélise les étapes de traitement de l'information visuelle, à partir de la stimulation rétinienne jusqu'à l'activité du cortex visuel primaire (voir la figure 4.1 (A)). Remarquons que la structure générale de cette architecture est aussi intégrée dans l'architecture neuronale d'auto-organisation que nous avons proposée dans la section 2.6, mais dans notre modélisation, l'activité de la couche corticale agit comme une contribution de feedback vis-à-vis de la couche d'entrée (c'est-à-dire la couche du module de détection).

Dans le modèle de Meyer, la réponse de la couche corticale est modélisée en utilisant une variante normalisée du modèle de champ neuronal de Wilson et Cowan (revoir la section 3.2.1). Notons que ce modèle de champ neuronal prend en compte la séparation des contributions excitatrices et inhibitrices de l'activité corticale. Ceci permet de quantifier la réponse globale de la couche corticale comme étant une combinaison linéaire des activités excitatrices et inhibitrices, et d'ajuster plus facilement les paramètres libres du modèle.

Comparant les résultats des données expérimentales avec les résultats obtenus par la simulation du modèle, Meyer met en évidence une haute corrélation entre les deux (voir les colonnes centrale et droite de la figure 4.1 (B)). Ces résultats permettent ainsi de considérer la pertinence de modélisation des réponses corticales par les modèles des champs neuronaux dynamiques, et de concevoir un mécanisme neuronal qui puisse expliquer l'émergence des phénomènes neurophysiologiques liés au traitement de l'information visuelle.

4.1.2 Modélisation de la préférence d'orientation visuelle

Dans [8], Ben-Yishai et al. modélisent la préférence d'orientation des neurones du cortex visuel primaire par une approche de type champ moyen. L'activité neuronale u d'une unité du champ ayant la préférence d'orientation θ , est modélisée par l'équation 4.1. Comme nous pouvons remarquer, ce modèle repose sur un traitement feedforward du flot d'information visuelle, modulé par les interactions latérales de la couche corticale. Pour un stimulus visuel d'orientation $\theta_\xi = 0$,

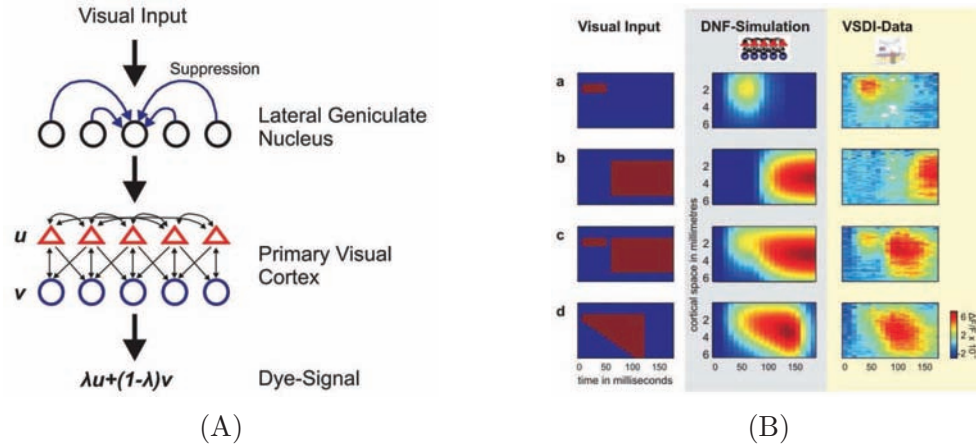


FIG. 4.1 – (A) Architecture neuronale proposée par Meyer, modélisant l'enchaînement des traitements neuronaux générés par une stimulation visuelle simple. Le signal visuel perçu par la rétine active le corps géniculé latéral, qui stimule ensuite le tissu cortical de l'aire Brodmann 17. L'activité de V1 est modélisée par un champ neuronal dynamique de type Wilson et Cowan. L'activité électrique corticale est quantifiée par une expression linéaire des contributions excitatrices et inhibitrices du champ. (B) Résultats expérimentaux (colonne centrale) et de simulation (colonne droite), obtenus pour différents stimuli visuels (colonne gauche). Ces résultats représentent des valeurs moyennes pour une série de 16 observations. Illustration graphique, extraite de [81].

la stimulation d'une unité du champ ayant la préférence θ est modélisée par la fonction h , définie dans l'équation 4.2. Cette fonction décrit la fonction de reconnaissance (tuning curve) de la préférence d'orientation d'une unité, et prend en compte l'amplitude A , le contraste lumineux c du stimulus visuel, ainsi qu'un facteur ϵ , qui module la distribution des stimuli perçus par chaque unité du champ (par exemple, pour $\epsilon = 0$, toutes les unités du champ reçoivent la même entrée).

$$\tau \frac{du(\theta, t)}{dt} = -u(\theta, t) + \sigma \left(h(\theta) + \int_{-\pi/2}^{\pi/2} \frac{d\theta'}{\pi} (-\lambda_0 + \lambda_1 \cos(2(\theta - \theta'))) u(\theta', t) \right) \quad (4.1)$$

$$h(\theta) = Ac(1 - \epsilon + \epsilon \cos(2\theta)) \quad (4.2)$$

Selon les auteurs, l'activité du champ ainsi modélisée est corrélée avec les réponses de préférence d'orientation visuelle qui ont pu être observées expérimentalement. Malgré ces résultats, nous observons que ce modèle réalise exclusivement un traitement de type feedforward, n'intègre aucun mécanisme adaptatif (d'apprentissage), et ne réussit pas à expliquer l'émergence des cartes d'auto-organisation de préférences d'orientation visuelle.

4.1.3 Modélisation des phénomènes neurophysiologiques anormaux

Si la dynamique des bulles stables d'excitation locale dans un champ neuronal peut expliquer plusieurs phénomènes observés expérimentalement au niveau physiologique, l'émergence et la

propagation spatio-temporelle d'ondes d'activité neuronale pourraient expliquer des phénomènes dynamiques qui caractérisent certains cas de fonctionnement cérébral anormal ou pathologique.

Ainsi, dans [12], Bressloff et al. utilisent une variante généralisée des équations de Wilson et Cowan (revoir les équations 3.12 et 3.13) pour modéliser l'activité neuronale du cortex visuel primaire et étudier l'apparition des phénomènes d'hallucination visuelle. Dans leur approche, les auteurs prennent en compte plusieurs aspects de modélisation. Premièrement, ils modélisent la projection rétinotopique déformée de V1, caractérisée par le fait que la fovéa est représentée par une densité de neurones corticaux plus grande que les autres zones rétinienne. Ensuite, les auteurs considèrent la distinction entre les connexions latérales liant des unités appartenant à la même hyper-colonne (c'est-à-dire la région corticale caractérisée par l'ensemble d'unités dont les préférences d'orientation couvrent tout le spectre de la sensibilité d'orientation), et les connexions latérales à longue distance, liant les unités avec la même préférence d'orientation appartenant à deux hyper-colonnes différentes.

Considérant ce modèle, Bressloff et al. montrent qu'en fonction de certaines conditions de paramétrage, le champ peut avoir deux modes dynamiques de fonctionnement : un cohérent avec les observations de Hubel et Wiesel, et un autre instable, dans lequel des ondes d'excitation émergent et se propagent dans la couche corticale. Étudiant plus en détail le comportement du modèle dans des contextes de stimulation visuelle avec des formes géométriques utilisées dans les expériences d'hallucination, les auteurs formulent des explications possibles pour l'apparition et la persistance des expériences visuelles réelles.

Intéressés par les cas d'anomalie de fonctionnement cortical, Golomb et Amitai étudient dans [50] l'influence de la vitesse de propagation de l'activité neuronale dans la dynamique corticale globale, et comparent des données expérimentales avec les données simulées d'un modèle cortical, basé également sur des équations de champ similaires à celles de Wilson et Cowan. Animés par les mêmes questions, Pinto et Ermentrout analysent plus en détail du point de vue théorique la problématique de la propagation des activités neuronales dans [92, 93], proposant le modèle de champ décrit dans la section 3.2.5. Les résultats de ces études convergent à conclure que la dynamique des ondes de propagation exhibée par leurs modèles, provoquant des rafales d'excitation, peut être mise en correspondance fidèlement avec les enregistrements électriques corticaux qui ont été observés dans les cas de crises d'épilepsie.

4.2 Modélisation des processus cognitifs élémentaires

Les paragraphes suivants présentent l'utilisation du paradigme des champ neuronaux dynamiques dans le contexte de modélisation des mécanismes cognitifs élémentaires, comme la mémoire, l'attention ou même les processus qui interviennent dans la phénoménologie de la conscience. Comme nous allons le constater, les auteurs utilisent dans leurs démarches de modélisation la dynamique des bulles d'activité des champs neuronaux à la fois comme attracteurs d'un système dynamique, comme mécanisme de compétition et de sélection neuronale, ou comme mécanisme dynamique favorisant la persistance ou permettant l'anticipation d'un certain processus ou événement neuronal.

4.2.1 Modélisation des mécanismes neuronaux de mémoire

Dans la section 2.2.4, nous avons présenté une approche de modélisation d'un mécanisme de mémoire, basée sur un modèle de réseau neuronal récurrent (le réseau de Hopfield). En résumé, l'approche de Hopfield consistait à associer chaque attracteur du système dynamique du réseau à un élément de mémoire. Dans la suite de ce paragraphe, nous présentons plusieurs approches

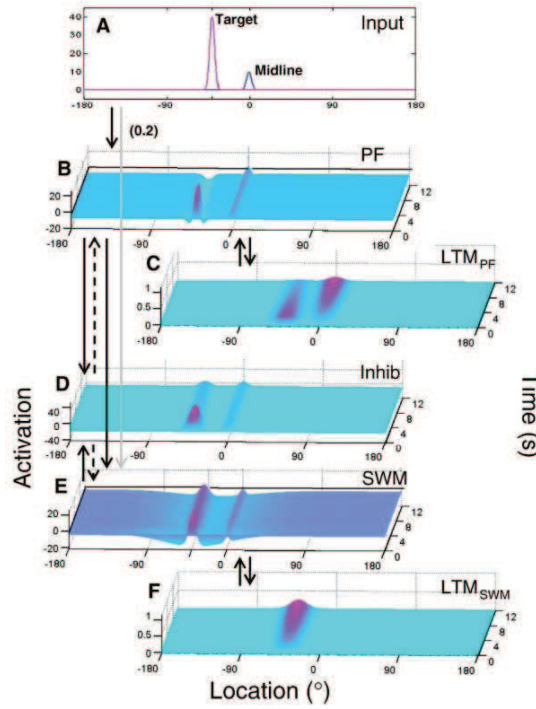


FIG. 4.2 – Architecture neuronale multi-couches proposée par Simmering et al., qui modélise un mécanisme de mémoire spatiale. Suivre le texte pour les détails. Illustration graphique, extraite de [105].

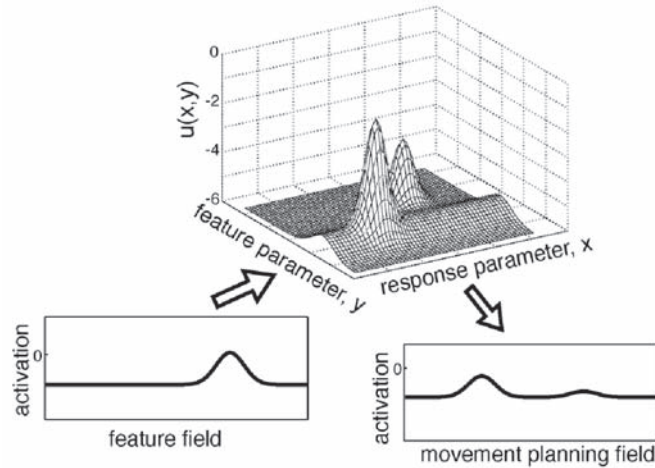


FIG. 4.3 – Implantation proposée par Wilimzig et al., pour un modèle de mémoire associative de type stimulus-réponse, qui associe à une caractéristique perceptive une action motrice possible. En cours du temps, une décision motrice peut être favorisée par l'augmentation progressive d'une *préforme d'activation*. Suivre le texte pour plus de détails. Illustration graphique, extraite de [124].

inspirées par le même principe, mais utilisant cette fois-ci la dynamique des champs neuronaux. Comme nous allons le constater, dans ce cas, les attracteurs du système dynamique du modèle (champ) correspondent aux bulles stables de l'activité neuronale du champ.

Dans [105] Simmering et al. proposent un modèle général de mémoire spatiale basé sur la dynamique conjointe de cinq couches de champs neuronaux (voir la figure 4.2) : un champ perceptif *PF*, un champ de type mémoire de travail *SWM* (ou mémoire à court terme), un champ inhibiteur partagé *Inhib*, et deux champs de mémoire à long terme, LTM_{PF} et LTM_{SWM} , respectivement couplés avec les deux premiers champs. La couche perceptive *PF* est stimulée par une distribution de stimuli, qui excite à son tour *SWM*, et imprime ainsi une trace temporelle. Les deux couches, *PF* et *SWM* sont interconnectées avec *Inhib*, ce qui se traduit par une augmentation en amplitude des zones de forte activité de *PF* et *SWM*, et une inhibition des autres zones de ces deux champs. Cette dynamique implémente ainsi un effet d'augmentation du contraste perceptif, ce qui aide à l'entretien de l'activité neuronale en absence du stimulus d'entrée initial. Intégrant également un processus d'apprentissage Hebbien, la dynamique des champs LTM_{PF} et LTM_{SWM} génère une activité qui augmente graduellement avec la durée de l'activité des deux couches afférentes, *PF* et *SWM*. Cette dynamique progressive influence les champs *PF* et *SWM*, déplaçant leurs zones d'activités existantes, ou en créant de nouvelles, qui correspondent à de nouveaux stimuli d'entrée.

En variant certains paramètres de ce modèle (comme par exemple l'influence des projections feedforward entre les couches neuronales, ou l'amplitude et la portée des influences excitatrices et inhibitrices), les auteurs examinent la capacité de ce modèle neuronal à mémoriser le positionnement d'un objet apparu dans la scène visuelle, ou la capacité à discriminer entre une nouvelle et une ancienne position d'un tel objet. En changeant ces paramètres, ainsi que les échelles de temps des interactions entre les différents champs neuronaux du modèle, les auteurs mettent en évidence plusieurs comportements dynamiques, similaires aux différentes phases du développement cortical. Selon [105], les résultats de ces expériences peuvent être mis en correspondance avec les résultats des tests de mémorisation, obtenus respectivement par des enfants, des adolescents ou des adultes.

Comme l'affirment les auteurs, les résultats de cette approche, basée sur la dynamique des champs neuronaux pour modéliser des mécanismes de mémoire, viennent renforcer le lien entre les neurosciences théoriques et comportementales. Toutefois, nous considérons que cette démarche est difficilement extensible pour permettre la mémorisation de plusieurs objets, ou pour concevoir une architecture avec des champs neuronaux bidimensionnels.

Dans une approche alternative de modélisation, qui utilise la variante bidimensionnelle du modèle de champ neuronal d'Amari (revoir la section 3.2.3), Wilimzig et Schöner proposent dans [124] un mécanisme de mémoire associative de type stimulus-réponse. Comme le montre la figure 4.3, chaque unité du champ 2D encode l'association entre une caractéristique perceptive et une décision d'action (motrice). Les auteurs implémentent dans la dynamique du champ le concept de *préforme* (en anglais *preshape*), qui représente une stimulation locale préférentielle du champ suite à une perception, mais préalable à la prise de décision. L'amplitude de cette stimulation augmente avec la probabilité d'un choix de décision, et facilite ainsi l'activation des réponses liées aux certains choix d'actions possibles. Si plusieurs actions sont candidates pour la même entrée perceptive, elles entrent en compétition via le mécanisme d'interaction des contributions excitatrices (locales) et inhibitrices (à longue distance) du champ neuronal. Comme les deux auteurs affirment, cette approche peut offrir une explication détaillée pour certains effets psychophysiques observés expérimentalement.

Dans ce modèle, les processus de perception, de décision et de planification interagissent en permanence, approche qui met en question la perspective traditionnelle conformément à laquelle

ceux-ci seraient indépendants, séquentiels et hiérarchisés. Cependant, notons que ce modèle de mémoire n'intègre pas, à l'instar du modèle de Simmering et al., un paradigme d'apprentissage des associations perception-action.

Comme évoqué dans la section 3.2.6, une autre alternative pour concevoir un mécanisme neuronal de mémoire, basé sur la dynamique des champs neuronaux, est formulée par Erlhagen dans [36]. Celle-ci considère un modèle de champ d'Amari, pour lequel le niveau de repos neuronal h s'ajuste dynamiquement selon l'équation 3.32. Un tel champ permet, dans un premier temps, la formation d'une forte activation du champ en présence d'une stimulation externe (par exemple une source perceptive). De plus, grâce au mécanisme de compétition neuronale, le champ peut sélectionner une seule zone d'activation, si nécessaire. Une fois l'activité du champ générée, elle devient persistante, même si l'entrée initiale cesse de stimuler le champ. Toutefois, si ce comportement dynamique est désirable pour implémenter un mécanisme de mémoire, il est nécessaire que l'activité du champ disparaisse aussi après un certain délai, laissant la possibilité de mémoriser un nouvel état perceptif. Pour résoudre ce problème, Erlhagen intègre au modèle d'Amari un mécanisme d'« oubli ». En absence d'une stimulation externe, le champ augmente graduellement son niveau global de repos neuronal, de manière proportionnelle au surplus courant d'excitation. Au cours du temps, ce phénomène entraîne la descente de l'activité du champ, et donc l'effacement de la mémoire, ce qui permet l'implantation correcte de la sémantique envisagée par un modèle de mémoire neuronale.

Si cette approche représente une source d'inspiration pour nos propres approches de modélisation (discutées dans le chapitre suivant), nous réitérons ici la remarque faite dans la section 3.2.6, concernant la procédure d'ajustement du niveau de repos neuronal du champ : cette adaptation est effectuée de manière centralisée, pour toutes les unités du champ, fait qui contrevient aux principes de modélisation neuronale que nous avons établis dans la section 2.5.

4.2.2 Modélisation des processus d'attention

Un autre intérêt des recherches basées sur la dynamique des champs neuronaux est représenté par la modélisation des mécanismes d'attention. Les approches explorant cet aspect considèrent que la compétition neuronale exhibée par cette dynamique possède les propriétés dynamiques pour implémenter le caractère sélectif du processus d'attention.

Utilisant la dynamique des bulles d'activité d'un champ neuronal de type Amari, Rougier et Vitay [99] proposent un mécanisme neuronal pour expliquer l'émergence du processus d'attention. Leurs expériences montrent la capacité du modèle à suivre des cibles statiques ou mobiles individuelles, en présence d'un bruit spatial de grande intensité, ou en présence d'autres stimuli spatiaux distracteurs. Basés sur ces résultats, Fix, Alexandre et Rougier proposent ensuite un mécanisme pour modéliser le mouvement saccadique oculaire volontaire, c'est-à-dire un processus neuronal qui vise à focaliser de manière séquentielle l'attention sur des cibles visuelles ayant différentes caractéristiques [38]. Le modèle, illustré graphiquement dans la figure 4.4, considère d'abord une étape d'extraction de certaines caractéristiques (couleurs, orientation, contraste, etc.) de la scène visuelle. Les résultats de ce traitement sont ensuite traités en parallèle par deux processus distincts : un pour l'analyse spatiale, et un pour l'analyse de ses caractéristiques. Le premier vise à sélectionner la zone d'intérêt spatial de la scène visuelle (via un mécanisme de compétition de type champ neuronal), la mémorisation de cet endroit, et l'anticipation des conséquences de la saccade suivante sur l'ensemble du traitement visuel. La deuxième voie de traitement vise à implémenter un mécanisme pour désigner le nouvel endroit le plus saillant concernant les caractéristiques de la scène visuelle. En fonction du résultat de ce traitement, une décision pour déplacer la zone de focalisation ou pour exécuter une saccade oculaire, est prise.

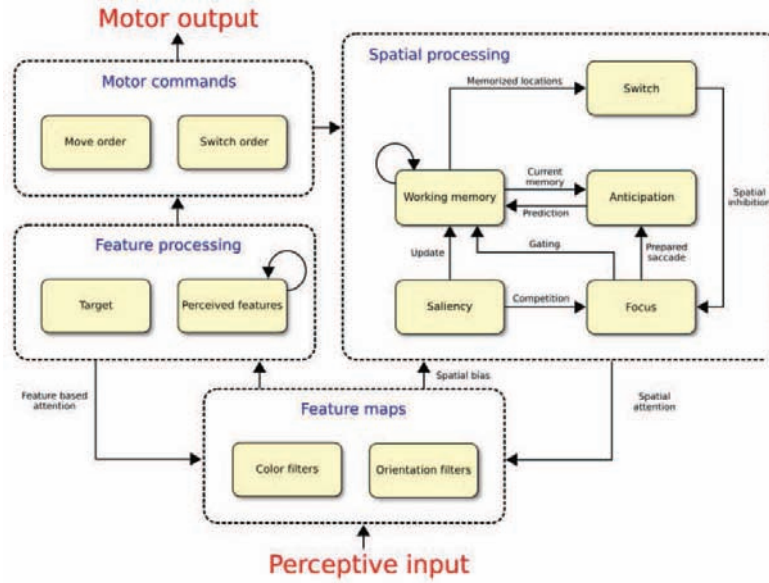


FIG. 4.4 – Architecture neuronale intégrant le traitement des saccades oculaires, proposée par Fix et al. pour modéliser un processus d’attention visuelle. Illustration graphique, extraite de [38].

L’effet de cette commande est intégré dans la première voie de traitement via une boucle de connexion récurrente.

Dans une autre approche, basée sur le modèle de Wilson et Cowan, Schneider et Erlhagen [103] proposent une modélisation des phénomènes de mouvement saccadique liés au colliculus supérieur. Les auteurs décrivent un mécanisme neuronal responsable de l’intégration et la sélection des stimuli visuels, ciblés par un sujet qui doit les suivre en permanence. La comparaison des résultats du modèle simulé avec les statistiques biologiques enregistrées expérimentalement conforte un certain nombre d’hypothèses sur le fonctionnement du colliculus supérieur. Nous remarquons également que, en dépit du fait qu’il ait été conçu pour modéliser la compétition des activités corticales, le modèle des champs neuronaux dynamiques est utilisé avec succès dans la modélisation d’autres structures cérébrales, ce qui montre que ce modèle pourrait être vu comme un paradigme général de traitement d’information, valable pour plusieurs structures anatomiques cérébrales. Notons également que le modèle des champs neuronaux dynamiques est une structure topologique, ce qui lui confère un avantage supplémentaire dans les démarches de modélisation de structures neuronales ayant une organisation anatomique bidimensionnelle.

4.2.3 Vers la modélisation des mécanismes neuronaux liés aux phénomènes de la conscience

Abordant un des défis les plus grands de la science actuelle, le problème de la conscience, Taylor [113] propose un modèle neuronal hiérarchisé capable de reproduire plusieurs aspects de la phénoménologie primaire de l’expérience consciente. Le fonctionnement de ce modèle repose sur la dynamique des champs neuronaux bidimensionnels de type Amari (plus spécifiquement, comme décrit dans la section 3.2.3). En résumé, ce modèle considère que la première étape de l’émergence d’une expérience consciente est définie par l’extraction et l’encodage des caracté-

ristiques du flot d'information perceptive. Au second palier de traitement, plusieurs sites de mémoire tampon sont couplés entre eux par des connexions inhibitrices, afin de permettre la compétition et la sélection des caractéristiques codées dans la première étape. Au dernier niveau, les modules de traitement établissent des connexions de feedback avec les modules des étages inférieurs, permettant l'émergence et le développement des processus d'attention qui modulent la sélectivité des étapes précédentes du traitement d'information. Utilisant le comportement dynamique des bulles d'activité des champs neuronaux, Taylor propose des mécanismes capables d'expliquer plusieurs phénomènes psychophysiques liés à l'expérience consciente. Pour illustrer cette approche, nous nous appuierons dans la suite sur deux exemples.

Dans un premier cas, Taylor examine le mouvement apparent d'un stimulus lumineux qui, après une brève disparition (moins de 200 *ms*), réapparaît dans la proximité de cet emplacement. L'auteur explique ce phénomène en utilisant la dynamique des bulles d'excitation. Plus spécifiquement, l'illusion optique ressentie au niveau conscient serait le résultat d'un processus de relaxation (une période transitoire) entre les deux états stables, correspondant à l'activation du champ induite par les deux stimuli visuels. Dans ce cas, l'évolution du champ se traduit par un déplacement lent (une trace d'activité neuronale) d'un endroit de stimulation à l'autre. L'auteur suggère que ce comportement dynamique peut se répercuter aux niveaux supérieurs du traitement visuel, induisant l'impression que l'événement initial de stimulation visuelle serait continu, et non pas discret, comme décrit ci-dessus.

Un mécanisme neuronal similaire serait responsable aussi de certains phénomènes manifestés quand une image est fixée sur la rétine pendant plusieurs secondes. Avec le temps, l'image perçue se dégrade, laissant place à l'apparition d'illusions optiques. Pour étayer cette hypothèse, Taylor propose un modèle de champ neuronal unidimensionnel, composé par deux couches neuronales s'inhibant réciproquement, qui encode deux caractéristiques antagonistes des objets visuels (dans son exemple deux couleurs très différentes). Une étude analytique de l'équation du champ montre que celle-ci admet au moins trois types de solutions (excitation étendue à l'infini, oscillations spatiales permanentes et bulles d'activités locales), chaque solution ayant un comportement dynamique qui peut être mis en correspondance avec la perception consciente des illusions visuelles ressenties dans le contexte des images stabilisées.

Contrairement aux études analytiques d'autres modèles de champ, qui s'intéressent souvent exclusivement à la stabilité des solutions, Taylor remarque qu'un aspect important dans un contexte de modélisation appliquée concerne l'étude de la durée de stabilisation et la durée de vie des bulles d'activité d'un champ. Analysant cette problématique, Taylor met en évidence l'influence de plusieurs facteurs liés à cet aspect : le rapport entre le taux de croissance de l'activité neuronale et le temps de saturation de l'activité du champ, la force des connexions latérales, ou la pente des fonctions de transfert sigmoïdes. Parmi d'autres travaux s'intéressant à ces aspects pratiques de modélisation d'un champ neuronal, nous mentionnons [3, 121].

Les résultats de ces travaux viennent renforcer l'opinion suivant laquelle l'exploitation des paradigmes neuronaux, comme les champs neuronaux dynamiques, peut apporter des éléments nouveaux dans la compréhension des processus cérébraux (et en particulier corticaux) décrits jusque récemment seulement au niveau phénoménologique. Notons toutefois que les résultats de Taylor concernant l'expérience consciente se limitent au traitement du flot perceptif élémentaire, et n'abordent pas des sujets essentiels comme les qualia, l'émergence du langage, la formulation des hypothèses (l'abduction), les mécanismes d'inférence logique (induction), etc.

4.3 Modélisation des traitements sensorimoteurs et des systèmes situés

Dans les sections suivantes, nous abordons le sujet de la modélisation d'inspiration corticale des traitements de nature sensorimotrice, un sujet essentiel pour le développement des systèmes autonomes et adaptatifs, capables d'explorer et d'agir dans un environnement a priori inconnu. Dans ce contexte, les champs neuronaux dynamiques sont utilisés principalement comme des mécanismes de compétition, de sélection, ou d'encodage neuronal des positions ou des commandes motrices (ou d'autres paramètres) des systèmes situés. Certains de ces systèmes intègrent également des modèles neuronaux (par exemple des modèles de mémoire) décrits précédemment, afin de développer un comportement plus complexe et mieux adapté aux tâches motrices considérées. Notons aussi que, grâce à la dynamique transitoire d'un champ neuronal, le temps de relaxation d'une activité de champ peut induire un ajustement graduel de certaines variables du système dynamique utilisant le champ. Comme nous allons le constater, cette propriété est utilisée par certains auteurs pour implémenter un déplacement moteur lisse des plates-formes robotiques conçues, ou pour implémenter des hiérarchies temporelles de traitement moteur.

4.3.1 Problématiques et approches de modélisation

Si les approches de modélisation énumérées ci-dessus visent, du point de vue théorique, à approfondir la compréhension des processus fonctionnels de la structure corticale, le point de vue pratique de ces démarches vise à intégrer ces résultats dans la conception des systèmes physiques, afin de les rendre capables de planifier, coordonner, décider et contrôler leurs actions d'une manière autonome et adaptative. Dans la suite de cette section nous analysons plusieurs de ces approches destinées à construire des systèmes situés (en particulier des systèmes robotiques).

Nous rappelons que pour les systèmes situés, l'action, c'est-à-dire l'interaction avec l'environnement extérieur, représente le facteur déterminant pour développer un comportement dynamique adaptatif. Dans l'espace physique, l'action, au sens le plus élémentaire, est motrice. Ainsi, le traitement sensorimoteur a pour objectif l'apprentissage de la corrélation entre diverses sources sensorielles (visuelles, auditives etc.) et proprioceptives (ou kinesthésiques), afin de pouvoir *percevoir et interpréter les conséquences* de ses actions, notamment les mouvements des membres et de leurs articulations. Dans le contexte le plus général, ce traitement sensorimoteur pose deux problèmes majeurs :

Le problème mécanique (PM) - déterminer le rapport bidirectionnel entre l'ensemble des actions motrices et l'espace physique, c'est-à-dire :

- le *problème cinématique direct* (PCD) : déduire la configuration future (dans l'espace 3D) des articulations du corps, suite à une action motrice ;
- le *problème cinématique inverse* (PCI) : concevoir une action motrice qui induira une configuration future particulière (dans l'espace 3D) des articulations du corps ;

Le problème d'utilité (PU) - interpréter et quantifier les conséquences des actions motrices effectuées.

Approches pour résoudre le problème mécanique (PM)

Comme présenté ci-dessus, le premier problème abordé par tout traitement sensorimoteur concerne son aspect mécanique. Cet aspect comporte en lui-même, deux problématiques fonctionnelles. Ainsi, dans un premier temps, le système robotique doit être capable de prédire les

positions de ses membres (et surtout de leurs extrémités) dans l'espace physique tridimensionnel, avant l'engendrement d'une action motrice (PCD). Deuxièmement, le robot doit être capable de concevoir une action motrice (ou une série d'actions motrices) qui puisse diriger ses membres d'une configuration posturale vers une autre, ou plus spécifiquement (pour une extrémité d'un membre), d'une position vers une autre position dans l'espace physique tridimensionnel (PCI). Remarquons ainsi que, si le problème cinématique direct est assez simple à résoudre, le problème inverse est très difficile à traiter. En effet, le plus souvent, les solutions du PCD ne peuvent pas explorer l'ensemble de toutes les actions possibles à effectuer, afin d'en proposer une comme solution du PCI, ou encore, il existe un nombre infini de possibilités pour atteindre un point donné de l'espace 3D avec l'extrémité d'un bras.

Pour résoudre PCD et PCI, deux approches différentes ont été proposées au cours du temps. La première est la solution automatique, qui formalise PCD par un calcul de matrices de translation et de rotation en coordonnées homogènes. Ainsi, selon cette approche, résoudre PCI se réduirait aussi à un calcul, celui des matrices inverses. Cette solution est employée typiquement dans la plupart des systèmes robotiques industriels. Toutefois, si un tel robot dispose de plus de 6 degrés de liberté, la matrice du modèle direct devient singulière, et par conséquent, le calcul de son inverse devient impossible. Un argument supplémentaire souligne encore plus radicalement les limites de cette approche : le comportement décrit par cette méthode de calcul matriciel est complètement « câblé », ce qui rend inutile toute stratégie d'adaptation à ce niveau de traitement d'information. Afin de trouver une alternative à cette approche, les recherches scientifiques se sont de plus en plus intéressées à trouver des variantes qui s'inspirent du traitement moteur des organismes vivants. Ces variantes s'affranchissent du recours à un modèle direct, qui serait exprimé explicitement par des équations.

Approches pour résoudre le problème d'utilité (PU)

Si l'action est un instrument grâce auquel un organisme vivant ou un système situé peut explorer son environnement (l'univers perceptible), l'interaction avec cet environnement a souvent un aspect téléologique. L'existence de tout système ouvert est contrainte par la satisfaction de plusieurs critères (imposés par exemple par les lois physiques, métaboliques, etc. et dans le monde du vivant, ces critères mènent à la formulation (implicite ou explicite) des *buts* (souvent contradictoires). Dans ce contexte, l'action représente l'outil permettant l'accomplissement de ces buts. Cependant, la complexité intrinsèque de ces critères, mais aussi l'impossibilité pratique d'explorer toutes les actions possibles pour atteindre ces buts, exige la conception d'une *stratégie* pour générer les actions, afin d'approcher le système qui les exécute de ses buts. Par conséquent, la stratégie a besoin d'*évaluer* en permanence l'état du système par rapport à ses buts. Remarquons ici que, pour un système ouvert, la conception des buts, ainsi que les stratégies conçues pour les atteindre, fondent le caractère *subjectif* d'un tel système. Trouver ce rapport essentiellement subjectif entre une action et ses conséquences définit ce que nous avons appelé précédemment le *problème d'utilité* (PU). Basés sur cette perspective, nous pouvons affirmer que, du point de vue de l'organisation de la matière, le PU est un des aspects essentiel qui différencie le monde minéral (inanimé) du monde vivant (animé).

Résoudre le PU revient alors à définir (de manière implicite ou explicite) les buts et à concevoir les stratégies d'un système ouvert donné. Toutefois, proposer une solution générale pour le PU, c'est-à-dire une stratégie générale pour satisfaire un but, quel que soit ce but, est une entreprise impossible. Pour ceci, remarquons que Entscheidungsproblem (concevoir un algorithme pour décider la valeur de vérité de n'importe quelle proposition exprimée dans la logique d'ordre un) est un cas particulier du PU, et ce problème a été prouvé insoluble par Turing [117]. Par

conséquent, le PU doit être abordé dans des contextes spécifiques.

Malgré ces difficultés inhérentes, la nature a réussi à développer au cours de son évolution différentes stratégies assez générales, pour aborder le PU. La sélection naturelle et l'apprentissage par renforcement sont deux tels exemples éloquentes. Bien qu'elles s'exercent à des échelles spatiales et temporelles très différentes, les deux stratégies reposent sur un principe commun : traiter et intégrer en permanence les influences de feedback du processus d'interaction système – environnement. Inspirée par ces aspects, la communauté scientifique a proposé au cours des années plusieurs approches pour aborder le PU, visant principalement la conception de systèmes situés.

La première méthode théorique pour résoudre le PU s'inspire de la théorie Darwinienne de la sélection naturelle, et emploie des *algorithmes évolutifs*. En résumé, une telle technique favorise le développement des traits fonctionnels qui permettent au système de satisfaire au mieux ses buts, supprimant (ou inhibant) en même temps les autres traits fonctionnels. Malgré le succès enregistré par cette approche (notamment dans le domaine robotique ou celui des systèmes multi-agent en tant que technique d'optimisation), elle requiert beaucoup de ressources de calcul, et par conséquent est peu efficace pour développer des systèmes adaptatifs, capables de réagir en temps réel. Pour cette raison, dans les paragraphes suivants, nous n'entrons pas dans les détails pratiques liés à l'implantation de cette méthode.

La deuxième méthode théorique employée pour résoudre le PU, part de l'hypothèse que tout signal moteur perturbé par certaines sources de bruit peut se corriger au cours du temps, favorisant graduellement la composante intrinsèque correcte, c'est-à-dire celle qui valorise le mieux le critère d'utilité. Cette méthode porte le nom de l'*apprentissage implicite* [110], et elle s'étaie sur un effet de type *bootstrapping*, où chaque nouvelle dextérité motrice optimise les capacités motrices déjà acquises.

La troisième et dernière méthode que nous mentionnons ici vise à minimiser les erreurs des actions effectuées par un système, intégrant dans sa dynamique une fonction de récompense. Cette fonction a le but d'évaluer la qualité des actions individuelles et des séquences d'actions exécutées par le système, récompensant le système quand ces actions induisent un rapprochement de ses buts, et le pénalisant dans le cas contraire. Selon [110], cette approche porte le nom de l'*apprentissage explicite* ou *apprentissage par renforcement*. Ainsi, cette technique vise à adapter le comportement d'un système, en corrigeant ou en optimisant sur le long terme la série de ses actions (par exemple le déplacement de ses membres), vis-à-vis des buts considérés.

Vu l'usage très répandu de cette troisième stratégie chez les espèces évoluées (celles disposant d'un cerveau), les paragraphes suivants mettent l'accent sur le développement des traitement sensorimoteurs intégrant cette stratégie comme méthode générale pour résoudre le PU.

4.3.2 Modélisation à l'aide d'un système dynamique dual

Une première approche de modélisation sensorimotrice que nous présentons ici, et qui vise la résolution du PM, appartient à Bergener et ses collaborateurs. Dans [10], ces auteurs abordent la problématique de conception d'une plate-forme robotique, capable de se déplacer dans un environnement a priori inconnu et d'éviter les différents obstacles fixes ou mobiles. Le robot a pour but de localiser, reconnaître et saisir certains objets, et même d'interagir avec des sujets humains.

Pour résoudre le PM, les auteurs proposent une approche basée sur l'implantation d'un système dynamique dual. Chaque changement de la configuration des articulations des membres du robot est « traduit » dans la dynamique du système dual. Cette approche facilite la résolution du PCD. Cette association action – dynamique ne représente pas la formalisation symbolique du

fonctionnement moteur du robot, mais elle représente une réverbération permanente de l'action du robot dans la dynamique du système dual. Comme nous allons le constater dans la suite, le système dual est aussi un système réactif, sa dynamique influençant en retour la dynamique du système robotique, afin de développer et de contrôler les stratégies motrices (pour résoudre le PCI). Bergener et al. choisissent comme système dynamique dual le modèle des champs neuronaux.

Pour implémenter le système dual, Bergener et al. définissent d'abord le concept de *variables comportementales*. Celles-ci sont des paramètres scalaires ou vectorielles, qui caractérisent les comportements élémentaires du robot (par exemple la position dans l'espace 3D de l'extrémité d'un bras du robot, l'angle de rotation de sa tête, etc.). Ensuite, la valeur de chaque variable comportementale est associée à un état dynamique d'un champ neuronal (le système dual). Utilisant le modèle de champ d'Amari comme support (voir la section 3.2.2), les auteurs imposent aux ϕ -solutions de l'équation de champ la sémantique des états moteurs sans but, et aux solutions locales de type bulle(s) d'activité neuronale la sémantique des états moteurs correspondant à des différents buts précis. Les orbites et les points de bifurcation dans l'espace des phases de la dynamique du champ neuronal permettent de faire le passage du système robotique d'un état moteur à un autre, développant ainsi une stratégie d'exploration motrice.

Comme nous l'avons précisé ci-dessus, cette stratégie générale est appliquée pour guider le déplacement d'un robot dans un espace tridimensionnel. Les cibles visuelles reconnues par le robot sont modélisées par des stimulations excitatrices du champ neuronal, en temps que les obstacles sont modélisés par des stimuli inhibiteurs. La prise de décision pour le déplacement du robot se réalise grâce au mécanisme de compétition neuronal, qui génère une seule bulle d'excitation à la fois. Dans le même temps, l'amplitude de cette bulle modélise le niveau de confiance de la décision prise, est augmentée, de manière proportionnelle, la vitesse de déplacement du robot. Pour éviter les impasses de revisiter les endroits déjà explorés, le modèle utilise un mécanisme de mémoire à court terme.

De manière similaire, le modèle intègre deux modules supplémentaires de traitement sensorimoteur, pour modéliser respectivement le mouvement du bras et de la tête du robot.

La stabilisation de l'état global du système se réalise grâce aux boucles de feedback entre les modules de traitement du modèle. L'intégration permanente de ces influences dans la dynamique du robot produit un ajustement graduel des différents paramètres du système, et induit un fonctionnement lisse et robuste de ses mouvements. Ainsi, un changement comportemental n'est pas un changement brusque, comme dans les approches de traitement à base de règles symboliques.

Les modules de traitement du système robotique évoluent à des échelles de temps différentes. Par exemple, le mouvement de la tête est plus rapide que le mouvement du bras du robot qui, à son tour, est plus rapide que le déplacement du corps du robot. Cette séparation temporelle permet un couplage « flou » entre les différents modules de traitement, tout en gardant un fonctionnement interne stable et cohérent, pour chacun de ces modules. Le même aspect favorise l'émergence des hiérarchies de traitement d'information, hiérarchies qui peuvent changer dynamiquement si les rapports entre les échelles de temps changent (ou s'inversent) au cours du temps. Ce concept de séparation temporelle entre les différents modules de traitement d'un système dynamique a été pris en compte dans la modélisation de l'architecture neuronale d'auto-organisation que nous avons décrite dans la section 2.6.1.

Pour conclure, soulignons le fait que cette approche de modélisation par un système dual s'inscrit dans l'approche générale de calcul neuronal basée sur la dynamique des réseaux neuronaux, approche que nous avons décrite dans la section 2.3.2.

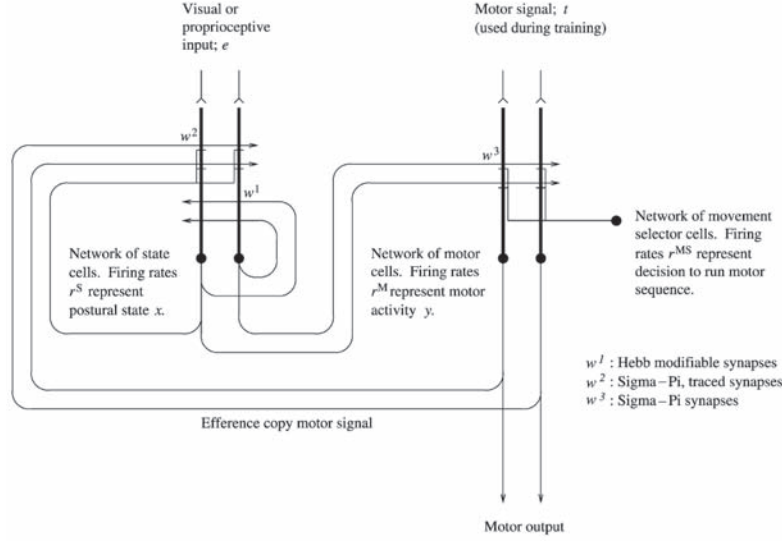


FIG. 4.5 – Architecture neuronale de traitement sensorimoteur, proposée par Stringer et al. dans [110]. Le modèle comprend un module postural, un module moteur (qui encode la commande motrice courante), et un module décisionnel (qui conçoit la commande motrice future). L’activité de chaque module est implémentée par la dynamique d’un champ neuronal. Suivre le texte pour les détails. Illustration graphique, extraite de [110].

4.3.3 Modélisation par duplication des commandes motrices

Partageant la même vision dynamique du traitement moteur que celle décrite par Bergener et al., Stringer et al. s’intéressent dans [110] à l’intégration de la copie efférente du signal moteur dans l’ensemble des processus d’apprentissage neuronal.

Selon les théories neuromotrices courantes (pour une vue d’ensemble des modèles de contrôle moteur de type feedforward, le lecteur est invité à consulter par exemple [82]), chaque fois qu’une commande motrice (c’est-à-dire un signal nerveux efférent) est envoyée aux effecteurs, le signal moteur serait dupliqué et traité en parallèle, afin de distinguer ses conséquences de celles dues aux actions externes au corps, déclenchées par une interaction avec l’environnement. Ainsi, ce mécanisme permettrait non seulement la conception d’un modèle moteur interne de type feedforward (qui contribuerait donc à la résolution du PCD), mais il faciliterait aussi la prédiction des actions motrices initiées par le système nerveux, pour améliorer le contrôle moteur.

Ayant le but d’implémenter ce mécanisme neuronal dans un système robotique autonome, Stringer et al. proposent dans [110] une architecture neuronale composée de trois modules de traitement (voir la figure 4.5), chacun gouverné par une dynamique similaire au modèle de champ neuronal d’Amari (revoir la section 3.2.2). Le premier module, appelé module d’état ou module postural, encode la position courante du robot, position qui est perçue grâce aux capteurs visuels et proprioceptifs du robot. Le deuxième module neuronal, le module moteur, encode la commande motrice courante effectuée par le robot. Pour finir, le module moteur décisionnel a le rôle de sélectionner la prochaine commande motrice à être exécutée.

Ces trois modules communiquent entre eux via deux principales boucles synaptiques. La

première voie synaptique part d'un couplage Sigma-Pi⁴ liant les unités des modules postural et moteur, et retourne vers le module postural. Celle-ci permet au modèle neuronal d'apprendre la corrélation entre une commande motrice réalisée et la position courante du robot dans l'espace, ce qui lui permet finalement de résoudre le PCD. Une seconde voie synaptique connecte un couplage Sigma-Pi des unités des modules postural et décisionnel avec les unités du module moteur. Cette deuxième boucle a le rôle d'implémenter le modèle moteur inverse (donc de résoudre le PCI) : chaque fois qu'une commande motrice ou un déplacement vers une nouvelle position dans l'espace 3D sont envisagés par le robot, cette boucle conçoit la nouvelle commande motrice (ou séquence de commandes), qui pourrait réaliser ce but.

Dans ce modèle (voir la figure 4.5), les poids w^1 encodent la posture du robot. Les poids w^2 du couplage Sigma-Pi de la première voie synaptique encodent les actions motrices du robot, et sont ajustés en fonction d'une trace temporelle de l'activité des unités posturales, et non pas de la réponse instantanée de ces unités. Cet aspect permet au module postural de créer un apprentissage robuste, et plus rapide, du modèle moteur direct. Deuxièmement, le niveau de repos de l'activité du champ neuronal, (le terme h de l'équation 3.15), s'ajuste aussi dynamiquement, variant de manière différente pour chaque unité individuelle du champ.

Afin de résoudre le PCD, le modèle suit d'abord une étape d'entraînement. Pendant celle-ci, des commandes motrices sont générées automatiquement, le robot percevant leurs conséquences (de manière visuelle et proprioceptive), ce qui lui permet d'apprendre le modèle moteur direct. Les poids synaptiques w^2 s'adaptent et se stabilisent au bout de ce processus. Comme particularité des résultats expérimentaux de cette étape, Stringer et al. remarquent que ces poids de connexion, qui intègrent la trace temporelle de l'apprentissage postural, deviennent asymétriques en cours du temps. Sur le long terme, cet aspect permet au modèle de découvrir plus facilement les motifs temporels dans le flot de l'information traitée.

Le PCI est résolu par l'apprentissage des poids w^3 de la seconde voie synaptique. Ainsi, après l'étape d'entraînement, le robot est soumis à la résolution de plusieurs tâches motrices, comme par exemple contrôler la force et la vitesse de ses mouvements, ou atteindre une cible fixe. Dans cette étape, le modèle d'apprentissage n'utilise pas des connexions de feedback, mais repose sur les capacités motrices acquises pendant la première étape. Cependant, le modèle laisse ouverte la possibilité d'utiliser un contrôle moteur de type feedback immédiat, qui pourrait intégrer pendant l'étape d'entraînement les informations sensorielles, directement dans le signal moteur.

Afin de résoudre le PU, c'est-à-dire pour optimiser le modèle cinématique inverse par rapport aux buts moteurs considérés, les auteurs considèrent deux approches générales de modélisation. Dans un premier temps, ils intègrent dans leur modèle un apprentissage implicite (par bootstrapping). Toutefois, pour que cette méthode ait du succès, il est nécessaire que les commandes motrices correctes dominent le bruit restant dans le signal moteur après la phase d'entraînement du système.

Dans une approche alternative, les auteurs utilisent le paradigme d'apprentissage explicite (par renforcement) pour résoudre le PU. Dans ce cas, le système est récompensé (et permet l'ajustement de ses poids synaptiques) seulement quand une séquence d'actions motrices conduit à la satisfaction du but demandé, et non pas à chaque instant que le robot s'approche de son but (par exemple une cible fixe).

⁴Deux couches neuronales A et B sont connectées par un couplage de type *Sigma-Pi* vers une autre couche neuronale C , si l'activité post-synaptique de chaque unité c_i de la couche C est influencée par une activité cumulée des deux autres couches, portant la forme $\sum_{j,k} w_{ijk} a_j b_k$. Dans cette expression, où a_j est le taux de décharge pré-synaptique d'une unité de la couche A , b_k est le taux de décharge pré-synaptique d'une unité de la couche B , et w_{ijk} désigne le poids de connexion post-synaptique de l'unité i de la couche C , qui est activée par les unités pré-synaptiques des unités j et k respectivement provenant des couches neuronales A et B .

Dans les deux cas d'apprentissage, le système réussit à minimiser, au fur et à mesure, ses erreurs par rapport à l'exécution des commandes motrices. Par conséquent, dans les scénarios considérés, le robot arrive à accomplir les tâches demandées avec une haute précision, même en absence des stimuli visuels et/ou proprioceptifs. De plus, le système est capable d'exécuter ses mouvements avec des vitesses et des forces arbitraires, produisant les séquences motrices toujours dans le même ordre, respectant ainsi les principes du programme moteur généralisé énoncés par Schmidt dans [102].

4.3.4 Modélisation par utilisation d'une carte cognitive de cellules de transition

Dans [26], Cuperlier et al. décrivent la conception d'une architecture neuronale modulaire pour résoudre à la fois PM et PU. Ce modèle a le but d'implémenter un mécanisme adaptatif pour le déplacement d'une plate-forme robotique mobile dans une enceinte fermée (divisée éventuellement dans plusieurs pièces, à l'intérieur desquelles peuvent être placés différents objets fixes), a priori inconnue par le robot.

Les auteurs implémentent un processus d'exploration de l'environnement, pendant lequel le robot construit une carte topographique de l'espace découvert. Cette carte est composée d'un ensemble de nœuds discrets, appelés *cellules de position*, chacune d'elle encodant la position du robot relative à l'environnement. Bien que cette structure aide à résoudre le PCD, elle s'avoue insuffisante pour planifier de manière efficace les actions du robot, surtout quand la complexité de l'espace exploré offre plusieurs possibilités pour atteindre une certaine position de l'environnement.

Afin de traiter le PCI, les auteurs proposent la conception d'une structure neuronale supplémentaire : une nouvelle carte, composée de plusieurs *cellules de transition*. Comme leurs nom le suggère, chaque cellule de cette carte encode une possible transition, d'une cellule de position vers une autre. Cette carte cognitive permet ainsi de définir des trajectoires entre deux positions quelconques de l'enceinte, qui ont été déjà explorées.

Selon les auteurs, l'architecture neuronale de leur modèle s'inspire des mécanismes de traitement séquentiel de l'information, implémentés dans les structures cérébrales du hippocampe et du cortex préfrontal.

Le robot vise à satisfaire plusieurs buts (donc résoudre un PU multiple), se qui se traduit par une exploration continue de l'environnement pour trouver les endroits qui lui permettent d'accomplir ces buts. Dans le modèle neuronal proposé par Cuperlier et al., chaque but a associé un indicateur de satisfaction (une valeur scalaire), qui décroît progressivement avec le temps passé depuis la dernière satisfaction du but. Le but courant à satisfaire est choisi selon un processus de compétition.

Dans ce contexte, l'action motrice à effectuer (la transition vers une nouvelle position) est réalisée grâce à un mécanisme de type champ neuronal dynamique. Le champ reçoit en entrée plusieurs angles de direction que le robot puisse choisir. Si ces angles possibles sont suffisamment proches, le résultat de la compétition neuronale indiquera un nouvel angle de direction moyennant ces valeurs. Au contraire, si ces angles diffèrent beaucoup entre eux, le mécanisme de compétition du champ sélectionnera un seul angle de direction parmi toutes ces variantes. Grâce à la dynamique intrinsèque du champ, le résultat de la compétition est toujours stable, ce qui permet au robot de construire une stratégie de déplacement robuste.

Comme les auteurs remarquent, l'usage du mécanisme de champ neuronal décrit ici peut s'étendre pour inclure le traitement de plusieurs sources d'information (visuelle, auditive, etc.), afin de permettre un traitement sensorimoteur multimodal.

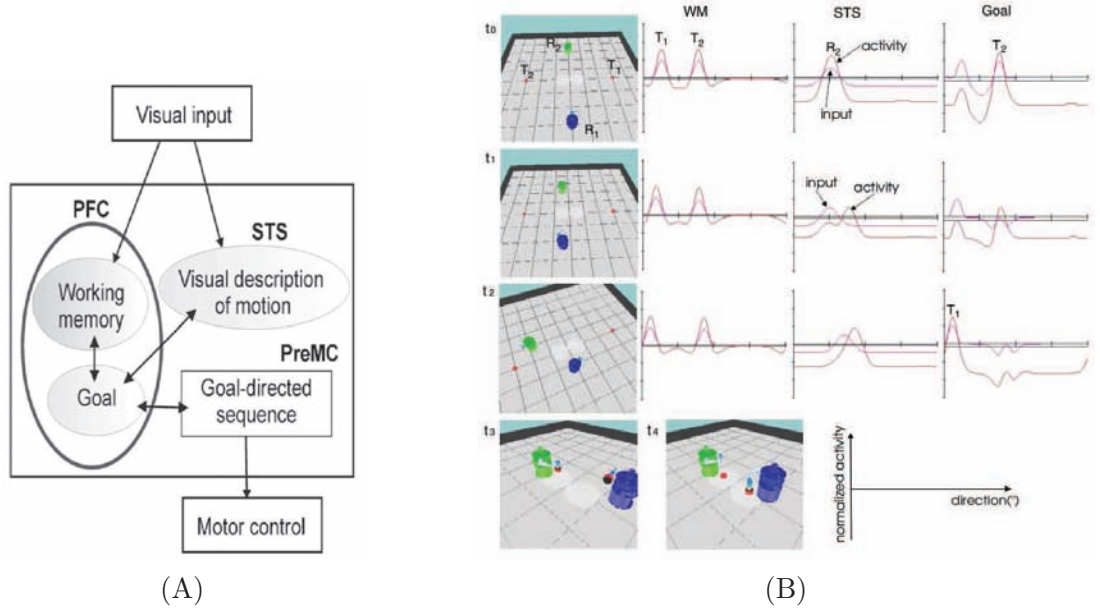


FIG. 4.6 – (A) Architecture neuronale de traitement sensorimoteur, proposée par Erlhagen et al. dans [36]. (B) Plusieurs phases de l'évolution de la dynamique des modules qui composent cette architecture, pendant un scénario de coordination entre deux robots ayant le but de collecter des objets de la scène visuelle. Suivre le texte pour les détails. Illustrations graphiques extraites de [36].

4.3.5 Modélisation de la coordination des tâches robotiques

Intéressés en général par l'intégration des approches de modélisation bio-inspirée dans le développement des plates-formes robotiques, Erlhagen et ses collaborateurs ont abordé dans [35, 36], d'une manière plus focalisée, la problématique de conception des systèmes situés utilisant la dynamique des champs neuronaux.

Traitement de la coordination motrice de deux robots

Un premier cas de modélisation vise la coordination de deux robots mobiles, qui doivent résoudre une tâche motrice commune : trouver et amener des objets dans un certain endroit de la scène spatiale. Les robots ne communiquent pas entre eux explicitement, mais ils sont capables d'observer et d'inférer l'activité de leur partenaire. Pour modéliser le traitement sensorimoteur de chaque robot, les auteurs proposent une architecture neuronale comme celle illustrée dans la figure 4.6 (A).

Dans ce modèle, l'entrée visuelle stimule d'abord une couche neuronale *WM* (provenant de l'anglais *working memory*), qui implémente le modèle de mémoire de travail présenté dans la dernière partie de la section 4.2.1, et qui est basé sur le modèle d'Erlhagen décrit dans la section 3.2.4. Ce champ neuronal encode le contexte visuel donné par le positionnement des objets dans la scène 3D. Cette mémoire de travail permet au robot de maintenir la cohérence du flot perceptif visuel, même dans les cas d'occlusion temporaire des objets de la scène visuelle. Cette couche influence directement un autre champ neuronal, qui encode le but moteur du robot. Ensemble, les deux couches composent le module *PFC* (provenant de l'anglais *prefrontal cortex*).

Un deuxième module de traitement, représenté dans la figure 4.6 (A) par la couche *STS*

(provenant de l'anglais *superior temporal sulcus*), encode le déplacement du robot partenaire. Pour implémenter ce module, Erlhagen et al. utilisent une dynamique de type onde de propagation d'un champ neuronal, qui a le rôle d'anticiper le mouvement du robot partenaire. L'activité de ce module s'ajuste en fonction du but moteur établi par le robot, mais influence à son tour la deuxième couche du module *PFC*, qui encode ce but. Ce module vise à maintenir la cohérence entre la perception visuelle et le but moteur suivi par le robot.

Le troisième et dernier module l'architecture présentée dans la figure 4.6 (A), *PreMC* (provenant de l'anglais *premotor cortex*), modélise une séquence d'actions motrices, qui doit être réalisée par le robot pour atteindre son but.

L'architecture décrite ici est utilisée par Erlhagen et al. pour coordonner les actions motrices de deux robots (c'est-à-dire un PU partagé). Chaque robot analyse la scène 3D. Les objets saillants de la scène visuelle stimulent la couche *WM* (voir la figure 4.6 (B)). En même temps, la couche *STS* qui modélise l'activité du robot partenaire, s'active. Un processus de compétition neuronale sélectionne le but du robot, c'est-à-dire choisi un objet à collecter de la scène, et induit un déplacement du robot vers cet objet. Par conséquent, la couche *STS* du robot partenaire s'active, générant une onde de propagation, qui aide à inférer cette décision. Si ce robot choisit le même objet, l'activité de cette couche neuronale inhibe sa décision. La capacité de prédiction de la couche *STS*, ainsi que la capacité de mémorisation de la couche *WM*, assurent la continuité du comportement moteur, pour chaque robot qui suit son but, ceci n'étant pas interrompu pendant les brèves occlusions du robot partenaire par les objets de la scène.

Notons que ce modèle de traitement sensorimoteur utilise exclusivement les propriétés dynamiques de l'interaction robot – environnement, sans intégrer un modèle d'apprentissage pour améliorer son comportement en cours du temps. Cette approche peut ainsi devenir assez limitée, quand les buts et les interactions du robot avec l'environnement augmentent en complexité.

Traitement sensorimoteur par imitation

Dans la deuxième application pratique, proposée par les mêmes auteurs de [36], un robot est censé saisir un objet et le placer à un endroit établi, derrière un obstacle qui est représenté par un pont (voir la figure 4.7 (A)). Le défi de la tâche oblige le robot à décider s'il doit déplacer l'objet par au-dessous ou par au-dessus du pont, selon sa hauteur. Pour résoudre cette tâche, donc pour résoudre le PCI, Erlhagen et al. font appel aux résultats expérimentaux de Rizzolatti et al. [97], qui avaient mis en évidence l'existence des neurones miroir (neurones qui s'activent quand un geste donné est effectué non seulement par le sujet même, mais aussi par un autre individu). Par conséquent, Erlhagen et al. proposent l'implantation d'un circuit neuronal miroir, qui a le but de simuler une action effectuée par un autre agent, utilisant ses propres moyens moteurs. Dans le cas de cette application, l'agent qui aide le robot, en lui montrant plusieurs exemples de résolution de cette tâche motrice, est une personne humaine.

Afin d'intégrer dans son comportement l'expérience exhibée par l'agent humain, le robot étend l'architecture décrite dans la figure 4.6 (A) (voir pour comparaison la figure 4.7 (B)). Dans ce nouveau modèle, la couche *F5* encode le répertoire des actions motrices élémentaires acquises par le robot, en même temps que la couche *PF* aide à combiner les actions qui seraient capables d'atteindre le but moteur du robot (codé par la couche secondaire du module *PFC*). Modélisant la caractéristique fonctionnelle des neurones miroir, la couche *F5* s'active à la fois quand une action motrice est effectuée par un agent, et à la fois quand elle est effectuée par le robot lui-même.

Pour apprendre l'association entre le but moteur du robot et les actions motrices effectuées par l'agent, les poids des connexions synaptiques entre les couches *STS* et *PF* s'ajustent au

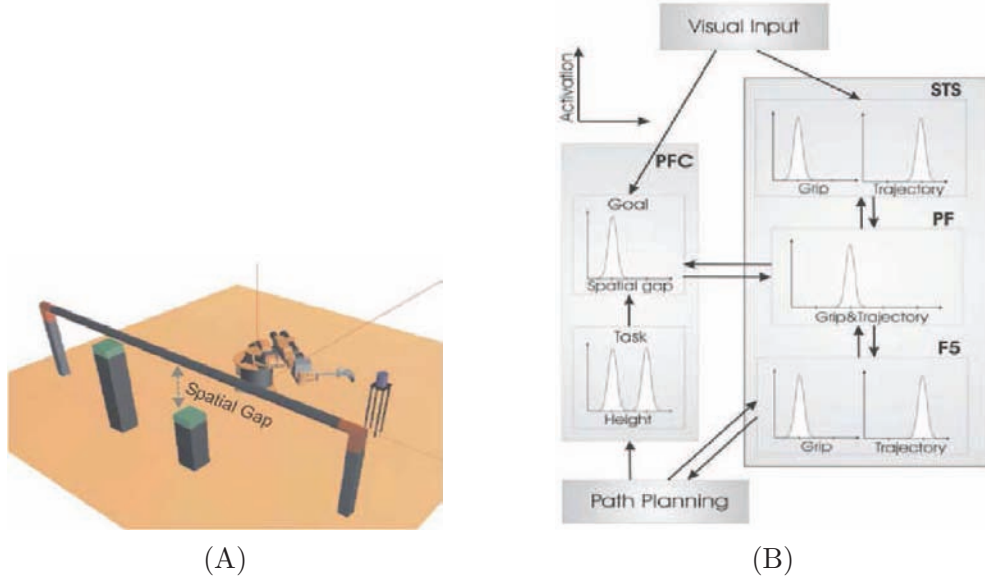


FIG. 4.7 – (A) Représentation visuelle d’une tâche robotique motrice, qui vise à déplacer des objets, d’une manière optimale, par dessous ou par dessus un pont, et les placer correctement sur d’autres objets. (B) Schéma de fonctionnement dynamique d’un modèle neuronal qui étend l’architecture illustrée dans la figure 4.6 (A), utilisée pour résoudre cette tâche motrice. Suivre le texte pour les détails. Illustrations graphiques extraites de [36].

cours du temps, selon une règle d’apprentissage de type Hebbien.

Comme les auteurs remarquent, cette approche d’apprentissage moteur par imitation est efficace dans le scénario considéré, mais elle est dépendante de l’expérience fournie par l’agent externe. De plus, si le contexte du scénario change d’une manière sensible, le robot trouve difficilement ou même impossible l’adaptation de son comportement, afin de résoudre les nouveaux buts. Pour cette raison, afin de concevoir des plates-formes robotiques, il est essentiel que le traitement moteur, facilité par ce paradigme d’apprentissage imitatif, soit couplé avec le développement d’un comportement intrinsèquement générique, capable d’explorer et de s’adapter en permanence aux changements perçus par le système situé.

4.3.6 Modélisation intégrant l’apprentissage par renforcement

Les deux approches décrites dans les paragraphes suivants présentent deux démarches qui, afin de concevoir une solution pour le PU, font appel au paradigme d’apprentissage par renforcement [112]. Dans ce contexte, les champs neuronaux dynamiques ont le rôle d’encoder dans la forme de leurs solutions stables (en particulier dans l’amplitude et l’extension spatiale des bulles d’activité) certains aspects numériques qui permettent d’implémenter une version distribuée de l’algorithme d’itération de la valeur ou de l’algorithme de Q-learning.

Implantation distribuée de l’algorithme d’itération de la valeur

Dans [115], Toussaint propose une architecture de traitement sensorimoteur d’inspiration corticale, capable d’apprendre et de contrôler l’influence des actions effectuées par un robot,

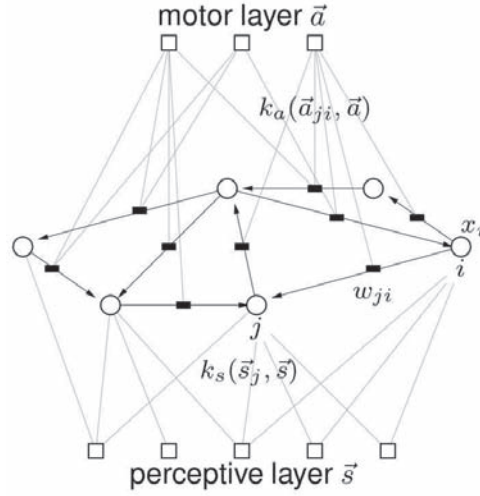


FIG. 4.8 – Architecture neuronale de traitement sensorimoteur, proposée par Toussaint, composée par trois couches neuronale : une perceptive, une motrice, et une corticale (générique), associant les deux autres couches. Le modèle vise l'implémentation de deux processus dynamiques : un responsable pour l'estimation de l'état courant du système, et un responsable pour la planification des actions motrices futures. Suivre le texte pour les détails. Illustration graphique extraite de [115].

vis-à-vis de la réalité qu'il perçoit. Le fonctionnement du robot est le résultat des interactions de trois couches de traitement : une couche perceptive, une couche corticale et une couche motrice (voir la figure 4.8).

Les unités de la couche perceptive répondent selon la reconnaissance d'une certaine caractéristique dans le flot d'information enregistrée par les capteurs du robot. Les unités de la couche motrice encodent les commandes motrices réalisées par le robot. La boucle perception-action est arbitrée par la dynamique de la couche corticale. L'activité neuronale de chaque unité de la couche corticale est modélisée par une équation différentielle de type champ neuronal. Celle-ci s'ajuste en permanence, prenant en compte d'une part les stimuli provenant de la couche perceptive, et d'une autre part l'influence cumulée et pondérée des activités des autres unités de la couche corticale, chacune modulée par un facteur d'activation dépendant de la commande courante encodée par la couche motrice.

Pour développer un comportement orienté vers un but, le modèle dynamique intègre un mécanisme d'apprentissage par renforcement. Plus précisément, la conception consiste à introduire un nouveau champ, appelé champ-valeur, qui implémente une version distribuée de l'algorithme d'itération de la valeur (voir [112] pour plus de détails), et améliore l'estimation de la récompense obtenue par le robot.

Le comportement du robot s'adapte selon l'ajustement des paramètres libres du modèle présenté, ajustement gouverné par trois processus dynamiques. Premièrement, grâce à l'algorithme de Growing Neural Gas [43], une nouvelle unité de traitement est ajoutée dans la couche perceptive, chaque fois que l'ensemble des unités actuelles de cette couche répondent de manière trop faible à un nouveau stimulus d'entrée. Deuxièmement, selon un apprentissage Hebbien, les champs récepteurs des unités de la couche perceptive et de la couche motrice s'ajustent au cours

du temps, s'organisant ainsi dans des cartes topographiques. En dernier lieu, une règle temporelle d'apprentissage Hebbien est implémentée pour ajuster les poids des connexions latérales de la couche corticale. Cette loi dynamique favorise les ajustements des poids qui entraînent l'augmentation de la corrélation entre les activations pré-synaptiques (provenant de la couche perceptive) et celles post-synaptiques (stimulant la couche motrice pour exécuter des commandes), et donc améliorent la prédiction des actions effectuées.

L'auteur expose les avantages de son modèle en étudiant le comportement du robot dans un scénario d'exploration d'un labyrinthe. Dans un premier temps, le robot explore aléatoirement mais exhaustivement l'environnement. Pendant cette étape, il construit de manière incrémentale une carte topographique discrète, représentant le labyrinthe découvert. Ensuite, quand on demande au robot de retrouver un objet placé dans un endroit quelconque du labyrinthe, son mode de fonctionnement change. Pendant cette étape, il utilise la carte construite précédemment pour concevoir des plans d'exploration du labyrinthe, ainsi que pour améliorer la stratégie de recherche guidée par le mécanisme de renforcement implémenté dans le modèle. Les résultats de cette expérience montrent la capacité du robot à se diriger dans le labyrinthe et à retrouver les objets rapidement. Cependant, les atouts de l'approche proposée par Toussaint sont mis en évidence quand la configuration initiale de l'environnement appris par le robot change à cause de l'insertion d'un ou plusieurs obstacles dans le labyrinthe. Comme le processus du robot responsable de la découverte du labyrinthe est un processus adaptatif, le robot est capable de traduire ces changements dans sa carte topographique, une fois qu'ils sont perçus. Grâce à cet aspect, la stratégie de planification change aussi dynamiquement. Par conséquent, le robot optimise en permanence son comportement, les résultats expérimentaux montrant aussi comment cette adaptation continue minimise les trajectoires employées par le robot pour atteindre les nouveaux objets, évitant les régions bloquées du labyrinthe.

Si ces résultats montrent la validité du modèle proposé par Toussaint, nous remarquons, comme dans l'approche d'Erlhagen et al., la nécessité du modèle de suivre une première étape d'apprentissage, entièrement distincte de l'étape suivante, d'exploitation motrice. Comme nous l'avons mentionné aussi précédemment, nous considérons que cet aspect impose certaines limites dans le fonctionnement du robot.

Implantation distribuée de l'algorithme de Q-learning

Afin de traiter le problème de la recherche d'un objet connu dans un environnement inconnu avec un robot mobile, Gross et al. proposent dans [53] une approche similaire à celle de Toussaint, basée sur le développement des stratégies d'apprentissage par renforcement avec des champs neuronaux. Pour stimuler le développement d'une capacité motrice adaptative, exigée par la complexité des scénarios pratiques, les auteurs imposent au fonctionnement du robot quelques contraintes supplémentaires. Ainsi, pour résoudre le PU, le robot dispose d'une quantité d'énergie limitée pour atteindre son but, il n'est pas en contact visuel avec la cible au début de l'expérience, et il reçoit une récompense positive seulement quand il réussit à reconnaître l'objet cible cherché. Le schéma architectural proposé par Gross et al. est décrit dans la figure 4.9.

Pour aborder le problème du traitement moteur, les auteurs utilisent une approche de modélisation dynamique basée sur l'algorithme de Growing Neural Gas [43], pour encoder de manière topographique les actions possibles du robot. Ensuite, Gross et al. utilisent des champs neuronaux bidimensionnels de type Amari (revoir la section 3.2.2), pour modéliser un mécanisme de prise de décision par un processus de sélection distribuée. La séquence des décisions prises définit la stratégie d'exploration du robot. Pour guider la stratégie du robot, les auteurs intègrent dans leur modèle une implantation distribuée de l'algorithme de *Q-learning* d'apprentissage par ren-

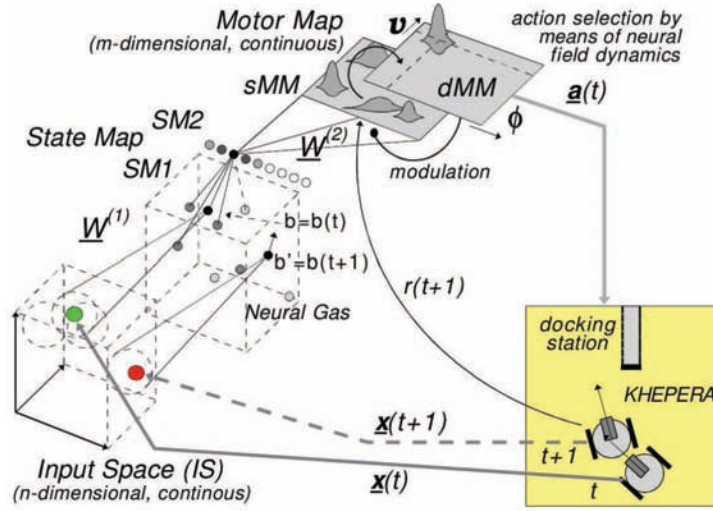


FIG. 4.9 – Architecture neuronale de traitement sensorimoteur, proposée par Gross et al. pour résoudre le guidage d'un robot dans une enceinte. Les paramètres qui définissent l'état courant du robot (IS) sont encodés par une couche neuronale (SM). Le modèle utilise une carte topographique (sMM) d'actions motrices, possibles à être réalisées par le robot, et un champ neuronal (dMM) pour choisir la meilleure action à effectuer. Suivre le texte pour les détails. Illustration graphique extraite de [53].

forcement, algorithme qui a le but d'évaluer les expériences passées pour apprendre la fonction valeur optimale (voir [112] pour plus de détails). Cette valeur et la probabilité d'exécuter une action donnée sont codées respectivement par l'amplitude et la variance des bulles d'activité d'un champ neuronal. Cette approche permet de développer une méthode d'exploration aléatoire et distribuée dans l'espace des actions.

Grâce à cette architecture de traitement d'inspiration corticale, les auteurs montrent que les résultats expérimentaux obtenus sont caractérisés par un taux de succès supérieur par rapport au cas d'implantation classique de l'algorithme Q-learning. Un aspect notable à mentionner est que le robot réussit à résoudre la tâche sans avoir la nécessité d'initialiser les valeurs d'action, une approche qui caractérise aussi la plupart des solutions traditionnelles du problème. De plus, le modèle de Gross et al. ne nécessite pas de faire la séparation entre l'étape de reconnaissance de l'environnement et l'étape de résolution effective de tâches motrices, comme dans le cas des approches de modélisation présentées précédemment.

4.4 Apprentissage d'architectures neuronales basé sur la dynamique des champs neuronaux

L'apprentissage réalisé par une architecture de traitement neuronal d'information se traduit le plus souvent par un processus d'ajustement dynamique des poids synaptiques des connexions reliant ses unités neuronales. Différents auteurs ont proposé en cours du temps des modèles neuronaux complexes, capables d'expliquer certaines propriétés de ces mécanismes d'apprentissage.

Dans les travaux que nous allons analyser dans la suite, l'activité des champs neuronaux

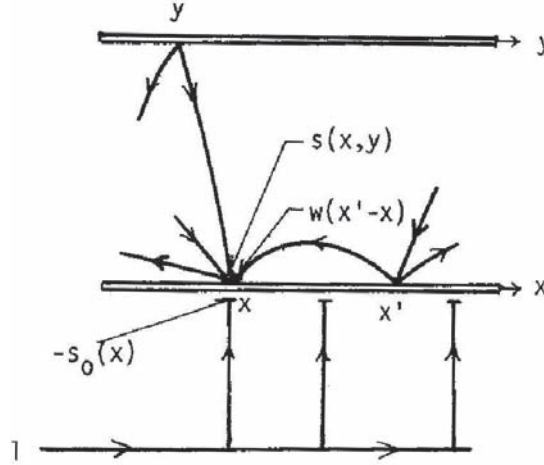


FIG. 4.10 – Architecture neuronale proposée par Amari pour modéliser le processus d’auto-organisation des poids synaptiques des connexions afférentes reliant différentes couches neuronales. Dans ce modèle, une distribution externe I stimule la couche Y via les connexions synaptiques afférentes $s_0(x)$, qui à sont tour, stimule la couche X via les connexions synaptiques $s(x, y)$. Suivre le texte pour les détails. Illustration graphique extraite de [5].

dynamiques est utilisée pour moduler (au niveau local) l’ajustement de certains paramètres de l’architecture d’apprentissage considérée. Les facteurs déterminants qui caractérisent la dynamique des champs et qui influencent le déroulement du processus d’apprentissage sont généralement : le temps de relaxation, l’amplitude, l’extension spatiale et la période de stabilité des bulles d’activité du champ. Comme nous allons le constater, dans ces contextes, le champ neuronal est aussi un mécanisme capable de favoriser l’émergence d’une cohérence de plusieurs activités neuronales, grâce à son mécanisme intrinsèque de compétition neuronale.

4.4.1 Apprentissage des connexions afférentes

La première approche résumée ici, concernant la modélisation des processus d’apprentissage, est celle d’Amari [5], qui propose une architecture neuronale capable d’exhiber une organisation topologique des connexions synaptiques.

Comme le montre la figure 4.10, l’architecture est composée par deux couches neuronales, notées X et Y . La couche Y fournit à la couche X une contribution excitatrice, via les connexions afférentes $s(x, y, t)$. En même temps, les unités de la couche X reçoivent également des contributions inhibitrices, provenant d’une autre couche neuronale I , de nature extra-corticale. La dynamique fonctionnelle des deux couches X et Y est décrite par l’équation 3.15 du modèle d’Amari de champ neuronal. Alors que les poids des connexions latérales relatives à chaque couche sont considérés fixes, les poids des connexions afférentes (reliant d’une part les unités des couches Y et X , et d’une autre part celles reliant les unités des couches I et X), s’ajustent en cours du temps. La plasticité de ces connexions ($s_0(x)$ et $s(x, y)$), est modélisée par une loi Hebbienne, et est modulée par l’activité des deux couches neuronale, X et Y :

$$\tau_1 \Delta s_0(x, t) = -s_0(x, t - 1) + \alpha_1 f(u(x, t - 1)) \quad (4.3)$$

$$\tau_2 \Delta s(x, y, t) = -s(x, y, t - 1) + \alpha_2 a(y, t - 1) f(u(x, t)) \quad (4.4)$$

où $a(y, t)$ désigne l'activité de l'unité y de la couche I , $u(x, t)$ désigne l'activité de l'unité x de la couche Y , $\tau_{1,2}$ représentent des constantes de temps, et $\alpha_{1,2}$ sont des valeurs scalaires positives.

Selon Amari, suite au processus d'apprentissage décrit par ces équations, les connexions afférentes entre les couches X et Y atteignent une auto-organisation topographique. Amari étudie analytiquement les conditions nécessaires pour l'émergence de ce phénomène, mettant en évidence également les cas d'auto-organisation de ces connexions afférentes dans plusieurs structures spatiales similaires aux colonnes corticales.

En dépit des résultats expérimentaux obtenu par cette modélisation, Amari remarque la difficulté d'obtenir une auto-organisation si les poids synaptiques sont considérés initialement aléatoires. Notons également que le processus d'auto-organisation modélisé par Amari ne concerne pas la capacité de quantification vectorielle, qui caractérise les cartes corticales d'auto-organisation.

4.4.2 Apprentissage multimodal

Une deuxième approche de modélisation pour un mécanisme d'apprentissage neuronal est formulée par Ménard et Frezza-Buet dans [79]. Dans cet étude, les auteurs s'intéressent à la conception d'une architecture de calcul générique de traitement multimodal, (appelée **bijama** [41]), tout en respectant les principes de traitement cortical présentés dans la section 2.5. Les auteurs proposent un modèle d'auto-organisation des modules de traitement cortical qui interagissent entre eux, et qui intègrent également un processus d'apprentissage par renforcement, afin de coordonner et maintenir la cohérence de l'activité globale du système.

Le modèle définit dans un premier temps un ensemble de cartes neuronales unimodales, chacune encodant une seule caractéristique (de perception ou d'action) du flot d'information traitée (par exemple la préférence d'orientation). Le modèle définit ensuite des cartes associatives qui rapprochent les informations provenant des différentes cartes unimodales.

Pour chaque carte neuronale, le cycle de traitement neuronal se déroule suivant les trois étapes énumérées ci-dessous :

- chaque unité de la carte évalue son niveau de sensibilité actuelle par rapport à une nouvelle stimulation externe ;
- les activités issues de la première étape entrent en compétition locale (grâce à un mécanisme de type champ neuronal), générant une réponse neuronale caractérisée par des bulles d'excitation dans des endroits correspondant à une forte activité d'entrée
- une règle d'apprentissage de type SOM ajuste les poids synaptiques entre les unités de la carte, renforçant proportionnellement les connexions qui correspondent à une forte réponse du champ neuronal considéré dans la deuxième étape

Comme dans le cas des SOM, le processus neuronal décrit ici converge vers l'auto-organisation des poids des connexions synaptiques entre les unités de la carte, modélisant ainsi par des relations topologiques une représentation distribuée et discrète d'un ensemble d'entrée.

Ce qui est plus spécifique à l'architecture du modèle **bijama** est la modélisation des connexions associatives entre les différentes cartes unimodales. Les auteurs du modèle proposent un mécanisme de connexion avec des directions préférentielles : deux cartes unimodales peuvent être connectées via des bandes directionnelles d'unités (voir la figure 4.11). Ce mécanisme assure l'implantation d'un processus de relaxation dynamique, ayant le but de trouver et maintenir la cohérence dynamique de plusieurs activités neuronales (grâce à un phénomène de résonance neuronale, similaire à celui proposé par Grossberg dans [54]). Notons également que les connexions associatives peuvent être considérées comme une approche de modéliser l'influence des connexions cortico-corticales à longue distance [55, 91], aspect souvent ignoré dans la modélisation classique des champs neuronaux dynamiques.

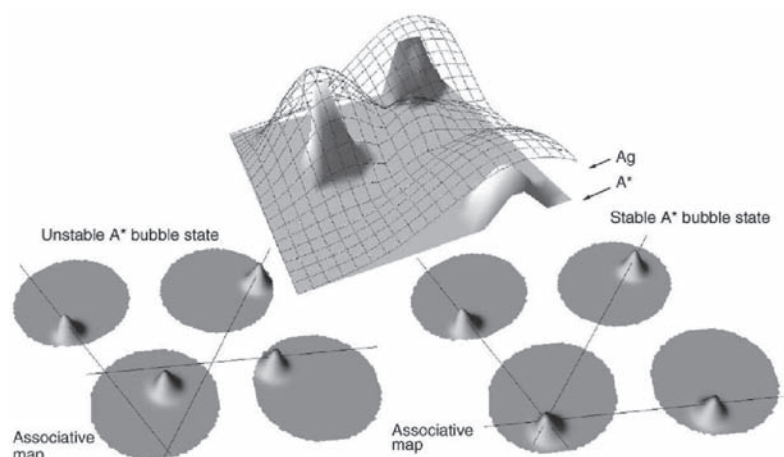


FIG. 4.11 – Architecture neuronale d’inspiration corticale, proposée par Ménard et Frezza-Buet pour modéliser des processus de traitement multimodal. L’activité neuronale de plusieurs cartes associatives unimodales est mise en correspondance via des connexions synaptiques préférentielles. Un processus d’apprentissage distribué de ces connexions vise à adapter en permanence l’activité de ces cartes, afin de maintenir la cohérence du traitement multimodal. Suivre le texte pour les détails. Illustration graphique extraite de [79].

Cette architecture neuronale à vocation à faire le passage au niveau de la modélisation multimodale, permettant de proposer une solution générique (ou un cadre général) pour traiter aussi le *problème de la liaison* (en anglais *binding problem*), souvent posé par la psychologie Gestalt.

Les auteurs utilisent cette plate-forme neuronale pour développer deux applications de traitement cognitif, notamment l’encodage phonétique dépendant de la sémantique des mots ou la résolution d’une tâche de mouvement d’un bras robotique pour atteindre une position dans l’espace visuel tridimensionnel. Les résultats obtenus montrent la capacité du modèle proposé à s’adapter et traiter de manière générique les flots d’information de différentes natures.

Le caractère modulaire et générique du traitement de l’information implémenté par ce modèle vise à offrir une approche de modélisation évolutive, capable de traiter d’une manière distribuée et interdépendante plusieurs sources sensorimotrices du flot informationnel.

L’architecture présentée ici décrit un modèle de traitement neuronal qui repose sur la capacité d’auto-organisation des cartes corticales unimodales. De ce point de vue, notre démarche de concevoir un mécanisme générique d’apprentissage d’auto-organisation, basé sur la dynamique des champs neuronaux (voir la section 2.6.1), vise à aider ce modèle à mieux exploiter ses capacités fonctionnelles.

4.5 Discussion

Comme nous avons pu constater le long de ce chapitre, le paradigme dynamique des champs neuronaux a été appliqué avec succès dans une grande variété d’approches de modélisation neuronale.

Cependant, différents aspects pratiques ou théoriques font que le spectre des comportements

dynamiques exhibés par les champs neuronaux est seulement partiellement mis en évidence expérimentalement, ou exploité dans les applications de modélisation neuronale. Pour cette raison, il est nécessaire que les recherches futures des champs neuronaux continuent à explorer les possibilités d'utiliser leurs propriétés dynamiques dans les contextes pratiques de modélisation neuronale.

Si certaines de ces approches visent à corrélérer au mieux leurs résultats avec les observations expérimentales, nous considérons que le fait de suivre strictement la plausibilité biologique de ces modèles ne doit pas être un but en soi, pour tous les modèles conçus. De ce point de vue, comme souligné dans la section 2.5, nous considérons plus important le fait de respecter certains principes de modélisation neuronale, afin de maintenir l'évolutivité des approches de modélisation proposées.

Analysant les approches de modélisation neuronale présentées dans ce chapitre, nous avons pu constater que la majorité de ces approches utilise la dynamique des bulles d'activité neuronale, et seulement une minorité fait usage des phénomènes d'ondes de propagation de l'activité d'un champ, ou d'autres phénomènes dynamiques instables exhibés par un champ neuronal. Comme les bulles d'activité représentent les attracteurs du système dynamique décrit par le champ, la majorité des approches examinées ici suivent les lignes générales du paradigme de calcul neuronal basé sur la dynamique des attracteurs, décrit précédemment dans la section 2.3.2.

Résumant l'analyse des modèles pratiques utilisant le paradigme des champs neuronaux dynamiques, nous pouvons affirmer que ce modèle général d'inspiration corticale permet d'implanter des mécanismes neuronaux d'encodage, de sélection, de compétition, de coordination et de traitement locale et distribué d'une ou plusieurs sources d'informations perceptives, d'actions ou de commandes motrices. En outre, nous soulignons le rôle des champs neuronaux dynamiques dans la modélisation des processus cognitifs élémentaires et des processus de plasticité neuronale (d'apprentissage).

Dans un contexte plus général de modélisation neuronale, toutes les approches résumées ici s'inscrivent dans la vision générale exposée par van Gelder concernant le traitement cognitif [120], vision qui considère que les processus cognitifs ont une nature intrinsèquement dynamique, essentiellement différente de la nature explicitement prédictive des représentations symboliques. Comme nous l'avons déjà évoqué dans la section 1.3, cette vision de calcul dynamique, que nous défendons également, a un rôle fondateur dans notre démarche générale de recherche.

Tout le long de notre discours, nous avons réitéré notre intérêt vis-à-vis de la modélisation des processus d'apprentissage qui régissent la dynamique du traitement cortical. Cependant, si les approches examinées dans ce chapitre ont abordé plusieurs aspects concernant ce sujet, elles n'offrent pas une réponse définitive à la problématique de modélisation des processus corticaux d'auto-organisation. L'analyse des approches présentées ici nous permet d'aborder dans la suite ce problème de modélisation, dans la perspective des champs neuronaux dynamiques, formulée initialement dans la section 2.6.1.

Chapitre 5

Méthodologie d'étude empirique des champs neuronaux dynamiques

Conçus comme un modèle de la compétition entre les activités neuronales de la structure corticale, les champs neuronaux dynamiques sont un formalisme mathématique suscitant un intérêt croissant de la part de la communauté des chercheurs des neurosciences. Bien que le chapitre précédent mette en évidence les avancées notables de ce domaine de recherche, différents aspects de modélisation (et en particulier le caractère non-linéaire du modèle) rendent la dynamique du modèle suffisamment complexe pour que la compréhension de son comportement devienne un problème d'étude en soi. Par conséquent, chaque nouveau modèle de champ proposé dans la littérature a bénéficié de nombreuses études analytiques concernant son comportement dynamique. Toutefois, ces études n'offrent pas aujourd'hui toutes les réponses nécessaires pour acquérir une compréhension exhaustive de ces comportements.

Si la nécessité de comprendre la phénoménologie corticale a conduit à l'apparition des formalismes théoriques de type champs neuronaux dynamiques, la nécessité de développer des systèmes situés, inspirés du fonctionnement cérébral, réclame l'utilisation pratique des résultats acquis par ces travaux de modélisation. Comme nous l'avons évoqué dans le chapitre 4, le paradigme des champs neuronaux dynamiques a été utilisé avec beaucoup de succès dans la conception bio-inspirée des systèmes situés.

Malgré les résultats obtenus par les deux axes de recherche (théorique et appliquée), nous considérons qu'au cours du temps, un certain écart s'est creusé entre leurs objectifs. Ainsi, dans la plupart des cas, les recherches théoriques ont été focalisées sur la dynamique des solutions homogènes (en absence de stimulation externe), ou décrivent parfois un contexte d'étude simplifié (linéarisant les équations de champ). À l'autre extrémité du spectre des recherches, les applications utilisant les champs neuronaux dynamiques comme « moteur neuronal » ont dû souvent se confronter à un problème de réglage de paramètres pour les équations de champ, afin de mettre en accord la dynamique d'un champ avec les contraintes imposées par le fonctionnement de l'architecture neuronale implémentée.

Concluant sur ces remarques, nous considérons nécessaire un rapprochement des deux approches, théorique et appliquée, afin de relier les exigences demandées par l'utilisation pratique de ce modèle avec l'étude théorique de son comportement dynamique.

Pour répondre à ces requêtes d'ordre pragmatique, nous proposons dans ce chapitre une méthode alternative d'analyse, pour examiner et régler le comportement dynamique des champs neuronaux dans des contextes pratiques. Dans la première section, nous présentons le contexte d'analyse des modèles de champs, introduisant aussi les notions de *scénario* et d'*expérience*,

relatives à l'évaluation dynamique d'un champ. Ensuite, nous proposons un ensemble de *critères de performance* et un *instrument de mesure* pour évaluer la capacité du champ à satisfaire ces critères. Ceci permet de concevoir une méthode générale d'évaluation empirique par simulation numérique, capable d'estimer la *qualité dynamique* d'un champ par rapport aux critères de performance considérés. Basé sur cette procédure, dans la dernière partie du chapitre, nous présentons une méthodologie d'ajustement de paramètres pour les équations de champ, afin d'améliorer la qualité dynamique du modèle de champ examiné.

5.1 Scénarios pertinents pour évaluer la dynamique d'un champ

Comme nous l'avons pu le constater dans le chapitre précédent, la vaste majorité des applications se basant sur la dynamique des champs neuronaux utilisent la capacité de ces modèles à générer des bulles d'excitation locale. Ce constat provient principalement du fait que dans un contexte pratique, un système fonctionnel (neuronal) quelconque traite un flot informationnel de nature externe au système et est influencé de manière déterminante par ce flot, perçu grâce à ses capacités perceptives. D'un autre point de vue, les travaux analytiques résumés dans le même chapitre montrent un intérêt d'étude de la dynamique des solutions homogènes des équations de champ, c'est-à-dire en absence de stimulation externe.

Par conséquent, nous reconsidérons dans ce chapitre la pertinence d'une étude approfondie de la dynamique des bulles d'excitation locale dans plusieurs contextes de stimulation. Bien qu'il soit impossible d'énumérer tous ces contextes, nous pouvons néanmoins mettre en évidence certains cas de stimulation qui résument de manière concise des contextes clé, rencontrés au cours d'évolution dynamique d'une application. Ces contextes définissent les *scénarios* de stimulation dans lesquels nous allons examiner en détail dans la section 5.3 la dynamique des champs neuronaux. Un tel scénario est défini par l'ensemble $\{i(x, t)\}_{x \in X, 0 \leq t < T}$, qui décrit l'évolution temporelle de la stimulation externe pour chaque unité de champ, pour un intervalle de temps fixé.

Dépendant de la nature de l'application où ils peuvent être rencontrés, nous distinguons plusieurs catégories de scénarios. Les paragraphes suivants présentent quelques-uns de ces scénarios, groupés selon la nature des applications qui s'y rapportent. Notons toutefois que tous ces scénarios sont indépendants un de l'autre, la classification que nous avons employée ici visant de mettre en évidence certaines propriétés dynamiques communes qui caractérisent généralement la dynamique des sources de stimulation externe des champs neuronaux. Basée sur ces observations, notre approche vise à traiter ensuite la problématique de la définition formelle d'un comportement attendu d'un champ neuronal, dans ces contextes de stimulation externe. Ces aspects sont abordés dans la section 5.2, où nous proposons également un instrument numérique d'évaluation du comportement d'un champ neuronal quelconque vis-à-vis d'un scénario arbitraire de stimulation. Grâce à cet outil d'évaluation, nous présentons dans la section 5.3 une analyse détaillée des comportements dynamiques exhibés par les modèles de champs neuronaux classiques dans chacun des scénarios que nous introduisons dans cette section.

Pour garder un niveau de description assez simple, nous illustrons ici seulement les scénarios de stimulation unidimensionnelle, notant cependant que les mêmes scénarios peuvent se généraliser pour le cas bidimensionnel.

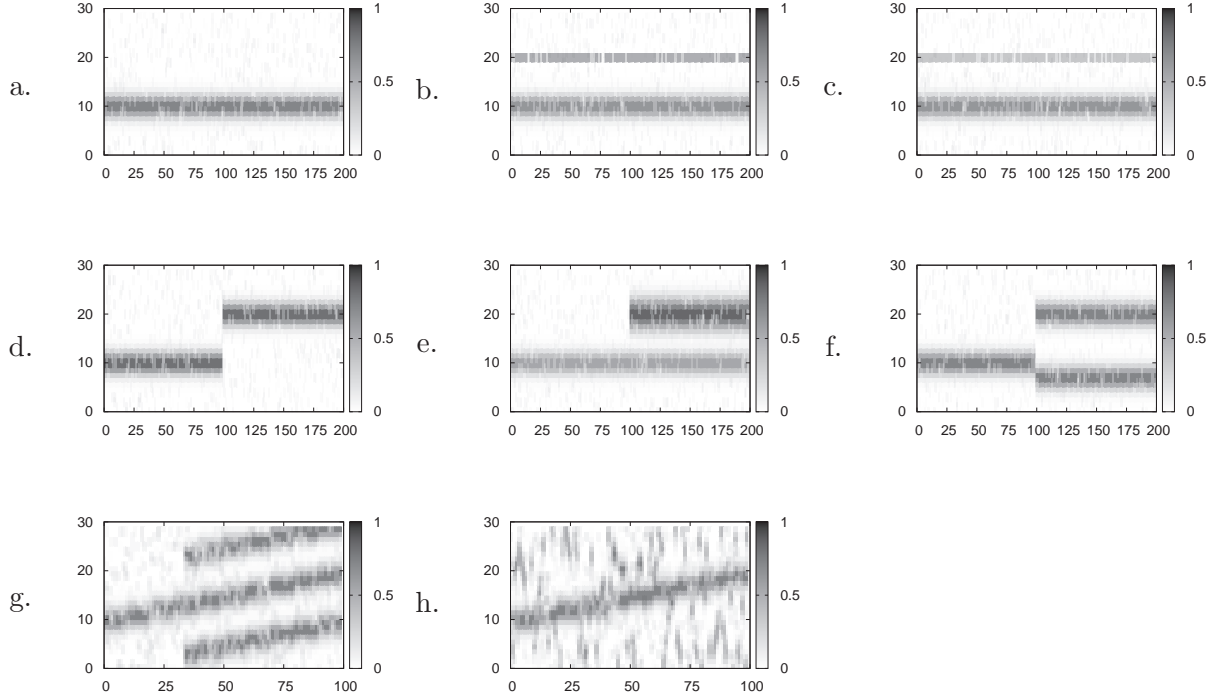


FIG. 5.1 – Plusieurs scénarios unidimensionnels, conçus pour examiner différents aspects dynamiques, souvent rencontrés dans la modélisation neuronale des processus d'attention. Pour chaque graphique, l'abscisse représente le pas de simulation t , alors que l'ordonnée représente les positions $x \in X$ du champ. L'intensité de stimulation $i(x, t)$ est représentée en niveaux de gris (le blanc symbolise une absence d'activité, et le noir symbolise le niveau maximal de cette stimulation). Description des scénarios de stimulation : a. stimulus gaussien individuel stationnaire ; b. stimulus gaussien stationnaire accompagné d'un stimulus distracteur voisin de la même amplitude ; c. stimulus gaussien stationnaire accompagné d'un stimulus distracteur voisin d'une amplitude inférieure ; d. stimulus gaussien individuel stationnaire, translaté avec 10 unités après 100 pas de simulation ; e. stimulus gaussien stationnaire accompagné, après 100 pas de simulation, d'un nouveau stimulus d'une amplitude et d'une variance supérieures ; f. stimulus gaussien stationnaire, qui après 100 pas de simulation, est translaté avec 5 unités et devient accompagné d'un nouveau stimulus de la même intensité ; g. stimulus gaussien qui suit une translation périodique de vitesse constante, et qui devient accompagné de deux stimulus gaussiens adjacents, du même type ; h. champ neuronal très bruité, avec un stimulus gaussien individuel qui suit une translation périodique de vitesse constante. Tous les scénarios présentés ici sont affectés par une source uniforme et aléatoire de bruit.

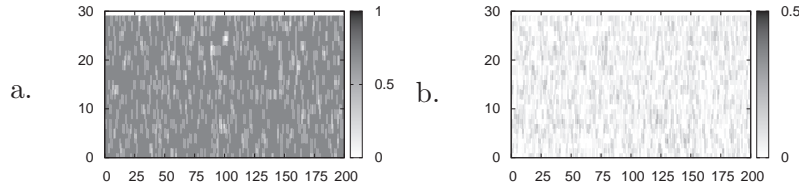


FIG. 5.2 – Scénarios d’exploration, conçus pour examiner la capacité des champs à générer des réponses sélectives, même dans des contextes où toutes les unités de champ sont stimulées de manière équiprobable, avec la même intensité. Les deux scénarios présentés ici sont affectés par la même source uniforme et aléatoire de bruit, mais ils diffèrent par l’amplitude des stimuli considérés.

5.1.1 Scénarios d’attention

Comme nous l’avons déjà remarqué dans le paragraphe 4.2.2, le paradigme des champs neuronaux dynamiques a été souvent utilisé dans la littérature du domaine des neurosciences pour modéliser différents aspects concernant les processus neuronaux, et notamment les mécanismes neuronaux d’attention. Pour cette raison, nous sommes intéressé dans notre analyse suivante à examiner plus en détail le comportement des champs neuronaux dans des scénarios appropriés pour les contextes de modélisation du processus de l’attention. La figure 5.1 présente une série de scénarios que nous considérons pertinents pour ce type d’applications.

Ces scénarios visent à mettre en évidence le comportement des mécanismes de compétition et de coopération neuronale modélisés par les champs neuronaux dans des scénarios d’attention, analysant la capacité de ces modèles de formuler des réponses sélectives et maintenir ou adapter ensuite leurs réponses, selon la dynamique induite par la distribution des stimuli d’entrée. Les scénarios considèrent les cas des distracteurs (fort (b) ou faible (c)), les cas de changements brusques survenus dans la distribution de stimuli (d-g), ou bien le cas des distributions très perturbées par une source de bruit uniforme (h).

5.1.2 Scénarios d’exploration

Par un scénario d’exploration nous entendons un scénario dans lequel soit toutes les unités du champ sont excitées par une stimulation externe de la même intensité, soit toutes ces unités sont stimulées par une source uniforme de bruit. La figure 5.2 présente deux exemples de scénarios d’exploration. Ces cas permettent d’étudier la manière dont un champ formule sa réponse en présence d’une stimulation uniforme. Autrement dit, ces scénarios sont appropriés pour examiner la capacité d’un champ à « explorer » une entrée équiprobable pour toutes ses unités, et à sélectionner ensuite un sous-ensemble de ces unités, formulant ainsi la réponse du champ. En conclusion, ces scénarios visent à étudier le caractère sélectif du champ, dans un contexte où quasiment toutes les unités du champ ressentent la même intensité de stimulation.

5.1.3 Scénarios auxiliaires

Finalement, la figure 5.3 illustre différents autres cas de stimulation neuronale qui permettent ainsi une étude étendue du comportement des champs. Ces scénarios visent à examiner la capacité du champ de choisir parmi plusieurs possibilités distinctes (a), la capacité d’utiliser un effet de

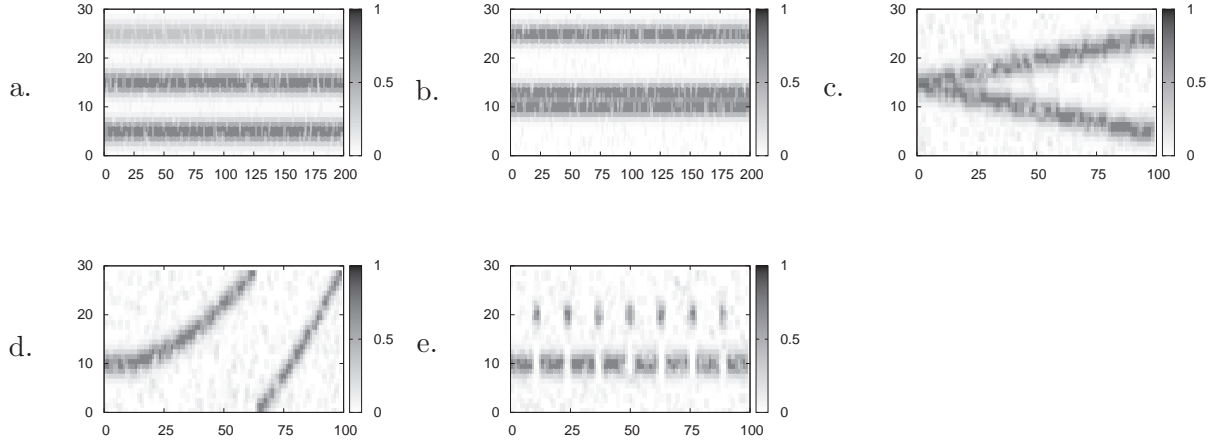


FIG. 5.3 – Plusieurs scénarios auxiliaires, mettant en évidence d’autres aspects dynamiques d’une distribution de stimulation neuronale. Description des scénarios de stimulation : a. trois stimuli gaussiens stationnaires de différentes amplitudes ; b. trois stimuli gaussiens stationnaires de la même amplitude, deux d’entre eux étant suffisamment proches pour induire un effet de coopération locale ; c. deux stimuli gaussiens qui partent d’une source commune et qui suivent une translation périodique de vitesse constante, dans deux directions opposées ; d. stimulus gaussien individuel qui suit une translation périodique accélérée ; e. deux stimuli gaussiens de la même amplitude et variance, qui apparaissent alternativement. Tous les scénarios présentés ici sont affectés par une source uniforme et aléatoire de bruit.

coopération locale dans le processus de compétition (b), ou la capacité à maintenir un choix, une fois que le processus de compétition en a généré un (c). Enfin, les scénarios (d,e) visent à examiner l’influence de la vitesse de déplacement ou changement d’un stimulus, par rapport à l’ajustement de la réponse du champ.

5.1.4 Scénario d’auto-organisation unidimensionnelle

Étant donné notre intérêt pour l’auto-organisation dans nos recherches, nous proposons ici un scénario conçu pour examiner la capacité de l’architecture décrite dans la section 2.6 à auto-organiser un ensemble unidimensionnel de prototypes, Ω . La figure 5.4 définit l’ensemble bidimensionnel $\Xi = \{\xi = (\xi_1, \xi_2) \in [0, 1] \times [0, 1] \mid 0.5 \leq \sqrt{\xi_1^2 + \xi_2^2} \leq 1\}$, représentant la distribution d’entrée qui est quantifiée par les prototypes $\Omega = \{\omega(x) = (\omega_1(x), \omega_2(x)) \in [0, 1] \times [0, 1] \mid x \in X\}$.

L’ensemble X des positions des unités du champ neuronal est un intervalle d’entiers $[1..n]$, et les connexions synaptiques entre ces unités décrivent un graphe cyclique. Nous avons décidé de choisir cette forme particulière pour X , afin que ses frontières n’influencent pas le processus d’auto-organisation (et donc elles n’induisent pas des effets de bord dans la dynamique du champ). Pour cette raison, l’ensemble Ξ est conçu en conséquence. Néanmoins, comme nous allons le constater dans la section des résultats expérimentaux (voir principalement dans 5.3.4), cette topologie de l’espace Ξ joue aussi un rôle important dans l’évolution du processus d’auto-organisation.

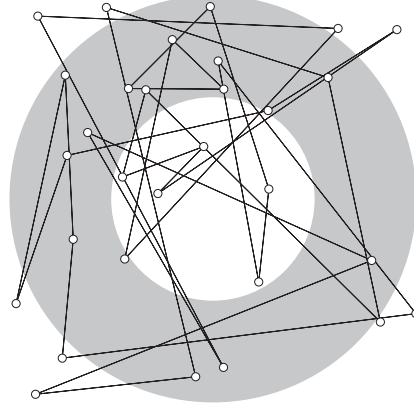


FIG. 5.4 – L'ensemble $\Xi = \{\xi = (\xi_1, \xi_2) \in [0, 1] \times [0, 1] \mid 0.5 \leq \sqrt{\xi_1^2 + \xi_2^2} \leq 1\}$, illustré par l'anneau gris dans la figure, définit la distribution bidimensionnelle considérée pour résoudre la tâche d'auto-organisation unidimensionnelle. L'ensemble Ω de prototypes à quantifier Ξ sont figurés aussi (par les petits ronds et leurs connexions latérales) avec leurs valeurs initiales aléatoires, définies au début du scénario.

Ainsi, à chaque présentation d'un nouvel échantillon $\xi \in \Xi$ à l'entrée de l'architecture neuronale d'auto-organisation, chaque unité x de la couche de reconnaissance répond avec une activité neuronale d'intensité $i(x)$, proportionnelle à la distance relative entre le prototype $\omega(x)$ et l'échantillon ξ :

$$i(x, t) = f_1(\omega(x), \xi) = e^{-\alpha_0 d^2(\omega(x), \xi(t))} \quad (5.1)$$

où d représente la distance euclidienne dans l'espace bidimensionnel \mathbb{R}^2 , et α_0 est une constante scalaire positive ($\alpha_0 = 5$ dans nos simulations). Notons que dans ce scénario, la stimulation externe $\{i(x, t) \mid x \in X\}$ de la couche corticale de l'architecture neuronale d'auto-organisation n'est pas imposée arbitrairement en tout x et à chaque pas de temps, comme dans les scénarios précédents, mais elle est calculée selon l'expression décrite ci-dessus.

5.1.5 Scénario d'auto-organisation bidimensionnelle

Afin d'examiner le comportement et la performance de l'architecture neuronale d'auto-organisation dans un contexte pratique plus élaboré et approprié à notre thématique de recherche, nous proposons ici un scénario qui modélise le processus de développement des cartes d'orientation visuelle, similaires à celles observées expérimentalement dans la structure neuronale du cortex visuel primaire.

Ainsi, dans ce nouveau scénario, les deux couches de l'architecture d'auto-organisation considèrent un champ X bidimensionnel. La distribution d'entrée Ξ est définie par l'ensemble continu d'angles d'orientation $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Chaque prototype de la couche de reconnaissance développe une sensibilité d'orientation préférentielle, relative à une orientation quelconque $\xi(t)$ d'un nouvel échantillon de Ξ :

$$i(x, t) = f_2(\omega(x, t), \xi(t)) = \begin{cases} \cos(|\omega(x, t) - \xi(t)|) & \text{si } |\omega(x, t) - \xi| \leq \frac{\pi}{2} \\ \cos(\pi - |\omega(x, t) - \xi(t)|) & \text{sinon} \end{cases} \quad (5.2)$$

Afin d'analyser l'influence de la forme de la fonction de reconnaissance dans le fonctionnement global du processus d'auto-organisation, nous considérons également le cas d'une deuxième

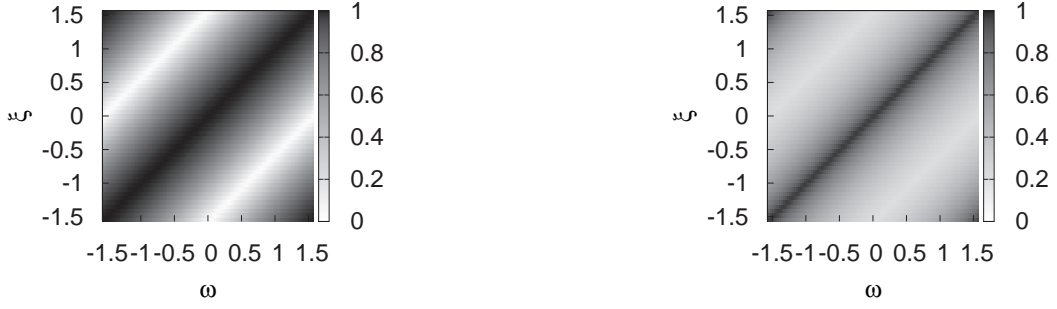


FIG. 5.5 – Les graphiques des fonctions de reconnaissance $f_2(\omega, \xi)$ (figure gauche) et $f_3(\omega, \xi)$ (figure droite), définies par les expressions 5.2 et 5.3, et utilisées dans les scénarios d'auto-organisation bidimensionnelle.

fonction de reconnaissance, qui définit la sensibilité d'orientation des prototypes selon l'expression :

$$i(x, t) = f_3(\omega(x, t), \xi(t)) = \begin{cases} e^{-|\omega(x, t) - \xi|} & \text{si } |\omega(x, t) - \xi| \leq \frac{\pi}{2} \\ e^{-\pi + |\omega(x, t) - \xi|} & \text{sinon} \end{cases} \quad (5.3)$$

Pour comparaison, les deux fonctions $f_2(\omega, \xi)$ et $f_3(\omega, \xi)$ sont illustrées sur la figure 5.5, pour $\omega, \xi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

5.2 Évaluation des capacités dynamiques d'un champ neuronal

Dans cette section nous présentons en détail une méthode d'analyse dynamique, capable de qualifier le comportement des champs neuronaux dans les différents scénarios de stimulation que nous avons introduits précédemment.

Cette approche s'étaie sur le fait qu'un modèle est considéré approprié pour un contexte pratique donné si son utilisation conduit à une certaine performance pour le système qui l'emploie. Ceci implique souvent l'existence d'une certaine attente du concepteur vis-à-vis des capacités fonctionnelles du modèle en question. Le paragraphe 5.2.1 décrit plusieurs comportements des champs neuronaux, attendus dans certains scénarios-clé. Les raisons pour lesquelles ces comportements peuvent être considérés comme souhaitables sont discutés dans les paragraphes 5.2.1 et 5.2.4, le dernier fournissant une forte justification réclamée par l'utilisation pratique des champs dans des scénarios d'auto-organisation. Les propriétés communes de ces comportements attendus sont ensuite extraites et abstraites dans 5.2.1, sous la forme de deux critères de performance. Cette formulation permet ensuite d'introduire dans 5.2.2 un instrument de mesure capable d'évaluer numériquement la capacité d'un champ neuronal à satisfaire ces critères. Le paragraphe suivant définit la notion de champ de référence, afin que le dernier paragraphe de cette section puisse décrire notre plate-forme d'évaluation empirique de la dynamique des modèles de champs neuronaux, utilisant les outils d'analyse présentés précédemment.

5.2.1 Propriétés dynamiques souhaitables et critères d'optimalité

Si dans les paragraphes précédents nous avons énuméré plusieurs cas d'utilisation pratique des champs neuronaux, nous analysons dans la suite ce que nous attendons de la dynamique

d'un champ dans ces contextes de modélisation.

Mikhailova et Goerick abordent dans [85] une problématique pragmatique, visant la réponse souhaitable d'un champ neuronal, afin d'utiliser celui-ci comme mécanisme de sélection d'activation neuronale. Inspirés par leurs travaux, nous proposons dans la figure 5.6 une liste étendue de comportements souhaitables d'un champ neuronal, qui correspondent aux réponses désirées d'un champ dans plusieurs scénarios-clé. Pour des raisons de simplicité visuelle, les cas illustrés dans cette figure se réfèrent exclusivement aux scénarios de stimulation unidimensionnelle. Toutefois, nous pouvons facilement imaginer les réponses des champs bidimensionnels correspondant aux scénarios représentés ici.

Toujours pour simplifier la compréhension de ces comportements attendus, nous présentons dans cette figure exclusivement des réponses de champ caractérisées par une unique bulle d'excitation. Nous justifions ce choix par le fait que ce qui nous intéresse dans ces contextes est la manière dont la stimulation externe influence la dynamique d'une bulle d'excitation (et son voisinage) dans le champ. Autrement dit, les cas présentés dans la figure 5.6 se focalisent sur la « zone d'influence » d'une seule bulle d'excitation de l'activité du champ. Toutefois, comme nous allons le constater prochainement, notre analyse ne se restreint pas à ce cas particulier.

Analysant la dynamique désirée des champs, présentée graphiquement dans la figure 5.6, nous pouvons déduire les arguments qui justifient de considérer comme souhaitables tels comportements. Tous les cas énumérés ici (et notamment les cas (a-e)) illustrent le principe fondateur des champs neuronaux, c'est-à-dire celui de la compétition neuronale. En effet, le champ est supposé s'activer le plus fortement (générer une bulle d'excitation) aux endroits correspondant aux sources de stimulation les plus significatives, ce qui lui donne un caractère de sélectivité (a). Si plusieurs sources sont d'amplitude comparable, les unités du champ doivent entrer en compétition afin de n'en retenir qu'une seule (b, c). Il est important de remarquer que, afin de « choisir » l'endroit où s'activer, le champ doit prendre en compte la contribution cumulée locale de la stimulation externe, et pas seulement la stimulation individuelle de chaque unité. Ainsi, des distracteurs comme ceux présents dans les cas (d,e) ne devraient pas influencer de manière décisive la réponse du champ. Plus spécifiquement, l'effet d'une collaboration entre les unités voisines est illustré dans (f). En résumé, nous pouvons affirmer que ces comportements s'inscrivent dans une dynamique de collaboration locale et compétition globale. Nous ajoutons à ces propriétés la nécessité du champ de reformuler sa réponse après chaque changement de configuration de stimulation, afin de rester cohérente avec les principes énoncés ci-dessus (voir le cas (g)). Finalement, nous imposons qu'un champ soit capable de maintenir son caractère sélectif même en absence de stimulation (voir le cas (h), similaire au cas (c) d'exploration). Même si ce comportement pourrait sembler inapproprié, il est tout à fait en accord avec les propriétés dynamiques mises en évidence jusqu'ici. Le simple fait que le champ soit stimulé avec une faible (voir nulle) source de stimulation ne devrait pas empêcher la sélectivité du champ. Autrement dit, le processus de compétition neuronale ne devrait pas être influencé par le niveau de stimulation. De plus, comme nous allons constater dans le paragraphe 5.2.4, cette propriété est essentielle pour la résolution des scénarios d'auto-organisation. Ce comportement nous conduit à souligner également un dernier point. Le fait que l'activation du champ ne doit pas être influencée par le niveau absolu de la stimulation externe nous fait considérer assez naturellement que l'amplitude de son état stable ne doit pas non plus dépendre de ce niveau de stimulation. Par conséquent, afin de simplifier l'analyse ultérieure, nous considérons ici seulement des réponses de type bulle d'excitation forte (atteignant le niveau de saturation).

Les raisons présentées ici nous invitent à considérer ces comportements dynamiques comme souhaitables. Même si cette liste n'est pas exhaustive, car bien d'autres comportements dynamiques peuvent être envisagés pour un champ, nous considérons néanmoins que ce spectre

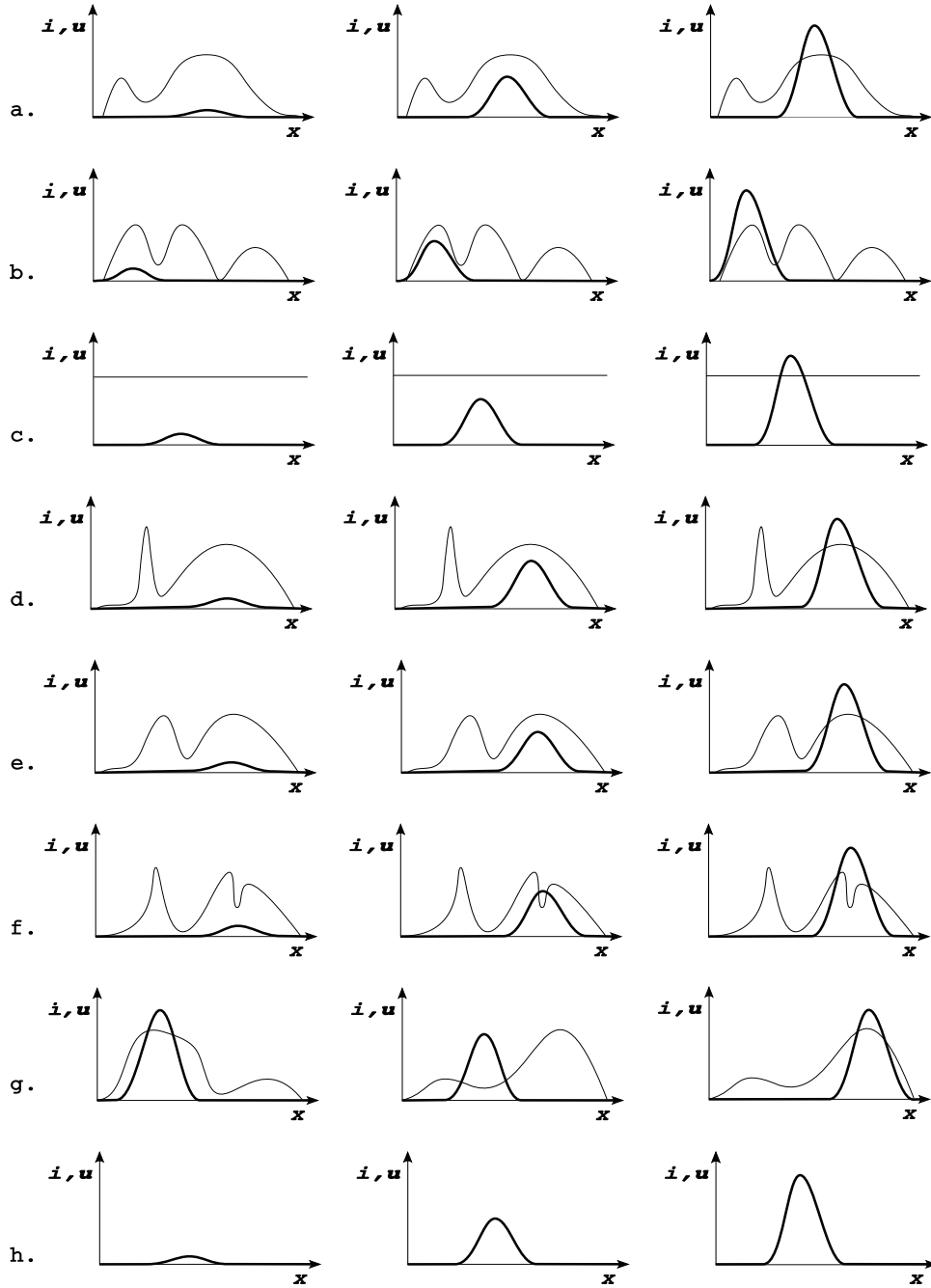


FIG. 5.6 – Les réponses attendues (figurées par les lignes en gras) d'un champ neuronal unidimensionnel qui est soumis aux différents scénarios de stimulation externe (figurés par les lignes fines). Les courbes présentées ici ont le but d'illustrer de manière qualitative plusieurs comportements souhaitables des champs neuronaux. Elles ne font pas l'objet d'une simulation numérique authentique. Chaque rang (a-h) décrit un nouveau scénario, et l'évolution temporelle de l'activation désirée du champ est résumée de gauche à droite par trois états (début, intermédiaire et final). Le champ est supposé générer une seule bulle d'activité neuronale. a. sélectivité, b. compétition, c. exploration (plateau haut), d. compétition (distracteur fort), e. compétition (distracteur faible), f. collaboration, g. adaptation, h. exploration (plateau bas).

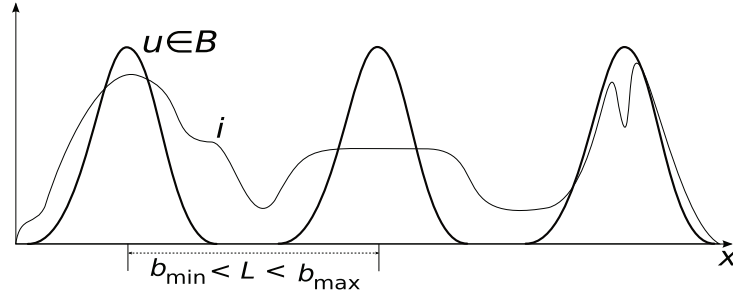


FIG. 5.7 – Un exemple de réponse stabilisée d'un champ neuronal unidimensionnel, qui satisfait les propriétés $(P \wedge Q)$.

comportemental mérite d'être examiné en détail, étant donné sa pertinence vis-à-vis des applications pratiques.

Synthétisant tous les comportements présentés dans la figure 5.6, et les extrapolant pour un champ capable de générer simultanément plusieurs bulles d'excitation, nous pouvons dériver deux propriétés dynamiques générales, valables pour tous ces comportements désirés.

Propriété P. *Pour la réponse stabilisée d'un champ neuronal, les bulles d'activité neuronale se situent dans des endroits où la stimulation externe est localement la plus forte.*

Propriété Q. *Pour la réponse stabilisée d'un champ neuronal, l'extension spatiale et l'amplitude des bulles d'activité neuronale ne dépendent pas du niveau de la stimulation externe i , et la distance entre deux bulles voisines doit être contenue entre deux bornes limites (b_{min} et b_{max} , avec $b_{min} < b_{max}$, comme illustré dans la figure 5.7), afin que les bulles ne soient ni trop éloignées, ni trop proches une de l'autre.*

La propriété (P) définit le caractère sélectif d'un champ, alors que la propriété (Q) définit une contrainte de forme et de densité des bulles d'excitation dans le champ, ce qui dicte la portée de l'influence d'une bulle d'excitation dans le champ.

Désormais, nous sommes intéressé à découvrir les champs neuronaux qui satisfont les deux propriétés énoncées ici, $(P \wedge Q)$. Par conséquent, dans le paragraphe suivant nous proposons un instrument de mesure capable d'évaluer la capacité d'un champ à satisfaire ces deux critères de performance, à travers son évolution temporelle.

5.2.2 Instrument d'évaluation

Afin de pouvoir évaluer numériquement la capacité d'un champ neuronal à remplir les conditions décrites par les propriétés (P) et (Q), nous devons d'abord traduire l'énoncé de ces propriétés sous une forme analytique.

Fixons t , l'instant de temps pour lequel nous conduisons notre analyse de mesure. Comme t est choisi arbitrairement, afin de simplifier les notations, nous omettons dans la suite de l'analyse l'indice temporel. Soit $\{i(x)\}_{x \in X}$ la distribution d'entrée de la stimulation du champ, et $\{u(x)\}_{x \in X}$ la réponse des unités du champ, enregistrées à ce moment de temps fixé. Notons \mathbb{R}^X l'ensemble des fonctions de X à valeurs dans \mathbb{R} , ensemble qui contient également l'entrée et la réponse du champ. Sur cet ensemble, considérons ensuite la fonction de distance euclidienne d :

$$a \in \mathbb{R}^X, b \in \mathbb{R}^X, d(a, b) = \sqrt{\int_{x \in X} (a(x) - b(x))^2 dx} \quad (5.4)$$

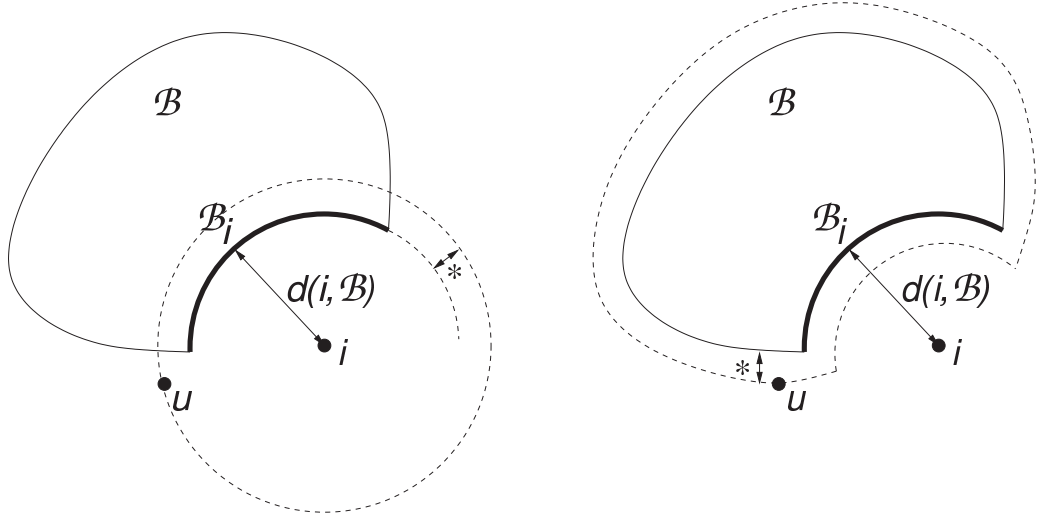


FIG. 5.8 – Interprétation géométrique du coefficient $\Delta_i^B(u)$, l'erreur résiduelle d'une réponse u de champ à satisfaire $(P \wedge Q)$, dans un scénario i donné. La courbe dessinée en gras sur les deux figures représente l'ensemble \mathcal{B}_i des réponses satisfaisant pleinement les propriétés $(P \wedge Q)$ par rapport à i . La réponse u du champ est satisfaisante vis-à-vis de $(P \wedge Q)$ si les écarts marqués par le symbole $*$ dans les deux figures : $(d(i, u) - d(i, \mathcal{B}))$ (sur la figure à gauche) et $d(u, \mathcal{B})$ (sur la figure à droite), sont petits.

Afin de faciliter la compréhension de notre approche analytique, la présentation de la méthode d'évaluation des champs est accompagnée dans la figure 5.8 de son interprétation géométrique.

Considérons une distribution quelconque u' d'activités neuronales. Si elle est caractérisée exclusivement par des bulles d'excitation d'amplitude maximale et de même extension spatiale, soit $\theta_{u'}$ l'ensemble des positions des centres de ces bulles dans l'espace X . Si la distance entre deux bulles voisines k et l , centrées aux positions $x_k, x_l \in \theta_{u'}$, respecte les inégalités :

$$\forall k \neq l, b_{min} \leq \|x_k - x_l\| \leq b_{max} \quad (5.5)$$

où $\|\cdot\|$ désigne la métrique (euclidienne) de l'espace X , alors nous pouvons affirmer que la distribution u' satisfait la propriété (Q). Nous notons que, afin de permettre l'émergence d'une unique bulle d'activité dans le champ, il suffit de considérer une valeur suffisamment grande pour le paramètre b_{min} ($b_{min} \rightarrow \infty$). Considérons ensuite $\mathcal{B} \subset \mathbb{R}^X$ (figuré aussi sur la figure 5.8), l'ensemble de toutes les réponses possibles d'un champ (composées seulement par des bulles d'excitation), qui satisfont correctement la condition (Q). Nous déduisons immédiatement que, afin de satisfaire (Q), u doit appartenir à l'ensemble \mathcal{B} , ce qui équivaut à :

$$d(u, \mathcal{B}) = \min_{u' \in \mathcal{B}} d(u, u') = 0 \quad (5.6)$$

D'autre part, la propriété (P) affirme que la formation des bulles d'activité dans le champ est réalisée par les zones de forte stimulation locale du champ. Ainsi, la réponse u doit être corrélée avec la stimulation i du champ, afin de considérer qu'elle satisfait la propriété (P). Bien que $d(u, i) = 0$ assurerait la parfaite corrélation entre u et i , cette condition imposerait que u soit une « simple copie » de la distribution i . Toutefois, cette condition devient incompatible avec

la condition (Q), qui demande $u \in \mathcal{B}$ (sauf pour le cas évident où i aussi est un élément de l'ensemble \mathcal{B}). Par conséquent, nous convenons de considérer plutôt :

$$d(i, u) - d(i, \mathcal{B}) = 0 \quad (5.7)$$

comme condition de satisfaction de la propriété (P).

Pour mieux comprendre le sens de cette condition, introduisons :

$$\mathcal{B}_i = \{u \in \mathcal{B} \mid d(i, u) = \min_{u' \in \mathcal{B}} \{d(i, u')\}\} \quad (5.8)$$

le sous-ensemble des éléments de \mathcal{B} (respectant donc (Q)), qui sont les plus proches de la distribution i (qui sont donc les plus corrélés avec i). Cet ensemble est mis en évidence par la ligne en gras sur les deux graphiques de la figure 5.8. On pourrait chercher les réponses respectant $(P \wedge Q)$ parmi les éléments de cet ensemble \mathcal{B}_i . Ceci revient à imposer comme condition nécessaire pour une distribution quelconque u :

$$d(u, \mathcal{B}_i) = 0 \quad (5.9)$$

afin qu'elle soit considérée satisfaisante vis-à-vis des deux propriétés $(P \wedge Q)$. Toutefois, d'un autre point de vue, nous pouvons considérer également que les distributions u convenables sont celles qui réussissent à être corrélées avec i de la même manière que les éléments de \mathcal{B}_i sont corrélés avec i . En effet, cette affirmation se traduit analytiquement sous la forme imposée par la condition 5.7.

En résumé, nous considérons que les conditions 5.7 et 5.6 décrivent correctement les critères souhaités par les propriétés $(P \wedge Q)$. Afin d'unifier ces deux conditions, définissons :

$$\Delta_i^{\mathcal{B}}(u) = \sqrt{(d(i, u) - d(i, \mathcal{B}))^2 + d^2(u, \mathcal{B})} \quad (5.10)$$

l'erreur résiduelle de la satisfaction de $(P \wedge Q)$ par une distribution u quelconque, en cas de stimulation i fixée. Définissons également la version normalisée de ce coefficient :

$$\bar{\Delta}_i^{\mathcal{B}}(u) = \Delta_i^{\mathcal{B}}(u) / d(0, \mathcal{B}) \quad (5.11)$$

Ces notations nous permettent de démontrer l'affirmation :

$$\bar{\Delta}_i^{\mathcal{B}}(u) = 0 \Leftrightarrow \Delta_i^{\mathcal{B}}(u) = 0 \Leftrightarrow d(u, \mathcal{B}_i) = 0 \quad (5.12)$$

La première partie de cette affirmation étant triviale, nous nous concentrons sur la démonstration de l'équivalence entre le deuxième et le troisième terme de cette proposition. Ainsi, $\Delta_i^{\mathcal{B}}(u) = 0$ implique d'une part $d(u, \mathcal{B}) = 0$, donc $u \in \mathcal{B}$, et d'une autre part $d(u, i) = d(i, \mathcal{B}) = d(i, \mathcal{B}_i)$, ce qui implique que u est un élément de \mathcal{B} qui minimise sa distance à i . Par conséquent, $u \in \mathcal{B}_i$, et donc $d(u, \mathcal{B}_i) = 0$. Réciproquement, si $d(u, \mathcal{B}_i) = 0$, alors $u \in \mathcal{B}_i \subseteq \mathcal{B}$, donc $d(u, \mathcal{B}) = 0$. De la même condition, comme $u \in \mathcal{B}_i$, nous déduisons immédiatement que $d(u, i) = d(i, \mathcal{B})$. Ces deux résultats confirment que $\Delta_i^{\mathcal{B}}(u) = 0$, ce qui finit de prouver l'équivalence entre le critère imposé par 5.9 et celui imposé par la valeur $\Delta_i^{\mathcal{B}}(u)$ nulle.

Afin de mieux évaluer la capacité d'un champ à respecter $(P \wedge Q)$, il est souvent nécessaire d'estimer aussi le degré de satisfaction de ces propriétés par la réponse du champ. Pour ceci, nous pouvons nous servir de l'implication :

$$d(u, \mathcal{B}_i) \leq \epsilon \Rightarrow \Delta_i^{\mathcal{B}}(u) \leq \epsilon \quad (5.13)$$

qui affirme que si $(P \wedge Q)$ sont satisfaites avec une marge d'erreur ϵ , alors la valeur $\Delta_i^{\mathcal{B}}(u)$ ne peut elle non plus dépasser ce seuil ϵ . En effet,

$$\begin{aligned} \mathcal{B}_i \subseteq \mathcal{B} &\Rightarrow d(u, \mathcal{B}) \leq d(u, \mathcal{B}_i) \leq \epsilon \\ \forall u' \in \mathcal{B}, |d(i, u) - d(i, \mathcal{B})| &= |d(i, u) - d(i, u')| \leq d(u, u') \Rightarrow |d(i, u) - d(i, u')| \leq d(u, \mathcal{B}_i) \leq \epsilon \end{aligned}$$

Toutefois, l'affirmation réciproque de 5.13 n'est plus valable cette fois-ci. Si le coefficient $\Delta_i^{\mathcal{B}}(u)$ est petit, ceci implique d'une part que u est proche de l'ensemble \mathcal{B} des réponses qui respectent (Q) (donc u minimise la distance $d(u, \mathcal{B})$ figurée par le symbole $*$ dans la partie gauche de la figure 5.8). D'autre part, la corrélation entre les deux distributions i et u , qui vise à minimiser la distance $d(i, u)$, est approximativement la même que la corrélation entre i et les éléments de \mathcal{B} les plus proches de i , qui est numériquement exprimée par $d(i, \mathcal{B})$. Ce deuxième aspect implique que u doit minimiser l'écart entre $d(i, u)$ et $d(i, \mathcal{B})$ (figuré aussi par le symbole $*$ dans la partie droite de la figure 5.8). Bien que ces conditions induites par $\Delta_i^{\mathcal{B}}(u) \leq \epsilon$ n'impliquent pas nécessairement une distance réduite entre u et l'ensemble \mathcal{B}_i des réponses « idéales », nous avons choisi de relaxer la contrainte $d(u, \mathcal{B}_i) \leq \epsilon$ par la condition $\Delta_i^{\mathcal{B}}(u) \leq \epsilon$. Ainsi, nous allons considérer toujours satisfaisantes toutes les réponses u de champ qui sont corrélées avec la distribution d'entrée i d'une manière similaire à ce que le sont les éléments de \mathcal{B}_i (qui respectent pleinement $(P \wedge Q)$).

Pour résumer, dans cette section nous avons présenté un critère analytique pour définir la capacité d'un champ neuronal à répondre à une stimulation externe i , respectant les propriétés souhaitables (P) et (Q) introduites dans le paragraphe précédent. Dans la suite de nos démarches, nous sommes intéressé à trouver des champs neuronaux capables de produire une valeur proche de zéro pour l'erreur résiduelle $\Delta_i^{\mathcal{B}}(u)$. Cet objectif est traité dans la section 5.2.5 de ce chapitre, qui présente en détail une méthodologie d'évaluation des champs neuronaux, basée sur l'instrument de mesure présenté ici.

5.2.3 Implantation de l'instrument d'évaluation. Champs de référence

L'expression de l'instrument d'évaluation des champs décrit au paragraphe précédent fait appel plusieurs fois au calcul d'une distance entre une fonction (la distribution u ou i) et un ensemble de fonctions (\mathcal{B} ou \mathcal{B}_i). Si cette approche est appropriée pour exprimer quantitativement, de manière simple et intuitive, la capacité d'un champ à satisfaire $(P \wedge Q)$, sa mise en pratique n'est pas tout à fait directe. Cette section décrit une implantation générique pour réaliser ce calcul numérique.

Soit u une distribution quelconque d'activités neuronales. Conformément à la définition 5.6, pour u donné, calculer $d(u, \mathcal{B})$ revient à trouver une distribution $u' \in \mathcal{B}$ qui minimise la distance $d(u, u')$. Afin de concevoir une procédure de minimisation pour cette distance, il est nécessaire de décrire l'ensemble continu \mathcal{B} sous une forme paramétrique.

Réexaminant la propriété (Q) des réponses attendues d'un champ, nous constatons qu'elle impose deux contraintes. La première se réfère à la forme des bulles d'excitation : leurs extensions spatiales et amplitudes sont indépendantes de la stimulation externe i . Ceci implique que la forme des bulles doit être une caractéristique intrinsèque au champ (en effet, celle-ci est souvent dictée exclusivement par le noyau des poids des connexions latérales, utilisé par le modèle de champ).

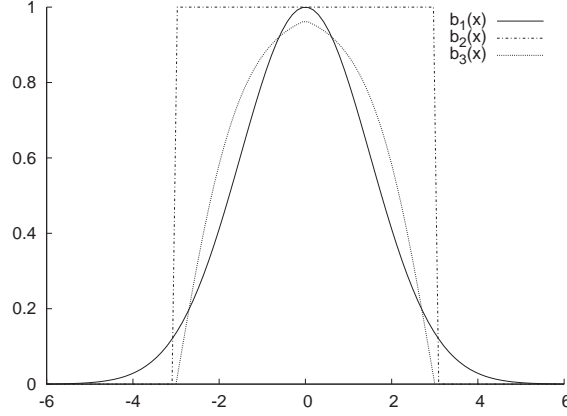


FIG. 5.9 – Les formes des bulles d'excitation définies par les expressions 5.14-5.16, qui pourraient caractériser une réponse stabilisée d'un champ neuronal quelconque (ici figurées avec un rayon $R = 3$).

Différentes formes de bulles d'excitation peuvent être envisagées, comme par exemple :

$$b_1(x, x_c) = e^{\frac{-2\|x-x_c\|^2}{R^2}} \quad (5.14)$$

$$b_2(x, x_c) = \begin{cases} 1 & \text{si } \|x - x_c\| \leq R \\ 0 & \text{sinon} \end{cases} \quad (5.15)$$

$$b_3(x, x_c) = \begin{cases} \frac{e^{4(1-\frac{\|x-x_c\|}{R})}-1}{e^{4(1-\frac{\|x-x_c\|}{R})}+1} & \text{si } \|x - x_c\| \leq R \\ 0 & \text{sinon} \end{cases} \quad (5.16)$$

Les fonctions $b_{1..3}$, également présentées de manière graphique dans la figure 5.9, désignent chacune la forme d'une bulle centrée en x_c (une position arbitraire du champ), R étant le rayon souhaité pour la bulle, et $\|\cdot\|$ la métrique (euclidienne) utilisée sur l'espace X . Dépendant du noyau des poids des connexions latérales, la forme des bulles générées par un champ neuronal peut ressembler à une des formes décrites par les fonctions $b_{1..3}$. Par exemple, la fonction b_3 décrit le plus fidèlement la forme des bulles générées par les modèles de champ que nous examinons dans nos expériences. Par conséquent, dans le cas particulier des expériences numériques ultérieures, nous considérons b_3 comme étant la forme attendue pour une bulle d'activité d'un champ neuronal.

Suite à cette première conclusion concernant (Q), nous pouvons affirmer que chaque élément $u' \in \mathcal{B}$ peut s'exprimer sous une forme générale de type :

$$u'(x) = \sum_{x_j \in \theta'} b_3(x, x_j) \quad (5.17)$$

où θ' représente une configuration pour les positions $x_j \in X$ des centres des bulles d'activité neuronale, générées par le champ. Afin de satisfaire pleinement (Q), cette expression doit respecter la contrainte de densité des bulles stables qui peuvent coexister simultanément dans le champ. Conformément à la condition 5.5, la distance $\|x_k - x_l\|$ entre les centres de deux bulles voisines k et l doit être contenue entre deux valeurs, b_{min} et b_{max} . Si nous considérons Θ comme étant l'ensemble de toutes les configurations possibles θ' (respectant donc la propriété (Q)), nous

```

00.  $d_{min} = \infty$  ;  $Sol \leftarrow \phi$ .
01. soit  $\{\delta_1, \dots, \delta_n\}$  une base canonique de  $\mathbb{R}^n$ , où  $n = \dim(X)$ .
02. générer toutes les configurations discrétisées possibles
     $\hat{\Theta} = \cup \{\hat{\theta}'\}$ , qui valident 5.5.
03. pour chaque  $\hat{\theta}' \in \hat{\Theta}$ 
    04. calculer  $u'(x) = \sum_{x_j \in \hat{\theta}'} b_3(x, x_j)$ .
    05.  $k = 0$ .
    06. pour  $(k < 2 * n * \text{card}(\hat{\theta}') + 1)$ 
        07. avec  $x_j = \text{Rand}(\hat{\theta}')$ ,  $\delta_l = \text{Rand}(\{\delta_1, \dots, \delta_n\})$  et  $\alpha =$ 
             $\text{Rand}(\{-1, +1\})$ ,
            chacun choisi aléatoirement,
            soit  $\hat{\theta}'' = \hat{\theta}' \setminus \{x_j\} \cup \{x_j + \alpha \delta_l\}$ .
        08. si  $(\hat{\theta}'' \text{ valide 5.5})$  alors
            09. calculer  $u''(x) = \sum_{x_j \in \hat{\theta}''} b_3(x, x_j)$ 
            10. si  $(d(u, u'') < d(u, u'))$  alors
                11.  $\hat{\theta}' \leftarrow \hat{\theta}''$  ;  $u' \leftarrow u''$ 
                12. revenir au pas 05.
    13. si  $(d(u, u') < d_{min})$  alors
        14.  $Sol \leftarrow \hat{\theta}'$ .
    
```

FIG. 5.10 – Procédure de calcul de la distance $d(u, \mathcal{B})$ par descente du gradient. Notons que $r = \text{Rand}(S)$ désigne un élément r choisi aléatoirement de l'ensemble S . Suivre le texte pour plus de détail.

pouvons affirmer que l'ensemble \mathcal{B} est décrit de manière bijective par l'ensemble Θ . Dans ces conditions, calculer la distance $d(u, \mathcal{B})$ se traduirait à chercher une configuration θ' de positions valides (en accord avec la propriété de densité) de bulles d'activité, qui minimise la distance $d(u, u')$.

L'analyse déroulée jusqu'ici nous permet de décrire une procédure de recherche combinatoire pour calculer la distance $d(u, \mathcal{B})$, basée sur un algorithme de descente de gradient. Cette procédure est décrite sous forme de pseudo-code dans la figure 5.10, et expliquée en détail dans la suite. Notons que cette méthode ne représente pas une procédure de simulation numérique du comportement dynamique d'un champ neuronal, mais elle permet d'examiner et de quantifier certaines propriétés de son comportement. Par conséquent, la procédure présentée ci-dessous ne fournit pas une alternative à l'implantation d'un champ neuronal, elle étant un instrument complémentaire d'analyse dynamique du champ.

Tout d'abord, l'ensemble Θ est discrétisé sous la forme $\hat{\Theta}$. Ensuite, plusieurs configurations initiales de type $\hat{\theta}'$, qui définissent les endroits possibles $x_j \in X$ des bulles d'activité dans le champ, sont générées de manière combinatoire (similaire au problème de placement de N reines sur un échiquier). Les deux inégalités imposées par 5.5 réduisent de manière significative le nombre de variantes possibles $\hat{\theta}'$. Néanmoins, leur génération combinatoire suppose l'implantation d'une procédure non-distribuée et très coûteuse du point de vue complexité de calcul.

Pour chaque configuration initiale $\hat{\theta}'$, la réponse u' du champ est calculée selon l'expression 5.17. Ensuite, une nouvelle configuration $\hat{\theta}''$ est générée à partir de celle-ci. Pour la générer, un déplacement spatial de l'ordre du pas de discrétisation est induit, seulement pour une position x_j , choisie aléatoirement de la configuration $\hat{\theta}'$. Si cette nouvelle configuration respecte

toujours la condition 5.5 de la densité de bulles d'activité neuronale, u'' est calculé de manière similaire à u' , utilisant donc la même expression 5.17. Si, suite à ce calcul, $d(u, u'') < d(u, u')$, la configuration $\hat{\theta}'$ est remplacée par la nouvelle $\hat{\theta}''$, et la descente du gradient est continuée. Si au contraire, aucun ajustement local de la configuration $\hat{\theta}'$ ne minimise $d(u, u')$, la procédure d'optimisation s'arrête, trouvant un minimum local. La descente stochastique (c'est-à-dire par ajustement aléatoire des positions x_j) augmente considérablement les chances de ne pas être « attiré » trop vite vers un minimum local, au détriment d'un autre minimum voisin, encore plus petit. Nous notons également que l'inégalité stricte présente au pas 10 de l'algorithme de descente du gradient peut être considérée non-strictes si, dans l'implantation de l'algorithme, on tient compte aussi des positions déjà visitées pendant le processus d'optimisation. Cet aspect permet une descente du gradient dans les régions de plateau, améliorant radicalement la recherche du minimum local (nous avons choisi de ne pas inclure cet aspect directement dans l'algorithme présenté dans la figure 5.10, afin d'illustrer exclusivement le principe de calcul, et non pas ses détails d'implantation).

Quand la procédure d'optimisation locale de chaque configuration initiale $\hat{\theta}'$ est achevée, une comparaison avec la solution provisoire, trouvée jusqu'à ce pas de l'algorithme, est réalisée. Si cette nouvelle configuration minimise à son tour la distance $d(u, u')$, elle est retenue comme solution potentielle. A la fin de la boucle 03, toutes les configurations initiales auront été examinées, et *Sol* désignera une des configurations u' pour lesquelles $d(u, u') = d(u, \mathcal{B})$.

Soit i une distribution de stimuli externes donnée. Une distribution $u' \in \mathcal{B}$ minimisant la distance $d(i, \mathcal{B})$, calculée par la procédure décrite ici, satisfait pleinement (par définition) les deux propriétés ($P \wedge Q$). Pour cette raison, appelons u' la *réponse de référence* pour l'activité d'un champ neuronal (cette réponse peut aussi être considérée comme étant une réponse « idéale » du champ, par rapport aux propriétés ($P \wedge Q$)). Par extension, nous appellerons aussi *un champ de référence* toute activité d'un champ neuronal, calculée par cette procédure, quelle que soit i . Remarquons que, pour une telle réponse u' du champ, $\Delta_i^{\mathcal{B}}(u') = 0$.

Le fait que toutes les réponses d'un tel champ soient *calculées* par une procédure combinatoire, ayant donc une grande complexité de calcul et étant totalement non-distribuée, ne nous satisfait pas à déclarer que nous avons trouvé, par cette procédure de descente de gradient, les moyens techniques de produire l'activité d'un champ neuronal. Ceci nous motive donc pour poursuivre la recherche d'un champ neuronal, capable de satisfaire en effet les critères souhaités, décrits par les deux propriétés ($P \wedge Q$). Nous soulignons néanmoins que cette procédure de calcul nous permet de réaliser une comparaison empirique des capacités des champs neuronaux à satisfaire à ces propriétés attendues.

Utilisant la procédure présentée ci-dessus pour calculer les réponses de référence d'un champ neuronal, nous justifions dans le paragraphe suivant que ($P \wedge Q$) sont en effet des propriétés souhaitables pour les contextes de modélisation d'auto-organisation.

5.2.4 Pertinence des critères d'optimalité

Dans le paragraphe 5.2.1 de ce chapitre nous avons présenté plusieurs comportements attendus de la part d'un champ neuronal, quand il est stimulé par différentes sources de stimulation neuronale externe. Ces comportements ont été formalisés sous la forme des propriétés dynamiques (P) et (Q). Si le même paragraphe offre plusieurs justifications pour considérer ces propriétés comme souhaitables, la plus forte argumentation vis-à-vis de ces propriétés nous est offerte par l'analyse des comportements des champs respectant ($P \wedge Q$), dans des contextes de modélisation d'auto-organisation neuronale, le sujet central de notre recherche, qui est abordé de nouveau dans ce paragraphe.

Afin de resituer ce contexte de modélisation, nous rappelons ici les trois étapes principales du processus d'auto-organisation, tel qu'il est mis en œuvre par l'architecture neuronale d'inspiration corticale décrite dans la section 2.6. La première étape consiste en la présentation d'un nouvel échantillon $\xi(t)$ au module de reconnaissance, à chaque pas de temps t . Deuxièmement, les unités de ce module répondent à cette nouvelle entrée, chacune avec une activité neuronale $i(x, t)$. Dans la dernière étape, cette distribution stimule la couche corticale de sélection, qui formule sa réponse $u(x, t)$, pour chaque unité du module de sélection. Dans le discours que nous avons développé jusqu'ici, nous avons apporté, un par un, plusieurs arguments concernant la nécessité de considérer le module de sélection de l'architecture neuronale proposée comme étant un champ neuronal. Notre principale motivation de recherche étant la mise en évidence des champs neuronaux capables de participer au processus d'auto-organisation, nous examinons ici la capacité des champs de référence, introduits dans le paragraphe précédent, à organiser correctement la topologie de plusieurs prototypes, initialisés de manière aléatoire. Par conséquent, nous examinons empiriquement dans la suite le comportement des champs de référence dans le scénario d'auto-organisation unidimensionnelle décrit dans le paragraphe 5.1.4. Afin de valider les résultats de cette expérience, une comparaison visuelle avec les résultats de l'algorithme SOM de Kohonen, décrit dans la section 2.4.3, est présentée dans la figure 5.11.

Considérons l'implantation du processus d'auto-organisation par des champs neuronaux de référence. Ainsi, pour chaque pas de simulation, après la troisième étape de ce processus, la réponse u du module de sélection est calculée utilisant la procédure combinatoire décrite dans le paragraphe précédent. Dans ce cas de modélisation, u est le résultat de la procédure de minimisation de la distance $d(i, \mathcal{B})$, ce qui revient à affirmer qu'il est un élément de \mathcal{B}_i , ensemble de réponses d'un champ satisfaisant pleinement les propriétés $(P \wedge Q)$.

Nous notons que, dans les simulations numériques de ce scénario, nous avons considéré une valeur $b_{min} \rightarrow \infty$, afin que la condition de densité de la propriété (Q) impose l'existence d'une unique bulle d'activité dans le champ. Nous avons fixé également un pas de discrétisation $\delta = 1$ (valable aussi pour toutes nos autres simulations numériques), et un rayon $R = 4$ pour la bulle d'excitation attendue. Ces paramètres permettent de définir complètement l'ensemble \mathcal{B} discrétisé.

Utilisant cette implantation de champ pour le module de sélection de l'architecture d'auto-organisation, nous pouvons simuler numériquement le déroulement du scénario unidimensionnel décrit dans 5.1.4. Ainsi, en réalisant ceci, nous pouvons constater que le réseau des prototypes converge vers un état stable, caractérisé par une topologie de connexion auto-organisée, comme illustré dans la figure 5.11 (b), toute à fait similaire à celle obtenue par l'implantation de l'algorithme SOM de Kohonen (voir la même figure, cas (a)), algorithme décrit dans la section 2.4.3, et appliqué en considérant une fonction de voisinage de taille fixe.

Les bons résultats présentés sur la figure 5.11 s'appuient sur deux principales caractéristiques dynamiques des champs de référence, que nous détaillons dans la suite.

La première caractéristique est liée à la capacité du champ à générer des bulles d'activité en condition de faible stimulation. En effet, au tout début de l'expérience, les distributions i générées par le module de reconnaissance de l'architecture neuronale d'auto-organisation sont caractérisées par seulement quelques pics d'activité. Cet aspect est dû au fait que les prototypes sont initialisés avec des valeurs aléatoires, ce qui implique que très peu d'unités du module de reconnaissance s'activent (concordent avec l'échantillon ξ présenté). Toutefois, même si certaines unités génèrent une activité i de forte amplitude, il est très improbable que toutes ces unités vont se trouver proches l'une de l'autre, afin qu'elles puissent stimuler suffisamment le module de sélection, et entraîner par conséquent une forte réponse de sa part. L'avantage des champs respectant les propriétés $(P \wedge Q)$ est qu'ils sont capables (par définition) de générer une bulle

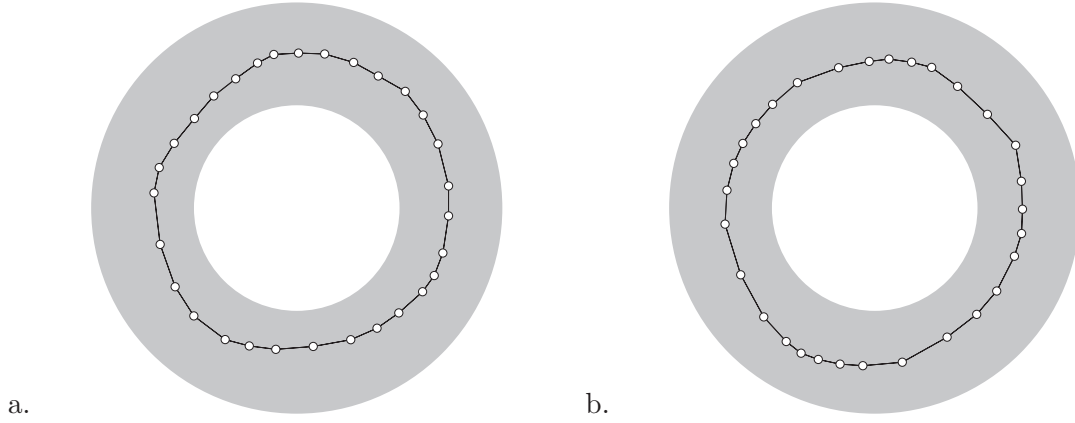


FIG. 5.11 – Comparaison entre les résultats d'auto-organisation obtenus utilisant l'architecture neuronale décrite dans 2.6 pour implémenter l'algorithme SOM de Kohonen (voir la figure a), et utilisant les champs de référence (satisfaisant pleinement les propriétés $(P \wedge Q)$) comme module de sélection de cette architecture (voir la figure b).

d'excitation forte, même dans des conditions de très faible stimulation externe. Ainsi, quel que soit i , le module de sélection mené par les champs de référence garde son caractère de sélectivité, et en particulier dans tous les cas similaires au cas (h) illustré graphiquement dans la figure 5.6, qui énumère plusieurs comportements attendus d'un champ neuronal. Cet aspect est crucial au début du processus d'auto-organisation, étant donné que l'apprentissage (c'est-à-dire l'ajustement) des prototypes se réalise en fonction de l'existence d'une bulle d'excitation dans la réponse u du champ.

La deuxième caractéristique dynamique du champ, également essentielle pour le processus d'auto-organisation, se réfère à l'adaptation de la réponse du champ suite à un changement de la distribution i provenant du module de reconnaissance. En effet, le champ de référence utilisé dans cette expérience est conçu (par définition) pour respecter $(P \wedge Q)$. Or, comme le montre le cas (g) de la figure 5.6 des comportements souhaitables, un tel champ est censé reformuler sa réponse en accord avec les modifications de la distribution i . Maintenant correctement sa capacité de sélectivité, le champ reste sensible aux changements enregistrés au niveau du module de reconnaissance, ce qui permet de favoriser toujours l'ajustement de la majorité des prototypes répondant le mieux aux échantillons $\{\xi_k\} \subset \Xi$ présentés au cours du déroulement du scénario d'auto-organisation.

En conclusion, nous pouvons affirmer que les résultats présentés ci-dessus valident empiriquement le fait que (P) et (Q) sont deux propriétés suffisantes pour un champ neuronal, afin qu'il puisse conduire correctement un processus d'auto-organisation, comme décrit par l'architecture d'inspiration corticale introduite dans la section 2.6.

Malgré les très bonnes performances des champs de référence, nous tenons à souligner encore une fois que ces champs ne sont pas de *vrais* champ neuronaux, mais les résultats d'un *calcul* par une procédure de recherche combinatoire, qui ne respecte pas les principes de calcul neuronal que nous avons définis dans le paragraphe 2.5. Ainsi, ces champs ne sont pas appropriés pour un usage pratique dans des applications de modélisation corticale, mais ont vocation à montrer la capacité des propriétés $(P \wedge Q)$ à garantir la bonne convergence du processus d'auto-organisation.

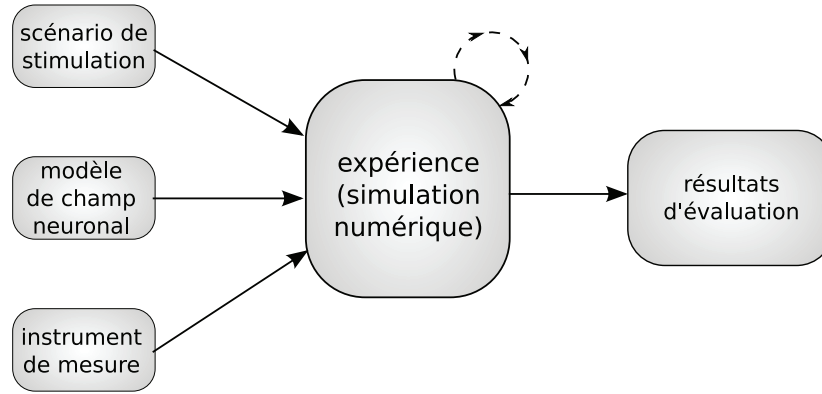


FIG. 5.12 – Schéma général d'évaluation empirique des champs neuronaux.

5.2.5 Plate-forme d'analyse empirique de la dynamique des champs neuronaux

Le paragraphe précédent avait conclu sur la pertinence des propriétés $(P \wedge Q)$, alors que le paragraphe 5.2.3 avait évoqué l'inefficacité pratique d'utiliser un algorithme de descente de gradient pour concevoir un champ neuronal capable de satisfaire ces propriétés. Changeant la perspective de ce problème, on peut se demander si un modèle quelconque de champ, décrit par un formalisme mathématique, est capable de satisfaire à ces deux propriétés. Par conséquent, dans ce paragraphe, nous présentons une plate-forme d'analyse empirique des modèles de champs neuronaux, qui a pour but d'évaluer leurs performances vis-à-vis de ces deux propriétés dynamiques, dans différents contextes de stimulation.

En résumé, cette méthode définit une approche empirique de type *banc d'essai* (en anglais *benchmark*), qui utilise la simulation numérique, et qui consiste à évaluer numériquement la capacité d'un champ neuronal à satisfaire aux critères de performance souhaités (en l'occurrence $(P \wedge Q)$), dans un ou plusieurs scénarios. La figure 5.12 illustre de manière schématique le principe de cette approche.

Le triplet (*scénario*, *modèle de champ*, *instrument d'évaluation*) définit une *expérience* de simulation. Le déroulement d'une expérience suppose qu'à chaque instant discret t du scénario i considéré, la réponse u du champ est calculée (par une méthode de simulation numérique du modèle de champ) et l'instrument d'évaluation mesure la capacité du modèle de champ à satisfaire aux critères de performance considérés. Une expérience a donc le rôle de simuler et de collecter les valeurs numériques calculées par l'instrument d'évaluation vis-à-vis de l'évolution du champ, à travers le scénario i examiné. Ces données peuvent servir ensuite à une analyse plus détaillée de la dynamique du champ.

Une expérience peut être répétée plusieurs fois (pour le même scénario), afin de constater l'éventuelle influence des facteurs aléatoires vis-à-vis du comportement du champ, facteurs qui ne peuvent pas être contrôlés par la simulation (par exemple l'effet de l'évaluation asynchrone, etc.). Ceci permettrait une meilleure observation de la statistique de ces aspects dynamiques aléatoires.

L'instrument d'évaluation que nous proposons d'utiliser pour cette plate-forme d'évaluation empirique est l'indicateur $\bar{\Delta}_i^B(u)$, défini par l'équation 5.11. Cet indicateur numérique dénote l'erreur résiduelle de la capacité instantanée d'un champ neuronal à satisfaire aux propriétés $(P \wedge Q)$. Ainsi, une valeur nulle de $\bar{\Delta}_i^B(u)$ indique une satisfaction totale des deux propriétés par

l'activité du champ, vis-à-vis d'un scénario i et un instant t donnés.

Comme nous pouvons le constater, cet indicateur dépend à la fois du moment t auquel l'activité u du champ est évaluée, et du scénario i examiné. Pour cette raison, une comparaison directe entre différentes valeurs de cet indicateur, pour différents scénarios i , peut s'avérer peu concluante. Comme nous allons l'observer dans la section suivante, au lieu de qualifier globalement la capacité d'un champ à satisfaire les deux propriétés ($P \wedge Q$), il est nécessaire d'analyser la dynamique d'un champ d'une manière individuelle, pour chaque scénario considéré.

Notons également que l'indicateur $\bar{\Delta}_i^B(u)$ ne mesure pas le degré de satisfaction des propriétés ($P \wedge Q$). Cependant, si sa valeur converge vers 0, nous pouvons affirmer que le champ améliore son degré de satisfaction de ces propriétés. En outre, notons que chaque modification du scénario de stimulation i se reflète par une période transitoire dans la dynamique du champ. Généralement, à la fin de ce processus, le champ atteint un nouvel état stable (un attracteur). Normalement, pendant les périodes transitoires, la valeur de $\bar{\Delta}_i^B(u)$ augmente, afin que le champ puisse « reformuler » son caractère sélectif par rapport à la nouvelle distribution de stimulation i . Si ce caractère sélectif se manifeste de nouveau (en rapport avec le nouveau i), après la stabilisation de l'activité du champ, la valeur $\bar{\Delta}_i^B(u)$ redescend vers 0. Pour analyser en détail cet aspect dynamique d'un champ, nous définissons sa *courbe de qualité dynamique*, pour un scénario i donné, comme étant la séquence des valeurs $\bar{\Delta}_i^B(u)$, enregistrées pour une série successive de pas de simulation $t \geq 0$.

Comme nous sommes intéressé principalement par les états stables de la dynamique d'un champ neuronal (surtout les bulles stables d'activité neuronale), nous introduisons également une mesure statistique de l'évaluation du champ. Ainsi, nous proposons de tracer l'*histogramme de qualité* d'un champ neuronal, examiné pendant une expérience. Un tel histogramme de qualité est déduit à partir de la courbe de qualité, en échantillonnant l'intervalle des valeurs $\bar{\Delta}_i^B(u)$ enregistrées pendant l'expérience, et en cumulant pour chaque sous-intervalle le nombre de pas où la réponse u atteint le niveau $\bar{\Delta}_i^B(u)$ dans le sous-intervalle respectif. En conclusion, si le champ neuronal est capable de satisfaire les propriétés ($P \wedge Q$), et en plus, le temps cumulé des périodes de stabilité du champ domine le temps cumulé des périodes des régimes transitoires, l'histogramme de qualité du champ affiche une distribution inclinée vers les valeurs faibles de $\bar{\Delta}_i^B(u)$. Cet aspect est une indication que, la plupart du temps, le champ est capable de satisfaire les critères imposés par les deux propriétés, ($P \wedge Q$).

Dans ce paragraphe nous avons décrit une approche empirique pour évaluer numériquement la capacité des champs neuronaux à satisfaire certains critères de performance souhaités, dans plusieurs scénarios de stimulation. Cette plate-forme automatique d'évaluation, implantée en pratique sous la forme d'une plate-forme logicielle, est utilisée dans la section 5.3 pour examiner le comportement des champs neuronaux classiques, vis-à-vis des deux propriétés dynamiques, ($P \wedge Q$), dans différents contextes clé de stimulation.

Enfin, mentionnons que cette méthode aide à s'atteler au problème du choix de paramètres pour une équation de champ, un aspect abordé dans la section 5.4.

5.3 Analyse du comportement dynamique des modèles classiques

Utilisant la plate-forme d'évaluation empirique des champs neuronaux dynamiques introduite au paragraphe précédent, nous examinons dans ce paragraphe de manière détaillée le comportement de deux modèles classiques de champs neuronaux, dans les contextes de stimulation décrits par les scénarios présentés dans la section 5.1. Les deux modèles analysés ici sont : le modèle

d'Amari, décrit dans la section 3.2.2, et référé ici par l'abréviation CNDA, et deuxièmement le modèle de Pinto et Ermentrout, présenté dans la section 3.2.5 du chapitre 3, et abrégé ici CNDPE.

Les valeurs numériques pour les paramètres utilisés pour simuler les expériences présentées dans les paragraphes 5.3.1 à 5.3.3 sont :

- *Paramètres généraux*
 - $X = 0, 1, \dots, 29$
 - nombre de pas de simulations par expérience : 200
- *Paramètres CNDA* (cf. l'équation 3.15)
 - constante de temps $\tau = 0.3^{-1}$
 - noyau des poids des connexions latérales de type chapeau mexicain, cf. l'équation 3.3 : $A^+ = 0.8, A^- = 0.5, a = 0.05, b = 0.001$
 - fonction sigmoïde σ , similaire à celle de l'équation 3.6, définie comme une fonction linéaire par morceaux :

$$\sigma(u) = \begin{cases} 0 & u = 0 \\ 0.4u & 0 < u \leq 0.25 \\ 1.6(u - 0.25) + 0.1 & 0.25 < u \leq 0.75 \\ 0.4(u - 0.75) + 0.9 & 0.75 < u \leq 1 \end{cases}$$

- niveau de repos neuronal $h = 0$
- *Paramètres CNDPE* (cf. les équations 3.26 et 3.27)
 - constantes de temps $\tau = 0.3^{-1}$ et $\eta = 0.85^{-1}$
 - noyau des poids des connexions latérales de type chapeau mexicain, cf. l'équation 3.3 : $A^+ = 0.67, A^- = 0.35, a = 0.05, b = 0.001$
 - la même fonction sigmoïde σ , linéaire par morceaux, utilisée pour CNDA
 - niveau de repos neuronal $h = 0$

Afin de minimiser l'influence des effets de bord de la topologie de X sur les réponses des champs, toutes les expériences de simulation considèrent une topologie de connexion circulaire pour les unités des champs, c'est-à-dire une topologie de type anneau pour le cas 1D, et de type tore pour le cas 2D.

5.3.1 Scénarios d'attention

Dans ce paragraphe nous présentons les résultats expérimentaux de l'évaluation des modèles CNDA et CNDPE, dans les scénarios d'attention considérés dans 5.1.1. La figure 5.14 résume les expériences déroulées pour ces scénarios. Le format de présentation de ces résultats est décrit sur la figure 5.13.

Comme nous pouvons constater analysant l'ensemble des courbes et des histogrammes de qualité tracées pour CNDA et CNDPE, les deux modèles examinés montrent une capacité élevée de satisfaction des critères dynamiques souhaités. Ils sont à la fois capables de sélectionner une zone pertinente des stimuli externes, et de s'adapter aux changements intervenus dans leur distribution au cours du temps. Le caractère sélectif et adaptatif leur confère une bonne dynamique dans l'ensemble, ce qui rend ces deux modèles de champ appropriés pour être utilisés dans des applications pratiques de modélisation des mécanismes d'attention. Ces résultats viennent confirmer le succès de ces modèles dans la littérature du domaine, et en particulier du modèle d'Amari, dans ce contexte de modélisation neuronale, ou dans des contextes cognitifs connexes (voir la modélisation de certains aspects ou mécanismes neuronaux de mémoire, de processus

scénario de stimulation i	réponse u du champ CNDA	réponse u du champ CNDPE
	<p>courbe de qualité exprimant graphiquement l'évolution temporelle du coefficient $\bar{\Delta}_i^B(u)$, pour le modèle CNDA</p>	<p>histogramme associée à la courbe qualité, calculée pour le modèle CNDA</p>
	<p>courbe de qualité exprimant graphiquement l'évolution temporelle du coefficient $\bar{\Delta}_i^B(u)$, pour le modèle CNDPE</p>	<p>histogramme associée à la courbe qualité, calculée pour le modèle CNDPE</p>

FIG. 5.13 – Légende générale détaillée pour les figures 5.14-5.16, représentant plusieurs résultats expérimentaux.

aidant à l'émergence de la conscience, à la sélection des actions motrices, etc.), fait déjà évoqué aux paragraphes 4.2 et 4.3.

5.3.2 Scénarios d'exploration

Dans ce paragraphe nous présentons les résultats expérimentaux de l'évaluation des modèles CNDA et CNDPE, dans les scénarios d'exploration considérés dans 5.1.2. La figure 5.15 résume les expériences déroulées pour ces scénarios. La signification des graphiques reste la même que celle décrite pour la figure 5.14.

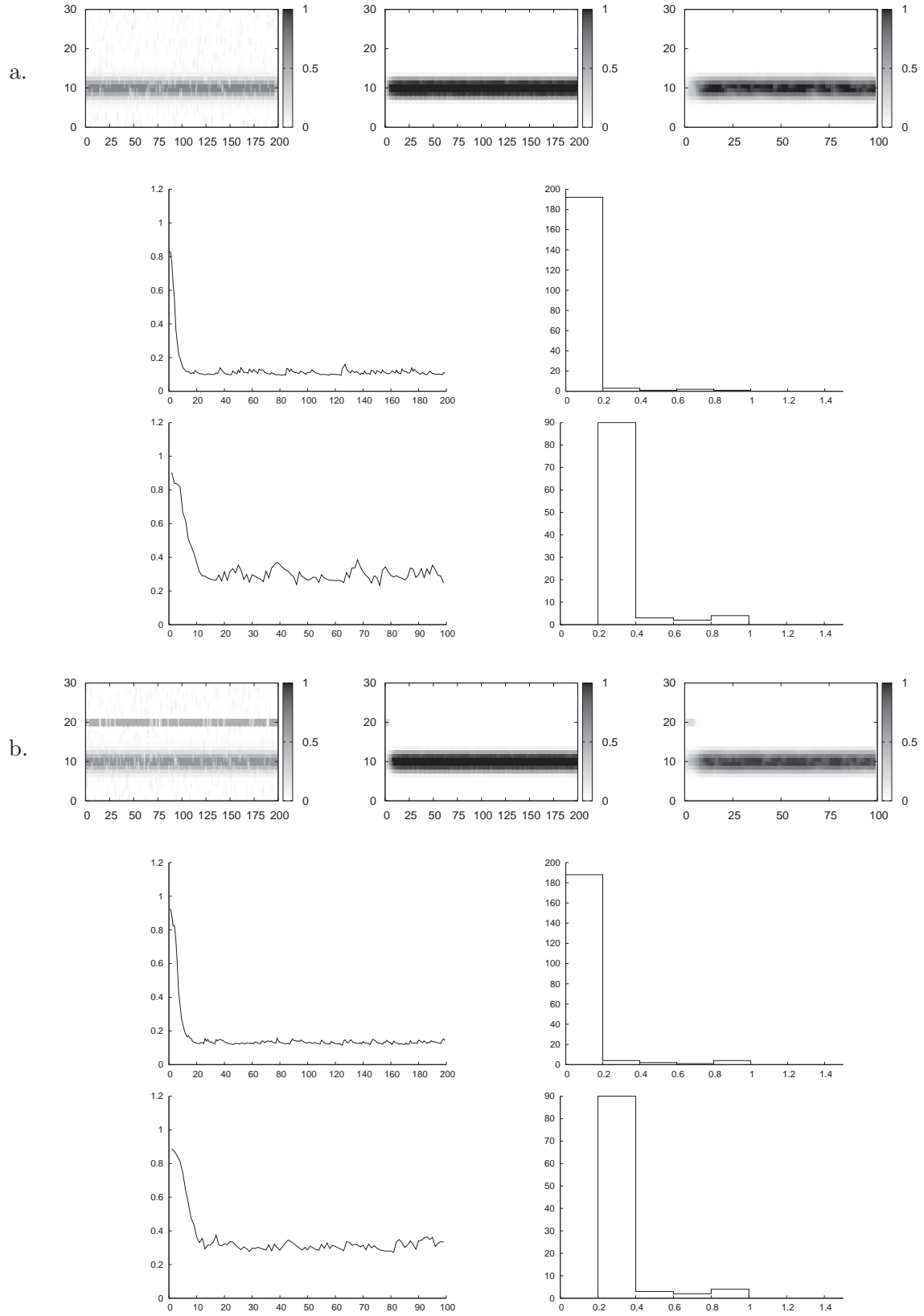
Comme nous l'avons précisé dans le paragraphe 5.1.2, ces scénarios ont été conçus pour tester la capacité des champs à maintenir leur caractère sélectif dans des conditions où, a priori, suite à une stimulation uniforme du champ, chaque unité a la même probabilité d'intensifier son activité afin de rendre le champ sélectif. Si les deux modèles, CNDA et CNDPE, réussissent à générer une bulle d'activité quand toutes les unités du champ reçoivent une forte stimulation (comme dans le cas (a) de la figure 5.15), aucun d'eux est capable d'avoir une réponse sélective dans des conditions de faible stimulation externe (voir le cas (b) de la même figure).

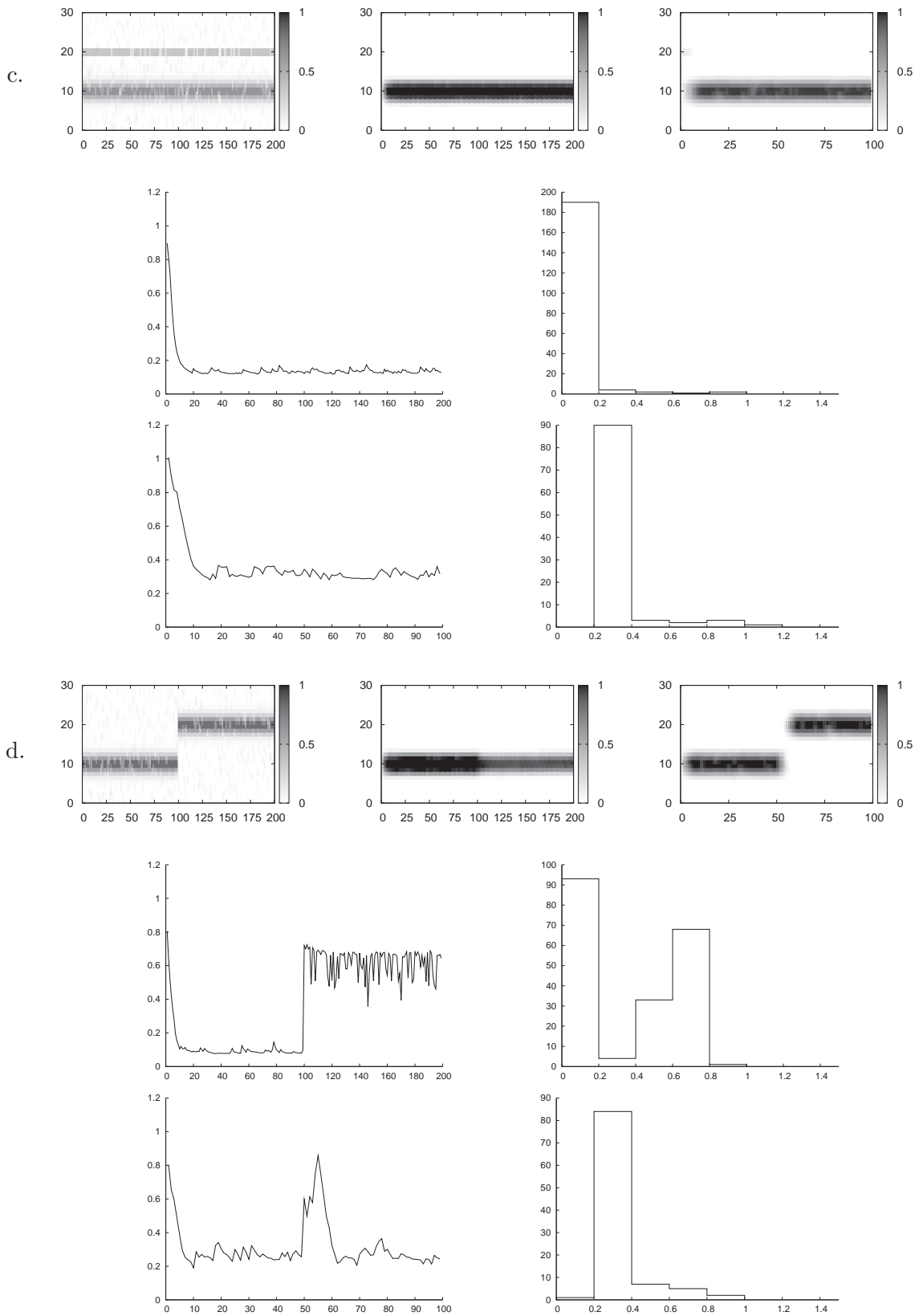
Comme nous allons analyser plus en détail dans les sections 5.3.4 et 5.3.5, dédiées aux scénarios d'auto-organisation, ce fait a une influence déterminante dans le processus d'auto-organisation, tel qu'il est conçu par notre approche cortico-inspirée.

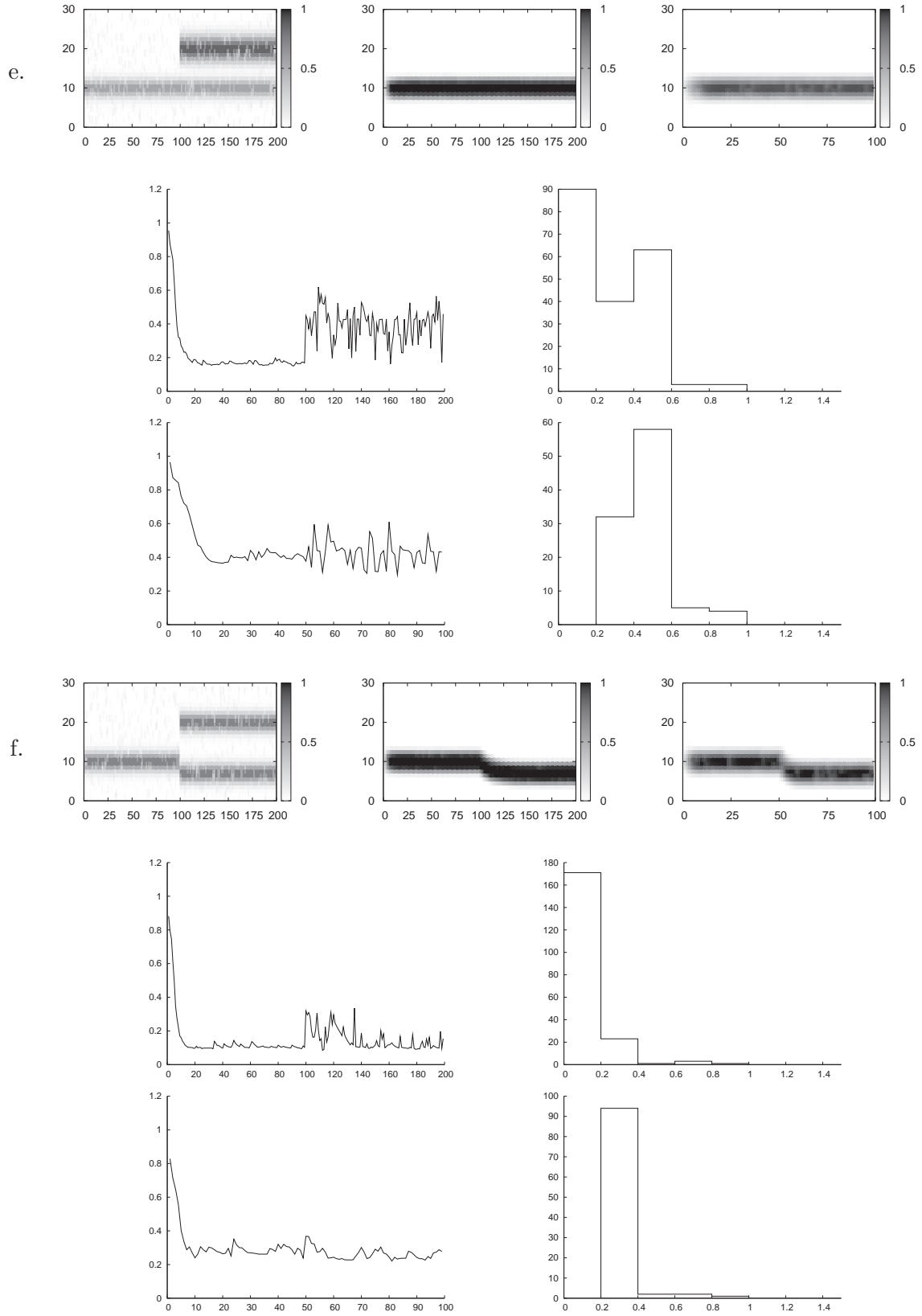
5.3.3 Scénarios auxiliaires

La figure 5.16 présente les résultats expérimentaux enregistrés en simulant les expériences des scénarios auxiliaires, précédemment définis dans la figure 5.3. La signification des graphiques reste la même que celle décrite pour la figure 5.14.

Même si ces scénarios pouvaient être considérés aussi comme scénarios d'attention, ils ont pour but d'étudier en particulier l'influence de certains facteurs dynamiques tels que : la capacité du champ à « décider » la région de stimulation à sélectionner, quand plusieurs réponses possibles entrent en compétition (voir les cas (a,c), qui font référence à la propriété (b) illustrée dans la figure 5.6) ; la capacité de collaboration évoquée par le cas (f) de la figure 5.6 (autrement dit la capacité du champ à « mettre en balance » la contribution de plusieurs stimuli locaux afin de







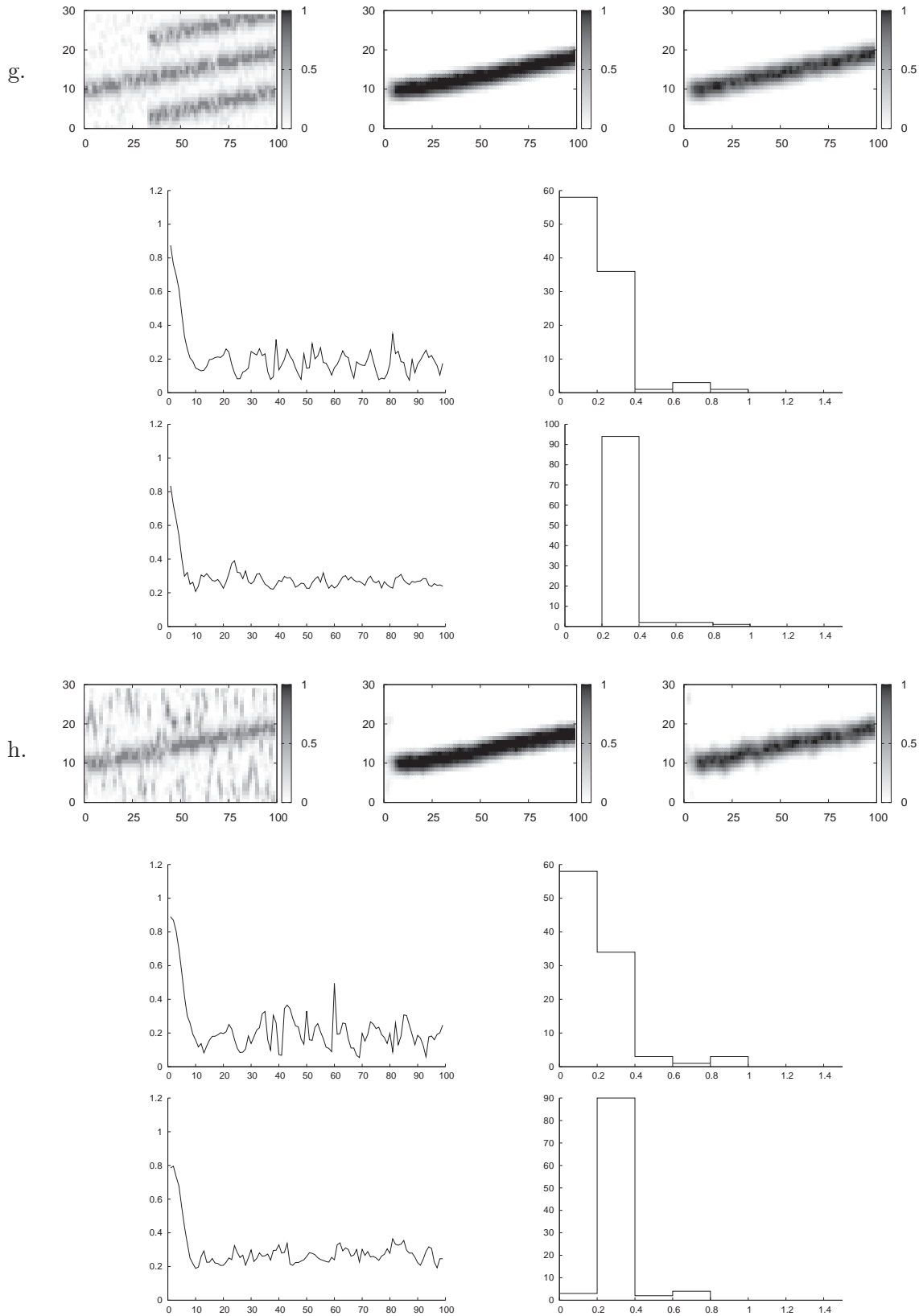


FIG. 5.14 – Résultats expérimentaux de l'évaluation des modèles CNDA et CNDPE, dans les scénarios d'attention considérés dans la figure 5.1. La signification de ces graphiques est décrite sur la figure 5.13. Suivre le texte pour plus de détails.

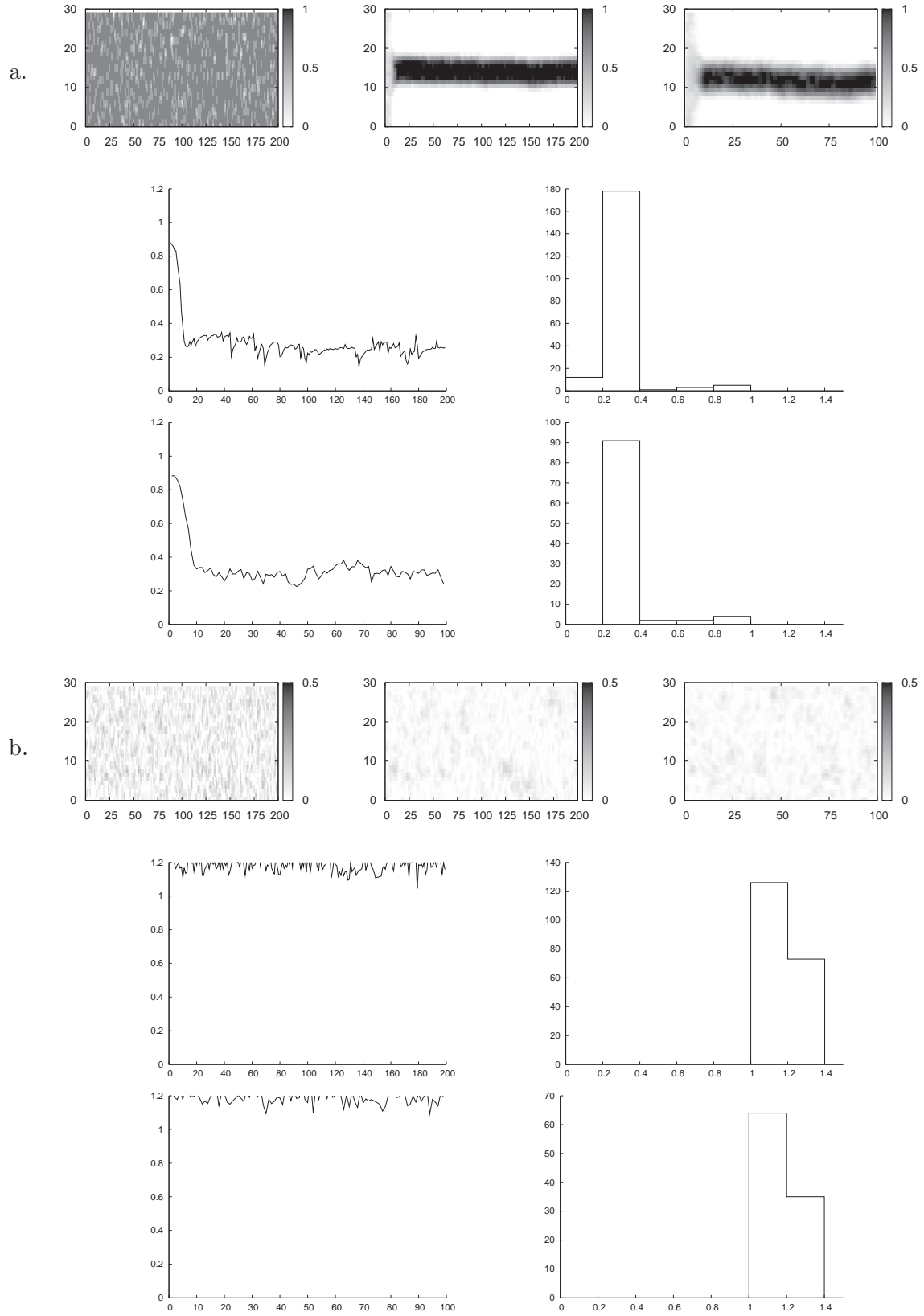
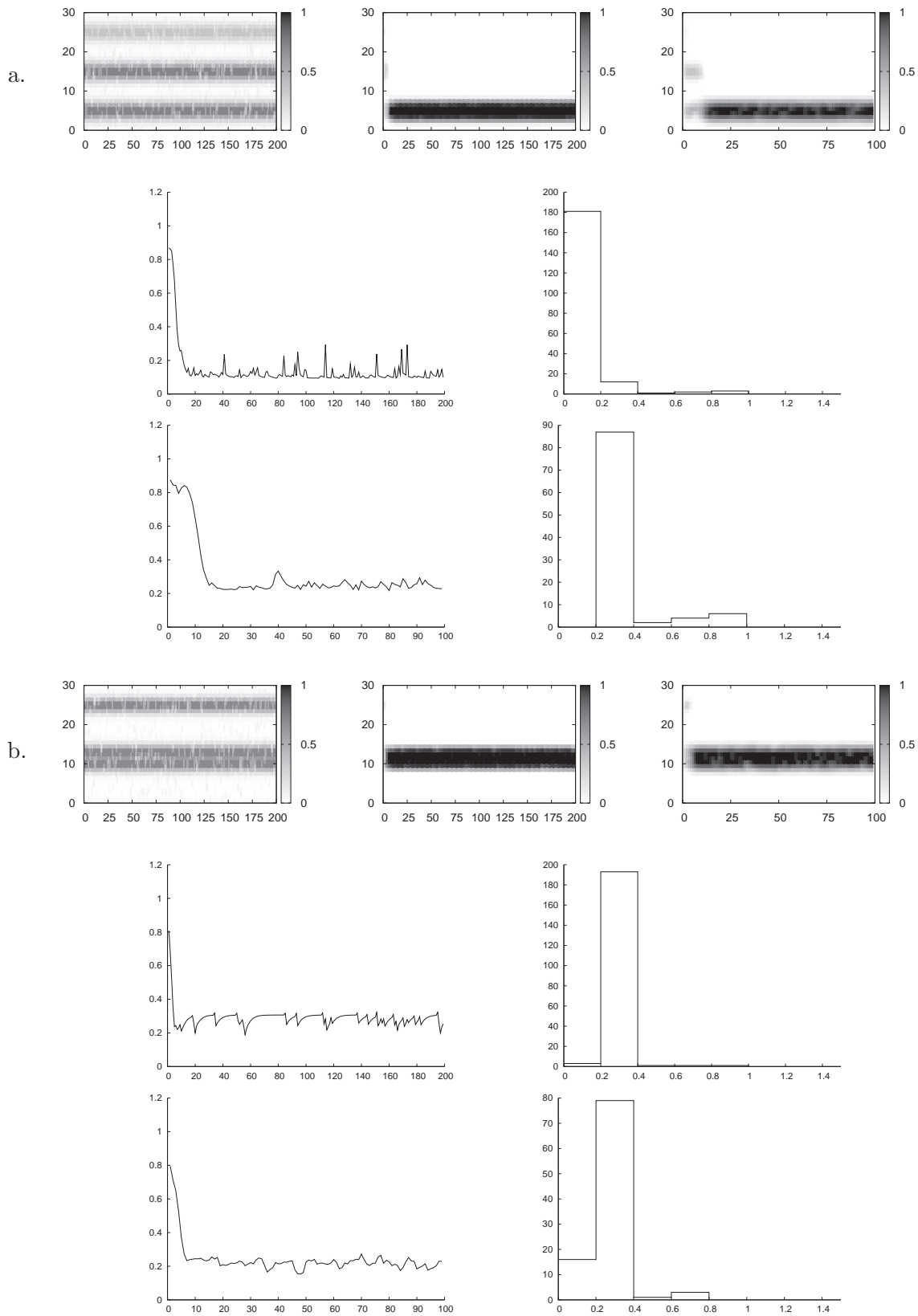
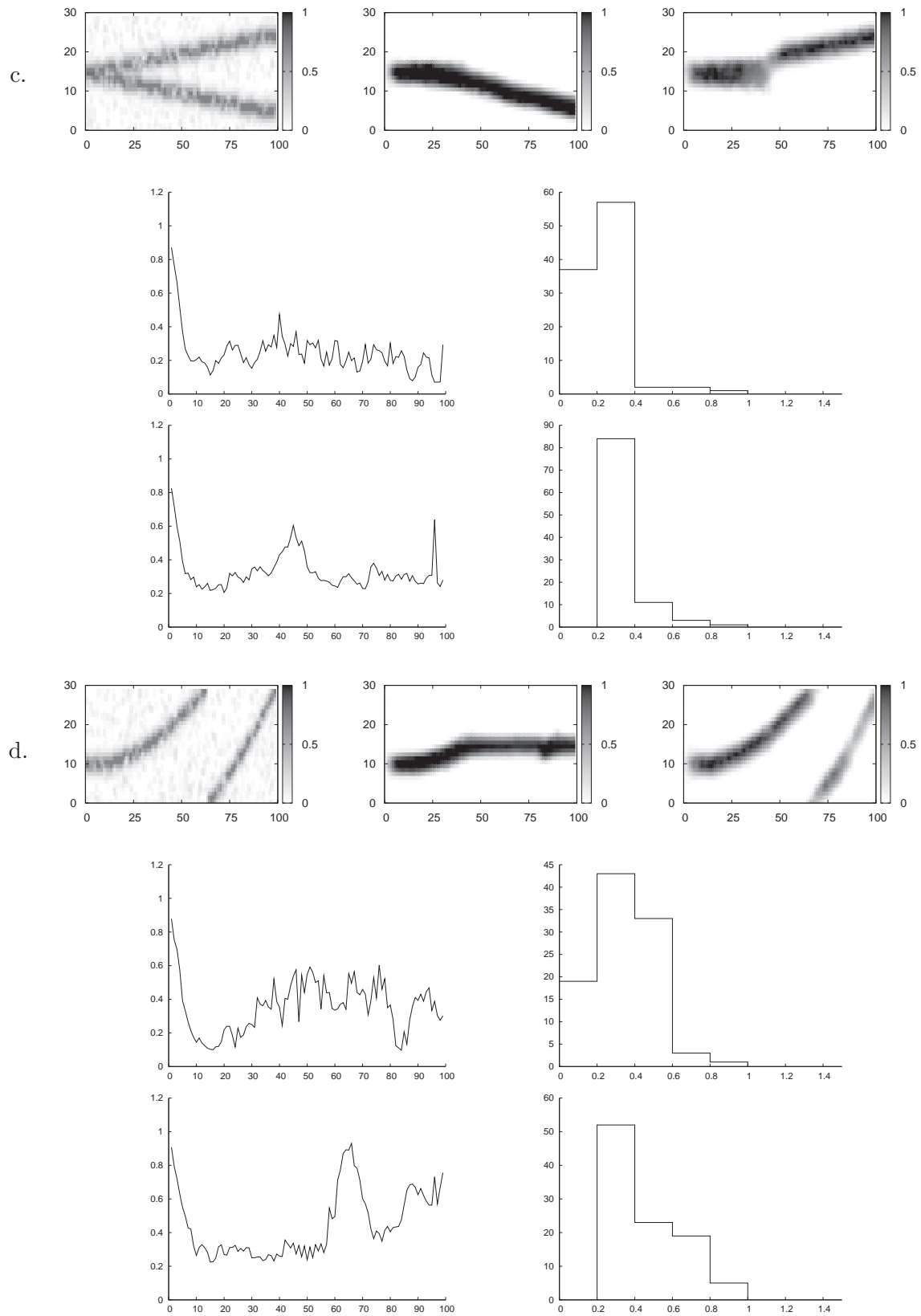


FIG. 5.15 – Résultats expérimentaux de l'évaluation des modèles CNDA et CNDPE, dans les scénarios d'exploration considérés dans la figure 5.2. La signification de ces graphiques est décrite sur la figure 5.13. Suivre le texte pour plus de détails.





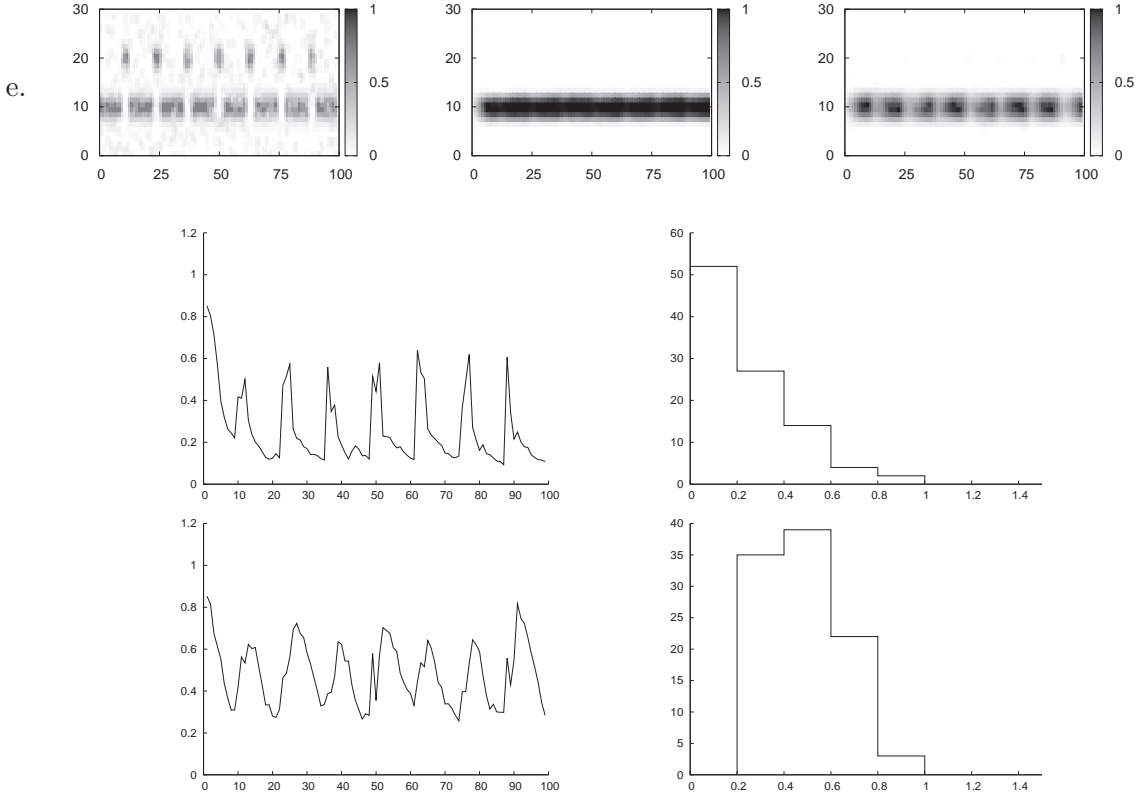


FIG. 5.16 – Résultats expérimentaux de l'évaluation des modèles CNDA et CNDPE, dans les scénarios auxiliaires considérés dans la figure 5.3. La signification de ces graphiques est décrite sur la figure 5.13. Suivre le texte pour plus de détails.

pouvoir sélectionner une zone pertinente de la stimulation externe) (voir le cas (b)) ; la vitesse de réponse du champ aux changements survenus dans la distribution des stimuli d'entrée (voir les cas (d,e)).

La qualité des deux modèles est très comparable dans les cas (a-c), mais comme nous pouvons le constater, le modèle CNDPE montre une adaptabilité légèrement meilleure face aux changements de la stimulation externe.

5.3.4 Scénario d'auto-organisation unidimensionnelle

Dans ce paragraphe nous abordons de nouveau le sujet d'auto-organisation, cette fois-ci en utilisant les champs neuronaux CNDA et CNDPE. Ainsi, nous sommes intéressé ici à examiner leur capacité à auto-organiser un ensemble unidimensionnel de prototypes, ayant le but de réaliser la quantification vectorielle du domaine d'entrée Ξ défini dans le paragraphe 5.1.4. La figure 5.17 montre les résultats du processus d'auto-organisation (après la présentation de 12000 échantillons, pour chaque 100 pas de simulation), pour chacun des modèles de champ considérés.

Afin de comprendre les causes de ces résultats insatisfaisants, nous résumons ici les conclusions de deux paragraphes précédents de ce chapitre, 5.2.4 et 5.3.2.

Premièrement, nous rappelons que l'expérience d'auto-organisation examinée ici a été pré-

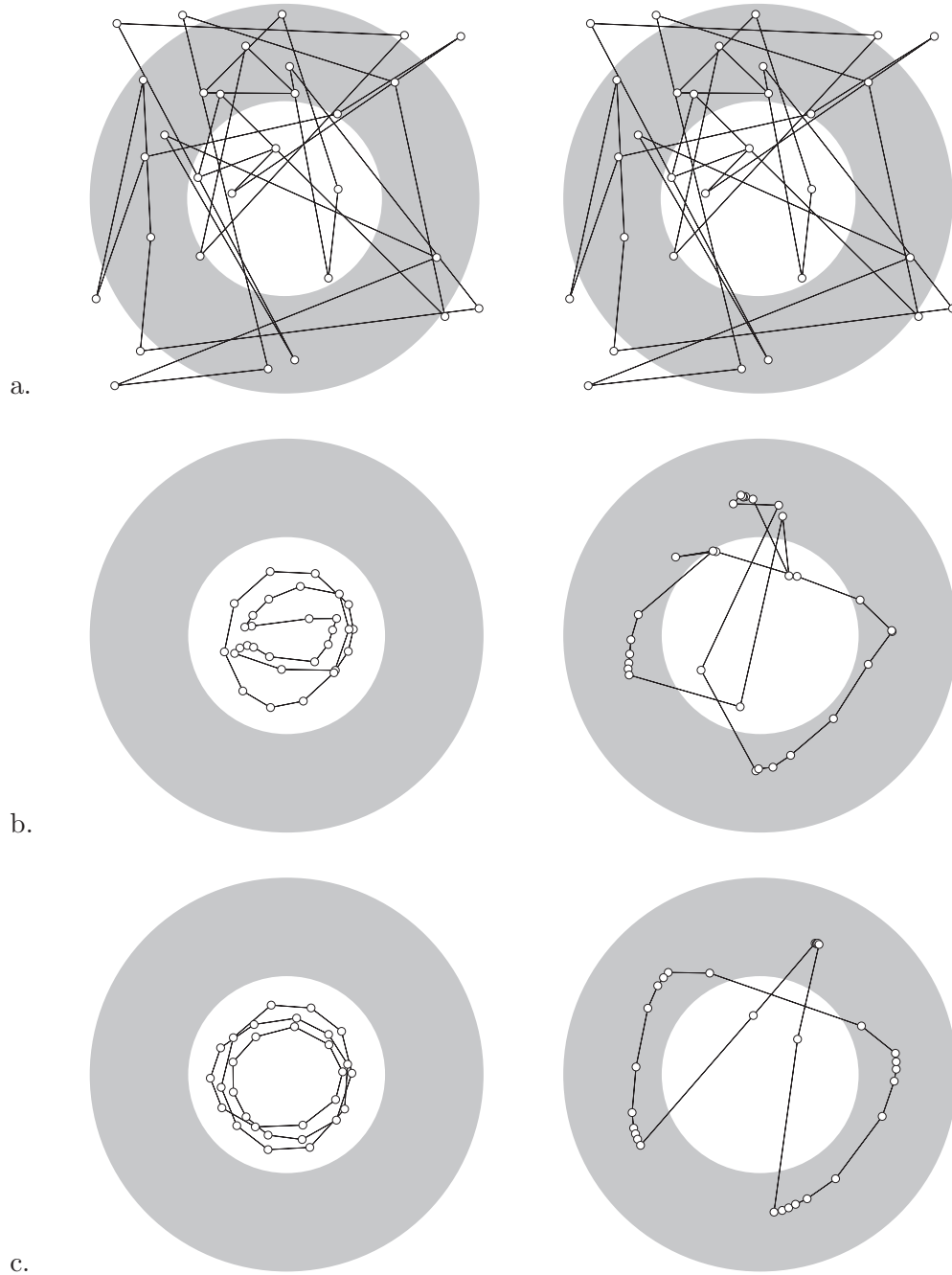


FIG. 5.17 – Les étapes initiale (a), intermédiaire (b), et finale (c), de l'expérience d'auto-organisation unidimensionnelle, déroulée conformément au scénario décrit dans 5.3.4. Comparaison entre les résultats obtenus utilisant comme module de sélection de l'architecture d'auto-organisation d'inspiration corticale le modèle CNDA (colonne gauche), et le modèle CNDPE (colonne droite).

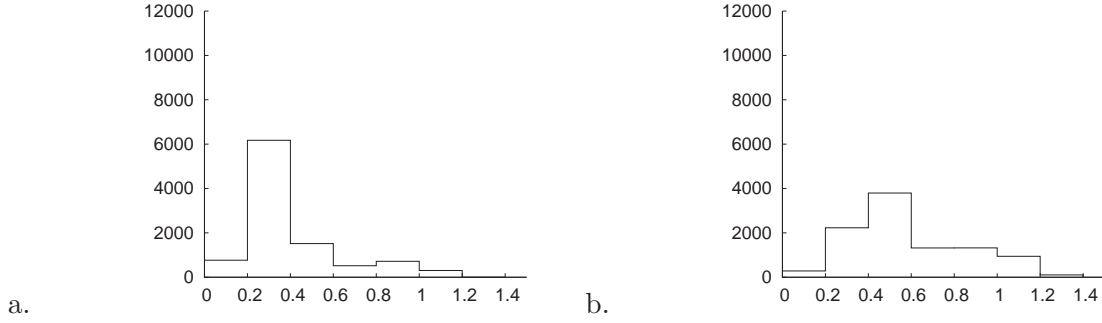


FIG. 5.18 – Les histogrammes de qualité correspondant respectivement aux modèles CNDA (a) et CNDPE (b), obtenues après la présentation de 12000 échantillons dans le scénario unidimensionnel d'auto-organisation.

cédemment réalisée dans la section 5.2.4, pour les champs neuronaux de référence (c'est-à-dire ceux conçus tels qu'ils puissent satisfaire pleinement les propriétés $(P \wedge Q)$). Les résultats de cette expérience ont conclu que les champs de référence sont capables de conduire correctement le processus d'auto-organisation, accomplissant proprement la réorganisation topologique des prototypes.

Deuxièmement, grâce aux résultats expérimentaux de l'analyse dynamique empirique conduite dans la section 5.3, nous avons pu constater que les deux modèles de champ, CNDA et CNDPE, sont capables dans la vaste majorité des cas d'avoir un comportement dynamique similaire à celui souhaité (c'est-à-dire respectant les propriétés dynamiques $(P \wedge Q)$). Néanmoins, le paragraphe 5.3.2 de cette section montre qu'aucun des deux modèles ne réussit à maintenir son caractère sélectif dans le cas de faible stimulation.

Remarquons à partir de ces deux observations que la différence qualitative majeure entre les comportements des deux types de champs, d'une part ceux définis par les modèles classiques CNDA et CNDPE, et d'une autre part les champs de référence, se réfère à leur capacité de sélectivité dans tous les scénarios de stimulation. Comme nous allons l'analyser en détail dans la section 5.3.6, cet aspect s'avère déterminant pour l'obtention d'une bonne performance dans le cas de modélisation du processus d'auto-organisation.

5.3.5 Scénario d'auto-organisation bidimensionnelle

Examinons finalement ici la capacité des modèles CNDA et CNDPE à accomplir correctement la tâche d'auto-organisation bidimensionnelle décrite dans le paragraphe 5.3.5.

La figure 5.19 présente les résultats de cette expérience quand la courbe de reconnaissance des prototypes est donnée par l'équation 5.2, et la figure 5.20 présente les résultats de cette expérience quand la courbe de reconnaissance des prototypes est donnée par l'équation 5.3.

Nous pouvons observer que dans le premier cas de simulation considéré, les deux modèles de champ sont capables d'assurer la convergence du processus de quantification vectorielle vers la bonne réorganisation des prototypes. Toutefois, les mêmes champs ne sont plus capables de garantir la même qualité des résultats, quand les prototypes changent légèrement la forme de leur fonction de reconnaissance. Les différences qualitatives observées dans ces expériences sont discutées en détail dans le paragraphe suivant.

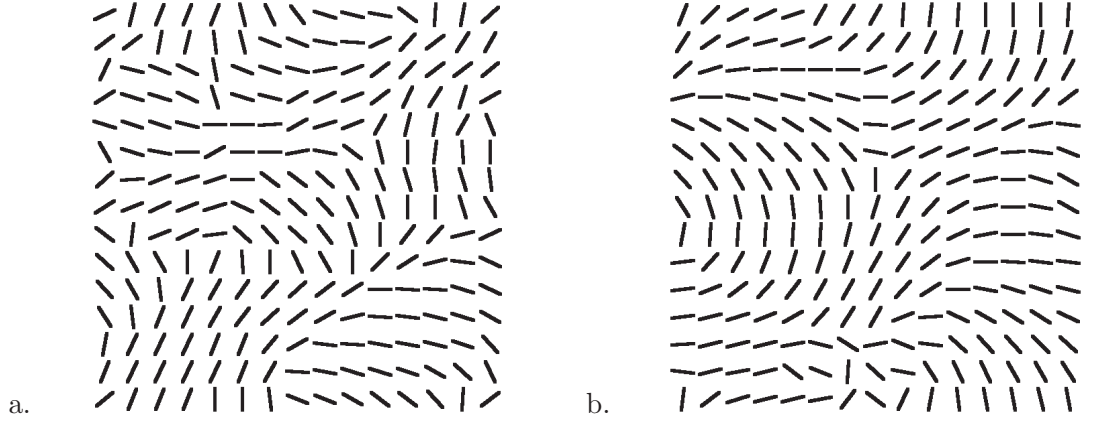


FIG. 5.19 – Expérience d’auto-organisation bidimensionnelle, déroulée conformément au scénario décrit dans 5.3.5, en considérant comme fonction de reconnaissance pour les prototypes la fonction f_2 définie par l’équation 5.2. Comparaison entre les résultats obtenus utilisant comme module de sélection de l’architecture d’auto-organisation d’inspiration corticale le modèle CNDA (figure (a)), et le modèle CNDPE (figure (b)).

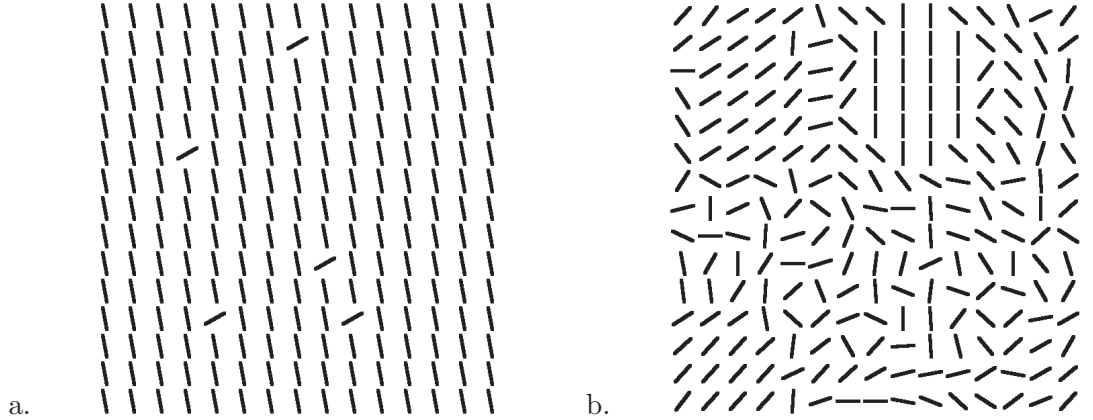


FIG. 5.20 – Expérience d’auto-organisation bidimensionnelle, déroulée conformément au scénario décrit dans 5.3.5, en considérant comme fonction de reconnaissance pour les prototypes la fonction f_3 définie par l’équation 5.3. Comparaison entre les résultats obtenus utilisant comme module de sélection de l’architecture d’auto-organisation d’inspiration corticale le modèle CNDA (figure (a)), et le modèle CNDPE (figure (b)).

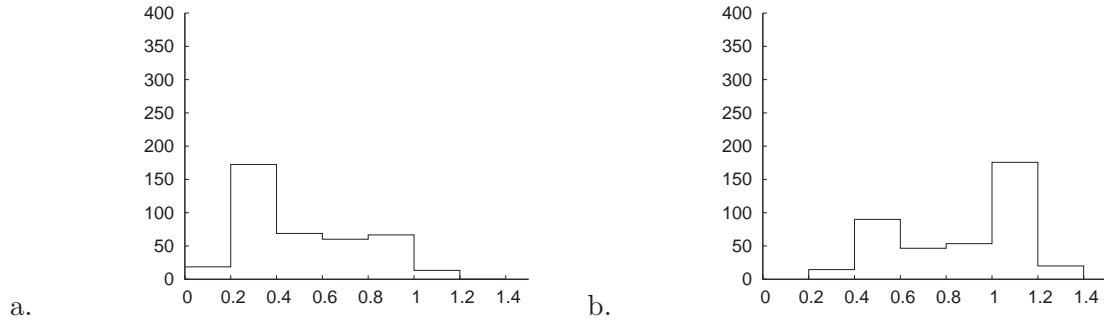


FIG. 5.21 – Les histogrammes de qualité correspondant respectivement aux modèles CNDA (a) et CNDPE (b), obtenus après la présentation des 400 premiers échantillons dans le scénario unidimensionnel d'auto-organisation.

5.3.6 Discussion

Nous analysons ici les résultats des expériences décrites au cours des paragraphes précédents, afin de synthétiser du point de vue qualitatif les capacités dynamiques des champs classiques examinés, CNDA et CNDPE.

Dans un premier temps, nous avons pu remarquer que les modèles classiques de champ réussissent à satisfaire les critères sous-jacents aux propriétés $(P \wedge Q)$, dans la grande majorité des scénarios examinés. La variété des scénarios d'attention montre que CNDA et CNDPE se prêtent à être utilisés dans la modélisation des mécanismes neuronaux qui ont besoin d'implanter des processus neuronaux sélectifs et persistants. Ces modèles de champs sont capables à la fois de s'activer localement, dans les zones correspondant aux sources de stimulation les plus pertinentes, et de suivre ces stimuli pendant leur évolution. Les résultats des scénarios auxiliaires mettent en évidence aussi la vitesse d'adaptation des réponses des champs, le modèle CNDPE ayant une latence d'adaptation très faible, ce qui lui confère une meilleure mobilité par rapport à la distribution de stimulation (pour les conditions de paramétrage considérées dans nos expériences). Le même groupe d'expériences montre également la capacité de ces modèles de champ à saisir correctement les zones pertinentes de stimulation externe, même quand plusieurs (un nombre fini) réponses équivalentes sont possibles. La même conclusion reste valable quand il s'agit d'explorer une plage continue de réponses équiprobables de haute intensité. Toutefois, l'activité de chacun des champs reste faible, dans le contexte clé défini par une faible stimulation externe. Par conséquent, dans ce cas, la propriété de sélectivité exigée par $(P \wedge Q)$ n'est plus respectée. Cet aspect s'avère crucial dans l'implantation des processus d'auto-organisation. Les deux modèles de champ ne réussissent pas à obtenir les résultats attendus dans le cas unidimensionnel, et s'ils les réussissent toutefois dans le cas bidimensionnel, leurs comportements globaux ne sont pas robustes : en changeant quelque peu un facteur externe au champ (la fonction de reconnaissance des prototypes), aucun d'eux n'est capable d'accomplir de manière satisfaisante l'auto-organisation des prototypes. Étant donné l'intérêt central de l'auto-organisation dans nos recherches, détaillons ci-dessous les causes de l'échec des deux modèles de champ, examinés dans ce type de scénarios.

Pour commencer cette analyse, nous rappelons que dans notre implantation d'inspiration corticale du processus d'auto-organisation, le champ neuronal dynamique a le rôle de mettre en œuvre une compétition de type WTM. La réponse du champ module la règle d'apprentissage des valeurs des prototypes (c'est-à-dire leurs ajustements graduels pour réaliser la quantification vectorielle du domaine d'entrée Ξ).

Au tout début de chaque expérience, les prototypes ont des valeurs aléatoires. De plus, ces valeurs n'appartiennent pas nécessairement à Ξ . Pour le scénario d'auto-organisation unidimensionnel, Ξ est une bande circulaire (couronne) des points du plan $\xi = (\xi_1, \xi_2)$, alors que les prototypes $\omega = (\omega_1, \omega_2)$ sont initialisés uniformément sur la région (le carré) du plan englobant Ξ (voir la figure 5.17 (a)). Une fois qu'un échantillon $\xi \in \Xi$ est présenté au module de reconnaissance de l'architecture d'auto-organisation, chaque prototype formule une réponse proportionnelle à son rapprochement de ξ (en terme de distance euclidienne), dans l'espace bidimensionnel. Initialement, peu de prototypes ont la chance de se trouver près de l'échantillon ξ , et il est peu probable que ces prototypes-là correspondent à des unités voisines dans X . Par conséquent, les éventuelles activités fortes qui stimulent le champ au début du processus d'auto-organisation sont éparpillées sur X , et agissent plutôt comme un bruit faible. Comme nous l'avons constaté empiriquement en examinant les scénarios d'exploration (revoir la figure 5.15 (b)), les deux modèles classiques de champ (CNDA et CNDPE) ne répondent pas de manière sélective dans des conditions de faible stimulation. Le champ reste « inactif », tant que les stimuli externes restent principalement faibles, avec seulement quelques pics éparpillés d'activité. Pendant tout ce temps, les valeurs des prototypes s'ajustent très peu, et toutes de la même manière, étant donné que leur ajustement est dicté directement par l'activité du champ. Au cours du temps, ce phénomène induit un rapprochement des valeurs des prototypes de la moyenne des échantillons ξ présentés. Dans le cas de notre scénario, cette valeur moyenne correspond au centre de la couronne Ξ . L'effet « d'attraction » des prototypes vers cette valeur peut être observé pour les deux modèles de champ, sur la figure 5.17 (b), qui montre une étape intermédiaire du processus d'auto-organisation unidimensionnelle.

Le fait que de plus en plus de prototypes aient des valeurs qui soient proches implique le fait qu'un échantillon, également proche de ces valeurs, induit un nombre significatif de réponses de reconnaissance $i(x, t)$ fortes. Cette modification lente de la distribution $\{i(x, t)\}_{x \in X}$ induit une réponse de plus en plus forte de la part du champ. Ainsi, comme le montrent les expériences des scénarios d'attention, la distribution u est capable de former dans ces conditions, une réponse locale de type bulle d'activité. Par conséquent, les prototypes qui correspondent à cette zone d'activité du champ réajustent leurs valeurs. Ce processus entraîne l'éloignement de ces valeurs de la moyenne des valeurs des prototypes, chacune essayant de se différencier, afin de réaliser la quantification vectorielle de l'espace Ξ . Néanmoins, les autres prototypes avec des valeurs similaires, continuent à garder ces valeurs. Au cours du temps, ce phénomène empêche un déploiement complet des prototypes sur l'ensemble de l'espace d'entrée Ξ , fait constaté expérimentalement et illustré sur la figure 5.17 (c). Notons que la topologie non convexe de l'espace d'entrée contribue à cet effet indésirable. Les différents échantillons « attirent » les prototypes qui existent dans leur proximité (dans la métrique de l'espace Ξ), mais en même temps, ces prototypes sont « attirés » entre eux, par leur proximité de connexion (donc par la métrique de l'espace X). Comme nous pouvons le constater empiriquement, dans le contexte unidimensionnel considéré pour le scénario d'auto-organisation, ces deux contraintes de proximité ne peuvent plus être satisfaites pour tous les prototypes. Vis-à-vis de ces résultats expérimentaux, remarquons aussi que, grâce à sa vitesse supérieure d'adaptation des bulles d'activité, le champ CNDPE réussit à déployer un nombre sensiblement plus grand de prototypes. Cependant, même dans ce cas, le processus d'auto-organisation ne converge pas vers un résultat final satisfaisant.

Un aspect supplémentaire que nous observons dans cette expérience d'auto-organisation, se réfère aux histogrammes présentées dans la figure 5.18. Ces histogrammes indiquent une qualité satisfaisante de la dynamique du champ, à la fois pour les modèles CNDA et CNDPE. Afin de comprendre ce résultat, nous rappelons que les histogrammes représentent, par définition, l'expression d'un caractère moyen de l'évolution de l'événement analysé. D'après notre analyse, les

deux champs ont une faible sélectivité au début de l'expérience. Pour vérifier cette affirmation, nous avons retracé les histogrammes de qualité pour les deux modèles de champ, exclusivement pour les 400 premiers échantillons présentés dans le scénario d'auto-organisation unidimensionnelle. Comme nous pouvons le constater dans la figure 5.21, les deux champs montrent un faible caractère de sélectivité au début de cette expérience. Cependant, comme nous l'avons précisé ci-dessus, après plusieurs pas de simulation (c'est-à-dire plusieurs échantillons présentés), la stimulation du champ augmente, et avec elle, le caractère sélectif des deux champs examinés. Par conséquent, *en moyenne*, les deux champs montrent en effet une sélectivité satisfaisante par rapport à leur stimulation externe, fait illustré graphiquement par les histogrammes de la figure 5.18.

De cette analyse détaillée nous pouvons conclure que le caractère sélectif d'un champ neuronal est une qualité essentielle pour assurer la réussite du processus d'auto-organisation, et il est déterminant qu'elle se manifeste *dès le début* de ce processus. A l'égard de cet aspect, nous constatons qu'une « simple » mesure statistique du processus d'auto-organisation n'est pas suffisante pour évaluer la performance d'implantation de ce processus. Cet aspect montre ainsi la nécessité d'examiner individuellement le comportement des champs neuronaux, dans des scénarios clé, capables de mettre en évidence les aspects dynamiques intrinsèques de ces modèles neuronaux.

Basé sur ces premières conclusions, nous analysons ensuite les résultats issus du deuxième scénario d'auto-organisation considéré. Cette fois-ci, le domaine d'entrée est unidimensionnel ($\Xi = [-\frac{\pi}{2}, \frac{\pi}{2}]$), mais l'ensemble X des unités de champ est bidimensionnel. Le fait que Ξ est dans ce cas contigu et convexe, et que X est bidimensionnel, augmente la probabilité que dès le début du processus d'auto-organisation, plusieurs prototypes voisins dans l'espace Ξ soient aussi voisins dans l'espace X (en fait les unités du champ qui leur correspondent). Ceci favorise l'apparition des réponses sélectives dans le champ neuronal de la couche corticale de l'architecture d'auto-organisation, plus tôt que dans le cas précédemment étudié. Comme affirmé, une bonne sélectivité du champ dans ces conditions lui confère la réussite de réorganisation des prototypes, comme montré dans la figure 5.19. Toutefois, reconsidérant le même scénario, mais cette fois-ci avec un petit changement dans la forme de la fonction de reconnaissance des prototypes, les réseaux des prototypes ne convergent plus vers des états désirables en ce qui concerne leurs topologies. La particularité de cette nouvelle fonction de reconnaissance étant qu'elle est légèrement plus sélective que la première utilisée, elle n'entraîne plus l'apparition des zones locales de forte stimulation. De manière similaire au cas du scénario unidimensionnel, le champ est stimulé dans ces conditions seulement avec des activités fortes plus rares et plus éparées, ce qui se prouve être en dessus du seuil de sélectivité des deux champs étudiés. Comme dans le cas unidimensionnel d'auto-organisation, les comportements des champs se ressemblent : les valeurs des prototypes se rapprochent de la moyenne de l'ensemble Ξ (aussi 0), mais la topologie convexe de Ξ favorise une convergence rapide vers la même valeur pour tous les prototypes dans le cas du modèle CNDA (voir la figure 5.19 (a)). Le modèle CNDPE réussit, tout comme dans le premier scénario analysé, à atteindre des partitions de prototypes réorganisées topologiquement, mais malgré la meilleure adaptation de ce modèle aux changements des stimulations externes, les résultats finaux restent peu satisfaisants (voir la figure 5.19 (b)).

Nous soulignons aussi que les deux expériences réalisées ici pour le même scénario bidimensionnel montrent également le manque de robustesse des modèles de champ, vis-à-vis de certains facteurs de modélisation (ici la fonction de reconnaissance des prototypes).

Comme conclusion finale, les résultats de l'évaluation empirique des champs classiques d'Amari et de Pinto et Ermentrout en font des candidats à la modélisation des processus de compétition, de sélection et de persistance neuronale, mais pas à la modélisation des processus d'auto-

organisation, tels que nous les avons conçus dans notre travail de thèse. De plus, le fait que les champs de référence montrent empiriquement la capacité de mener correctement un tel processus d'auto-organisation nous offre un argument pour rechercher un modèle de champ capable d'implanter proprement le module de sélection de l'architecture d'auto-organisation proposée, tout en respectant les principes de calcul neuronal énoncés dans paragraphe 2.5. Cet aspect est abordé dans le prochain chapitre, dans lequel nous proposons un nouveau formalisme de champ neuronal, capable en effet d'accomplir ces desiderata.

5.4 Méthodologie d'ajustement fin des paramètres d'un modèle de champ

Dans cette section, nous proposons d'utiliser l'instrument de mesure introduit dans la section 5.2.2, ainsi que la plate-forme d'évaluation empirique présentée dans la section 5.2.5, pour implémenter une procédure d'ajustement des paramètres d'un modèle de champ neuronal. Cette procédure vise à trouver les paramètres d'un modèle de champ neuronal, capables d'optimiser son comportement, vis-à-vis les propriétés dynamiques ($P \wedge Q$) présentées dans la section 5.2.1. Cette procédure est composée de plusieurs étapes, qui sont résumées ci-dessous. Nous illustrons cette approche d'ajustement des paramètres en considérant le modèle CNDA, présenté dans la section 3.2.2.

Dans un premier temps, nous déclarons comme *paramètres libres* un certain nombre de paramètres du modèle, que nous allons faire varier au cours de cette procédure. Ensuite, nous précisons les plages de valeurs pour ces paramètres libres. À cause du fait que la procédure décrite ici repose sur une simulation numérique, il est nécessaire que les valeurs possibles des paramètres libres soient des valeurs discrètes.

Pour le modèle CNDA, défini par les équations 3.3 et 3.15, nous considérons les paramètres fixes suivants :

$$\tau = 0.3^{-1}, h = 0, a = 0.35, b = 0,$$

et le triplet $(\alpha_1, \alpha_2, \alpha_3)$ des paramètres libres du modèle, tel que :

$$\begin{aligned} F(x) &= 1/(1 + e^{-\alpha_1(2x-1)}) \\ A^+ &= \alpha_2 \\ A^- &= \alpha_3 \end{aligned}$$

où $\sigma(x) = (F(x) - F(0))/(F(1) - F(0))$ est la fonction sigmoïde qui apparaît dans l'équation 3.15.

Nous considérons les plages de valeurs discrètes suivantes, pour les trois paramètres libres du modèle CNDA : $\alpha_1 \in [1, 9]$ (5 valeurs), $\alpha_2 \in [0.2, 1.2]$ (11 valeurs), $\alpha_3 \in [0, 0.3]$ (31 valeurs).

Dans la deuxième étape de la procédure d'ajustement des paramètres, il est nécessaire de définir un scénario i pour lequel on souhaite optimiser le comportement du champ. Les paragraphes précédents de ce chapitre nous ont montré empiriquement que le succès d'utilisation d'un champ neuronal dans la modélisation des processus d'auto-organisation dépend de son comportement dynamique relatif aux scénarios d'attention et d'exploration. Afin d'optimiser le comportement du modèle CNDA dans les deux contextes de stimulation externe, nous proposons comme scénario i pour cette procédure d'ajustement de paramètres, le scénario présenté dans la figure 5.22 (a). Nous pouvons affirmer que ce scénario met en évidence les situations clé de stimulation du champ, rencontrées pendant le déroulement du processus d'auto-organisation. Ce scénario vise à tester si le champ est capable de générer une bulle d'activité neuronale dans des conditions de faible stimulation (premier tiers du scénario), ainsi que dans des conditions de stimulation

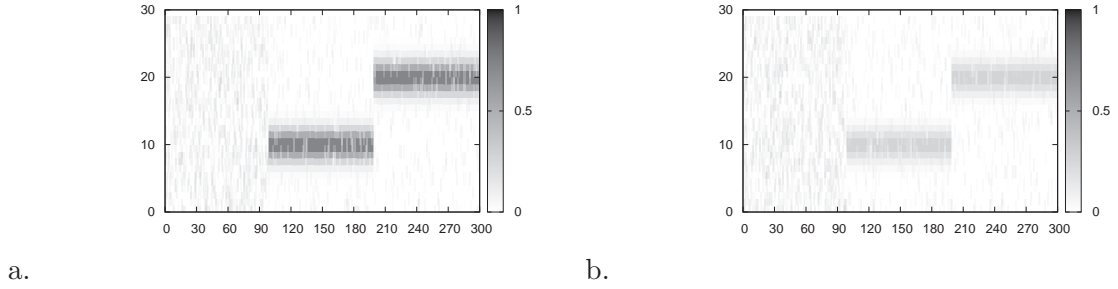


FIG. 5.22 – Deux scénarios similaires, conçus pour prendre en compte la dynamique de plusieurs situations clé de stimulation, rencontrées pendant le déroulement du processus d'auto-organisation. Les premiers 100 pas de simulation du scénario i représentent un bruit aléatoire, les 100 pas suivants représentent un stimulus gaussien local, et les derniers 100 pas de simulation représentent toujours un stimulus gaussien, mais placé à un autre endroit de la distribution i . Les deux scénarios (a) et (b) diffèrent par l'amplitude des stimuli gaussiens.

locale forte (le reste du scénario), et si le champ est capable de modifier sa réponse, en accord avec les changements dynamiques, intervenus dans la distribution des stimuli externes.

Dans la troisième étape de la procédure d'optimisation, la capacité du champ à satisfaire les propriétés $(P \wedge Q)$ vis-à-vis du scénario i considéré, est évaluée en utilisant la plate-forme d'évaluation empirique décrite dans la section 5.2.5. Pour chaque combinaison de valeurs des paramètres, la moyenne des valeurs $\Delta_i^B(u)$ enregistrées pendant le déroulement de l'expérience est calculée, et considérée comme l'erreur moyenne du champ à satisfaire les critères imposés par les deux propriétés dynamiques. Notons que $\Delta = 0$ pour un champ respectant entièrement les propriétés $(P \wedge Q)$, et $\Delta = 1$ si le champ manifeste une absence totale d'activation vis-à-vis du scénario considéré. Notons aussi que la valeur de Δ peut même dépasser le seuil unitaire, si l'activité du champ est évaluée comme étant considérablement insatisfaisante par rapport aux propriétés $(P \wedge Q)$. Dans le cas de notre exemple, cette valeur est symbolisée par $\Delta(\alpha_1, \alpha_2, \alpha_3)$.

La dernière étape de la procédure décrite ici comprend une recherche exhaustive, pour trouver la combinaison des valeurs des paramètres libres du modèle de champ, qui minimise la valeur moyenne Δ ⁵. Pour le cas illustré ici, la valeur minimale de $\Delta(\alpha_1, \alpha_2, \alpha_3)$ est 0.40, et elle est atteinte pour le triplet $\alpha_1^* = 1, \alpha_2^* = 0.6, \alpha_3^* = 0.09$. Remarquons néanmoins que, si on impose $\alpha_1 > 1$ et on répète la procédure décrite ci-dessus, le minimum de $\Delta(\alpha_1, \alpha_2, \alpha_3)$ est atteint chaque fois pour $\alpha_2 = 0.3, \alpha_3 = 0.03$. Le fait que $\Delta(1, \alpha_2^*, \alpha_3^*) = 0.40$ et $\Delta(\alpha_1, 0.3, 0.03) \approx 0.49$, pour les autres valeurs de α_1 relève l'influence des contributions non-linéaires dans le comportement dynamique du champ. Afin de mieux percevoir l'influence des paramètres dans le comportement du modèle CNDA, la figure 5.23 montre graphiquement les valeurs de $\Delta(\alpha_1, \alpha_2, \alpha_3)$, pour $\alpha_1 = \alpha_1^* = 1$.

La figure 5.24 présente plusieurs comportements typiques, exhibés par le modèle CNDA pour $\alpha_1 = \alpha_1^* = 1$, et différents cas de paramétrage pour α_2 et α_3 . Comme nous pouvons l'observer sur cette figure (voir le graphique (A)), le comportement considéré optimal du champ (selon la

⁵Notons que la complexité de calcul de la procédure décrite ici peut être réduite si, au lieu d'employer une recherche exhaustive pour trouver la combinaison des valeurs des paramètres qui optimisent le comportement du champ neuronal, on utilise un algorithme de recherche de descente de gradient, similaire à l'algorithme présenté dans la figure 5.10.

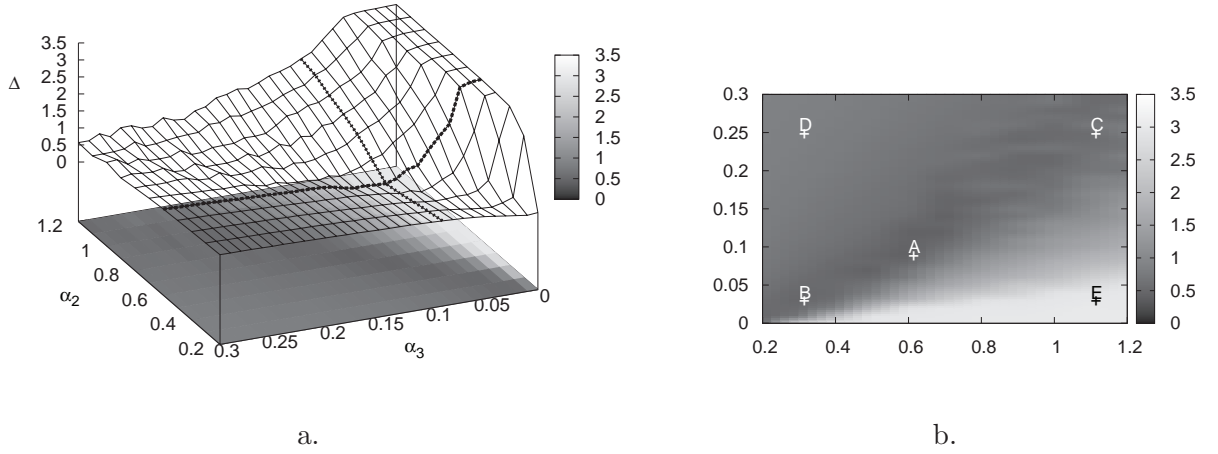


FIG. 5.23 – Les valeurs de l'indicateur $\Delta(1, \alpha_2, \alpha_3)$, obtenues suivant la procédure d'ajustement de paramètres du modèle CNDA, pour le scénario considéré dans la figure 5.22 (a). Les paramètres α_2^* et α_3^* pour lesquels $\Delta(1, \alpha_2, \alpha_3)$ atteint le minimum se trouvent à l'intersection des deux lignes tracées sur la graphique (a). L'illustration est reprise sur le graphique (b), qui représente en outre plusieurs points significatifs du comportement dynamique du champ. Le point A de ce graphique correspond au triplet $(\alpha_1^*, \alpha_2^*, \alpha_3^*)$, pour lequel la valeur $\Delta(1, \alpha_2, \alpha_3)$ atteint le minimum. Suivre le texte pour plus de détails.

valeur minimale de l'indicateur Δ) répond aux attentes formulées vis-à-vis de sa dynamique, par rapport au scénario considéré. Dans d'autres conditions de paramétrage du modèle CNDA, les comportements dynamiques du champ diffèrent considérablement. Ainsi, pour le point B symbolisé sur la figure 5.23 (b), le champ formule une réponse similaire à la distribution d'entrée i perçue (voir la figure 5.24 (B)). Le point C de la figure 5.23 (b) montre un champ qui est capable de formuler une réponse locale de type bulle d'activité dans des conditions de faible stimulation externe, mais qui est incapable d'adapter sa réponse face aux changements brusques, intervenus dans la distribution de cette stimulation (voir la figure 5.24 (C)). La réponse du champ correspondant au point D de la figure 5.23 (b) est similaire à la réponse du champ représenté par le point B de la même figure (voir la figure 5.24 (D)). Enfin, pour certaines valeurs des paramètres α_2 et α_3 , le champ exhibe une réponse entièrement saturée, c'est-à-dire une activité ayant la valeur $u(x) = 1$, pour chacune de ses unités $x \in X$. Cette réponse, correspondant au point (E) de la figure 5.23 (b), n'est pas illustrée sur la figure 5.24.

Si les paramètres $(\alpha_1^*, \alpha_2^*, \alpha_3^*)$ qui minimisent la valeur de l'indicateur $\Delta(\alpha_1, \alpha_2, \alpha_3)$ correspondent à un champ neuronal d'Amari ayant un comportement souhaitable (comme l'illustre la figure 5.24 (A)), nous sommes intéressés à examiner dans la suite la robustesse de ce comportement. Par conséquent, dans un premier temps, nous examinons l'influence de l'évaluation asynchrone, ainsi que celle du bruit de la source de stimulation, sur le comportement global du champ. Pour ceci, la capacité du champ optimal (c'est-à-dire le CNDA avec $\alpha_1 = \alpha_1^*, \alpha_2 = \alpha_2^*, \alpha_3 = \alpha_3^*$) est réévaluée 50 fois, pour le même scénario illustré sur la figure 5.22 (a). L'évolution dynamique de l'indicateur $\bar{\Delta}_i^B(u)$ pour le comportement moyen du champ optimal dans ces conditions, est représentée graphiquement sur la figure 5.25 (a). Comme nous pouvons l'observer sur cette figure,

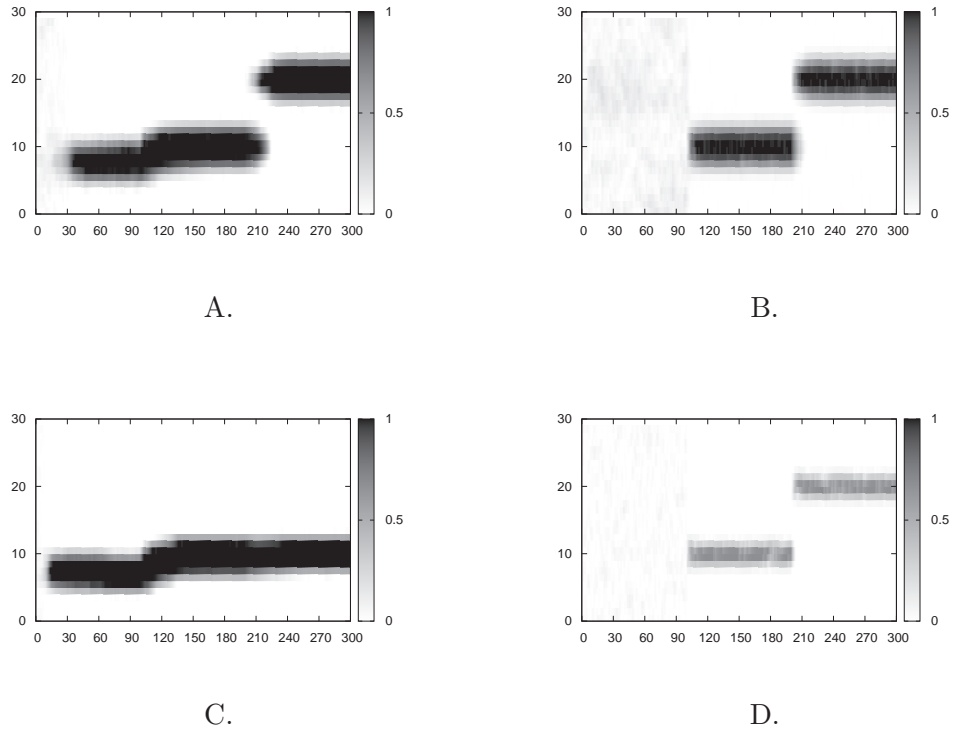


FIG. 5.24 – Plusieurs comportements représentatifs du modèle CNDA, obtenus pour $\alpha_1 = \alpha_1^* = 1$, et différentes valeurs des paramètres α_2 et α_3 , comme réponse au scénario de stimulation présenté dans la figure 5.22 (a). Ces cas de paramétrage correspondent aux points A-D, marqués sur la figure 5.23 (b). Le point E de la figure mentionnée correspond à une réponse du champ avec un niveau d'activité $u(x, t) = 1, \forall x, t$, mais ce comportement du champ n'est pas illustré sur cette figure.

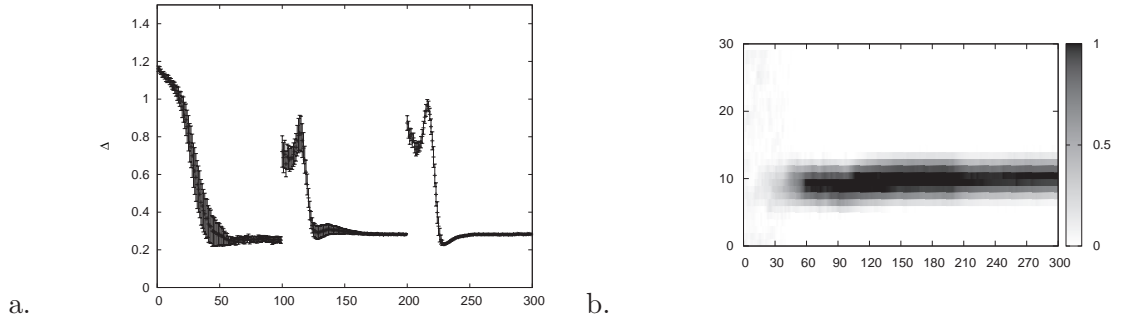


FIG. 5.25 – a. La moyenne et la déviation standard de l'indicateur $\bar{\Delta}_i^B(u)$, obtenues après l'évaluation répétée (50 fois) du comportement du champ optimal, vis-à-vis du scénario présenté dans la figure 5.22 (a). b. Le comportement dynamique du même champ optimal, vis-à-vis du scénario considéré dans la figure 5.22 (b).

le champ est capable de générer une bulle d'excitation dans des conditions de faible stimulation, et d'adapter sa réponse quand cette stimulation change son profil, fait confirmé par la convergence de l'indicateur $\bar{\Delta}_i^B(u)$ vers des valeurs inférieures, après chaque changement significatif intervenu dans la distribution de stimulation i du champ. Cet aspect confirme la robustesse du champ optimal par rapport aux sources de bruits ou aux effets de l'évaluation asynchrone.

Dans un second temps, nous examinons le comportement du champ vis-à-vis d'un nouveau scénario de stimulation, considéré dans la figure 5.22 (b). Comme nous pouvons le remarquer en analysant la figure 5.25 (b), le champ est incapable d'adapter cette fois-ci sa réponse, face au nouveau scénario. Cet aspect montre que le comportement du champ optimal n'est pas robuste, par rapport aux scénarios de stimulation considérés.

En conclusion, cette procédure nous confirme empiriquement la difficulté pratique pour trouver les paramètres appropriés pour que le modèle de champ d'Amari soit capable de satisfaire, d'une manière robuste, les propriétés $(P \wedge Q)$, dans différents contextes de stimulation. Cependant, cette procédure peut être utilisée avec succès pour ajuster les paramètres d'un modèle de champ neuronal, afin de rendre le champ capable d'exhiber un caractère sélectif dans les scénarios individuels, comme les scénarios d'attention.

5.5 Discussion

Dans ce chapitre nous avons proposé une approche d'analyse empirique, pour examiner la capacité des champs neuronaux classiques à participer à la modélisation de différents processus neuronaux, et notamment ceux liés à l'auto-organisation corticale. Cette approche, issue d'une nécessité pratique d'utilisation des champs neuronaux, s'est concrétisée dans l'analyse de plusieurs comportements souhaitables d'un champ neuronal, formalisés sous l'expression de deux propriétés dynamiques, (P) et (Q).

Grâce à un instrument d'évaluation numérique, conçu pour quantifier la satisfaction des propriétés $(P \wedge Q)$ par un champ neuronal quelconque, nous avons examiné empiriquement la capacité de champs neuronaux classiques (CNDA et CNDPE) à respecter ces propriétés dynamiques, dans différents contextes pratiques de stimulation. Les résultats expérimentaux de cette procédure d'évaluation ont pu confirmer que ces modèles de champ sont appropriés pour être utilisés dans la modélisation des processus de mémoire ou d'attention.

Cependant, une étude expérimentale étendue du comportement de ces champs nous a révélé

l'incapacité de ces modèles à accomplir des résultats satisfaisants dans les scénarios d'auto-organisation. Cet instrument nous a permis ensuite d'examiner plus en détail la cause de ces résultats insatisfaisants. Ainsi, nous avons pu mettre en évidence que la qualité de ces résultats est liée à la capacité du champ à générer une bulle d'activité neuronale dans des conditions de faible stimulation externe. Afin d'améliorer le comportement des champs dans ce type de scénarios, nous avons conçu une procédure d'ajustement des paramètres du modèle. En résumé, les résultats expérimentaux de cette approche nous ont montré que le comportement des champs classiques dans les conditions de faible stimulation externe s'inscrit dans un des trois cas de figure : (1) le champ est incapable de générer une bulle d'activité ; (2) le champ est capable de générer une bulle d'activité, mais cette bulle demeure au même endroit du champ, une fois qu'elle a été générée, et le champ est incapable d'adapter sa réponse vis-à-vis d'autres sources de stimulation ; (3) le champ est capable de générer une bulle d'activité en conditions de faible stimulation et il est capable d'adapter sa réponse seulement dans certains autres cas de stimulation, ce qui implique l'instabilité de ce comportement par rapport aux conditions de stimulation.

En conclusion, nous pouvons affirmer que pour faire usage des propriétés dynamiques des champs neuronaux dans la modélisation des processus d'auto-organisation, il est nécessaire de proposer une alternative aux modèles classiques des champs neuronaux, capable de contourner les problèmes de fond énumérés ci-dessus. Plus spécifiquement, ce modèle nécessite l'implantation d'un mécanisme dynamique robuste, capable à la fois de générer une bulle d'activité, indépendamment de l'amplitude des stimuli perçus par le champ neuronal, et de limiter la durée de vie des bulles qui sont générées à partir d'une faible stimulation externe.

Dans le chapitre suivant, nous abordons la problématique de modélisation d'un tel mécanisme, en exploitant les avantages illustrés par les modèles classiques examinés, mais en s'inspirant aussi d'autres approches de modélisation qui ont été présentées dans les chapitres précédents.

Nota : les travaux décrits dans ce chapitre ont fait l'objet des articles 1, 2 et 5, énumérés dans l'annexe A.

Chapitre 6

Champs neuronaux dynamiques pour l'auto-organisation

Intéressé par la modélisation des processus d'auto-organisation s'inspirant du traitement cortical de l'information, nous avons proposé dans 2.6 une architecture de calcul neuronale, qui intègre les principaux aspects fonctionnels de ce processus. Nous avons ensuite prêché dans 2.5 pour une mise en œuvre de cette architecture qui respecte certains principes de modélisation de calcul neuronal. Conduit d'une part par ces principes, et d'une autre part par une approche de modélisation d'inspiration corticale unifiée et générique, nous avons choisi le paradigme des champs neuronaux dynamiques comme modèle de calcul neuronal pour implémenter le module de sélection de l'architecture d'auto-organisation proposée.

Une analyse empirique détaillée présentée dans le chapitre précédent nous a permis de constater certaines déficiences comportementales des modèles classiques de champs neuronaux, quand il s'agit d'implémenter une architecture d'auto-organisation. Dans 5.3.6 nous avons conclu que l'incapacité de ces modèles à garder un caractère dynamique sélectif dans les contextes de faible stimulation externe empêche de les considérer comme appropriés pour notre objectif de modélisation. En même temps, grâce aux outils théoriques développés et présentés dans ce même chapitre, nous avons pu observer que les champs capables de maintenir ce comportement dynamique dans toutes les conditions de stimulation (les champs ainsi nommés de référence), réussissent en effet à atteindre ce but avec des résultats satisfaisants. Néanmoins, comme nous l'avons précisé, ces champs de référence sont le fruit d'une procédure de calcul en désaccord avec les principes de calcul neuronal promus tout au long de notre discours.

En résumé, d'une part l'incapacité des champs classiques à assurer des résultats satisfaisants d'auto-organisation, et d'autre part l'incapacité des champs de référence à respecter les principes de modélisation que nous avons considérés, nous motivent une fois de plus à poursuivre la recherche d'un véritable modèle de champ neuronal, capable de produire un effet d'auto-organisation.

Comme réponse à ces remarques, dans ce chapitre nous proposons un nouveau modèle champ, conçu pour améliorer le comportement des modèles classiques de champ dans notre contexte. Le chapitre est organisé comme suit. La première section décrit le formalisme du modèle proposé, la deuxième présente les résultats de la modélisation du processus d'auto-organisation, la troisième examine en détail le comportement du champ dans d'autres scénarios, et la dernière section conclut sur les aspects qualitatifs et émergents de ce nouveau modèle.

6.1 Champs neuronaux dynamiques rétro-inhibés (CNDRI)

La diversité des approches de modélisation et d'exploitation pratique, présentées dans les chapitres 3 et 4, ont constitué de fortes sources d'inspiration pour notre propre démarche de modélisation neuronale. Parmi celles-ci, l'approche d'Erlhagen et celle de Pinto et Ermentrout, nous ont inspiré le plus dans la conception du modèle que nous proposons.

Dans [36], Erlhagen présente une approche de modélisation de mémoire à court terme utilisant un modèle de champ neuronal dynamique. Pour rendre fonctionnelle une logique de type mémoire à court terme, Erlhagen affirme que le champ doit avoir la capacité d'empêcher la persistance des bulles d'excitation en cas d'absence de stimulation. Pour cela, il propose dans la [36] (revoir aussi la section 3.2.6) un ajustement du modèle d'Amari qui consiste en la modification dynamique du niveau global de repos de l'activité du champ, afin de déclencher l'évanouissement des bulles du champ qui ne sont plus entretenues par une source de stimulation externe. Même si la condition imposée par Erlhagen se place en totale opposition avec les comportements dynamiques que nous souhaitons, l'approche employée par Erlhagen pour améliorer la dynamique du modèle de champ d'Amari constitue un exemple notable d'adaptation de conception d'un modèle théorique, et c'est ce qui a constitué une inspiration de départ pour notre propre approche de modélisation de champ. Nous soulignons également que cette approche promue par Erlhagen est issue, comme dans notre cas, d'un besoin essentiellement pratique d'amélioration dynamique du comportement des champs neuronaux.

Notre deuxième source majeure d'inspiration se réfère à l'approche de modélisation de Pinto et Ermentrout, décrite dans [92]. Celle-ci, détaillée dans 3.2.5, étend les capacités dynamiques d'un champ en intégrant différentes contributions neuronales dans sa modélisation. Ces contributions, et notamment celles inhibitrices de feedback, modélisées par les équations 3.26 et 3.27, pourraient être vues comme un mécanisme de contrôle dynamique du comportement local du champ. Malgré les résultats insatisfaisants de ce modèle dans les scénarios d'auto-organisation, nous considérons néanmoins que le principe de modélisation proposé par Pinto et Ermentrout mérite à être reconsidéré et exploité également dans ces contextes. De plus, la mise en évidence des déficiences dynamiques de ce modèle par l'analyse empirique décrite dans le chapitre précédent constitue un avantage qui nous permet de repenser l'utilisation de ce principe, afin d'améliorer le comportement des champs dans les contextes d'auto-organisation.

Également inspirés par l'approche de Wilson et Cowan, nous considérons une séparation entre les noyaux qui modélisent les poids des connexions latérales excitatrices et inhibitrices :

$$w^+(r) = e^{-ar^2} \quad (6.1)$$

$$w^-(r) = e^{-br^2} - e^{-cr^2} \quad (6.2)$$

Afin de maintenir les deux noyaux positifs (pour bien différencier les valeurs absolues respectivement des contributions excitatrices et inhibitrices), nous imposons $b < c$. Les deux fonctions w^+ et w^- sont représentées aussi dans la figure 6.1. Cette modélisation séparée des influences excitatrices et inhibitrices nous permet, sans compter l'argument de la plausibilité biologique, de garder un meilleur contrôle sur la contribution de chacune de ces influences sur la dynamique de l'activité globale du champ. Toutefois, contrairement à l'approche de Wilson et Cowan, les deux contributions sont incluses dans une unique équation, comme dans le formalisme d'Amari.

Soit $\mathcal{E}(x, t)$ et $\mathcal{I}(x, t)$, respectivement les influences instantanées excitatrices et inhibitrices,

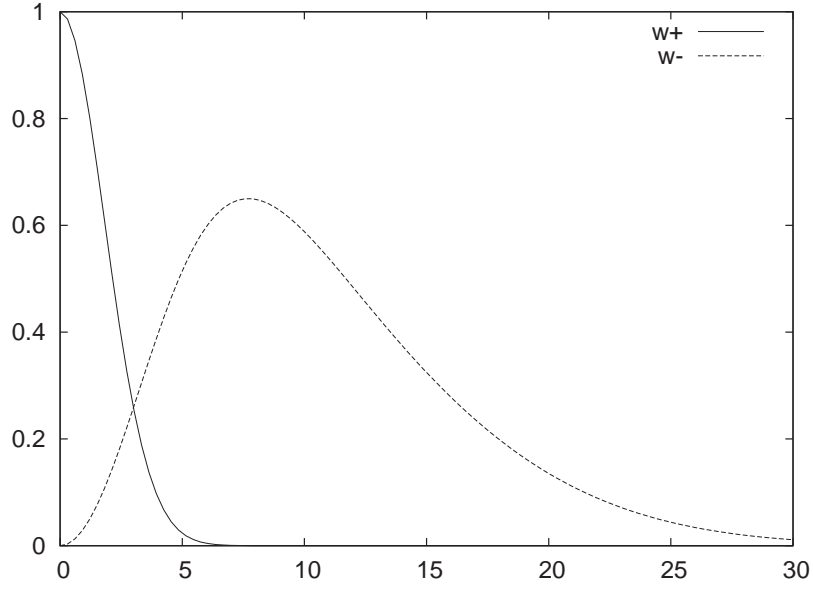


FIG. 6.1 – Les noyaux des poids des connexions synaptiques excitatrices et inhibitrices, définis par les fonctions w^+ et w^- .

ressenties par chaque unité x du champ :

$$\mathcal{E}(x, t) = \sigma^+ \left(\int_{x'} w^+(|x - x'|) f(u(x', t)) dx' \right) \quad (6.3)$$

$$\mathcal{I}(x, t) = \sigma^- \left(\int_{x'} w^-(|x - x'|) f(u(x', t)) dx' \right) \quad (6.4)$$

où σ^+ , σ^- et f sont des fonctions sigmoïdes.

Avec ces notations, nous pouvons introduire le formalisme du nouveau modèle de champ, décrit par les équations suivantes :

$$\tau \frac{du(x, t)}{dt} = i(x, t) + \alpha \mathcal{E}(x, t) - \beta \mathcal{I}(x, t) - \gamma g(i, v) \quad (6.5)$$

$$\frac{dv(x, t)}{dt} = h(\mathcal{E}(x, t)) \quad (6.6)$$

où α, β, γ sont des constantes réels positives. τ est une constante de temps, comme dans d'autres formalismes de champs, qui module le taux de croissance de l'activité u . La variable v modélise une activité auxiliaire qui suit l'évolution dynamique de l'excitation locale \mathcal{E} . Enfin, g et h sont des fonctions sigmoïdales, ayant des formes similaires à celles montrées dans la figure 6.2, modulant respectivement la dynamique d'une influence de feedback de nature inhibitrice, appelée *rétro-inhibition*, et la dynamique de l'activité auxiliaire v .

Nous mentionnons que, telles qu'elles sont définies, les équations du champ permettent aux variables u et v de dépasser les valeurs 0 ou 1. Pour empêcher le dépassement de ces seuils, les valeurs de u et v sont saturées explicitement entre 0 et 1.

Les étapes successives de l'évolution dynamique du modèle CNDRI, en réponse à une faible source de stimulation externe, sont illustrées dans la figure 6.3.

Remarquons d'abord que l'activité v est une « variante retardée » de l'excitation locale, \mathcal{E} . En effet, si $\mathcal{E}(x, t)$ ne dépasse pas un certain seuil, la fonction $h(\mathcal{E}(x, t))$ reste négative, ce

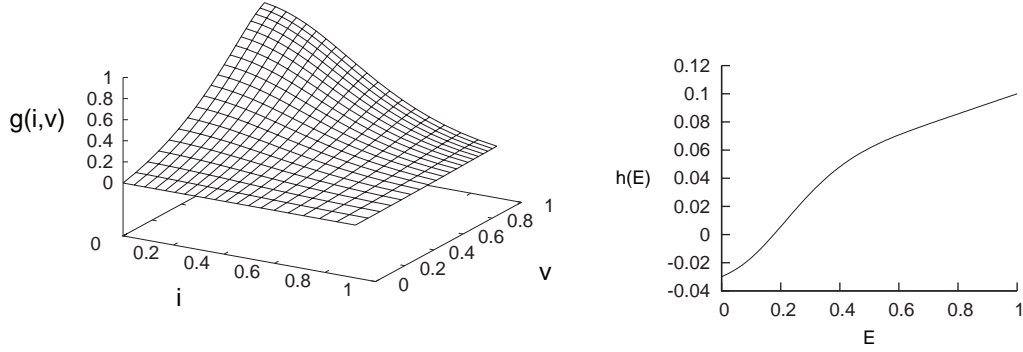


FIG. 6.2 – Les fonctions g (gauche) et h (droite), introduites dans les équations 6.5 et 6.6, entraînant d'une part la dynamique de rétro-inhibition, et d'une autre part celle de l'activité auxiliaire v du champ.

qui entraîne une décroissance vers 0 de l'activité v . La valeur de v se maintient nulle jusqu'au moment où $\mathcal{E}(x, t)$ est suffisamment grande pour rendre la fonction $h(\mathcal{E}(x, t))$ positive. À partir de ce moment, si $\mathcal{E}(x, t)$ continue à croître, l'activité v suit la même tendance, mais avec un rythme plus lent. Au final, v atteint le niveau de saturation avec un certain délai après $\mathcal{E}(x, t)$.

Une deuxième constatation se réfère au fait que, si $g(i, v) \approx 0$, c'est-à-dire $i \approx 1$ ou $v \approx 0$, l'activité u du champ est gouvernée par une dynamique similaire à celle du modèle de champ d'Amari. En effet, dans ces conditions, les seuls termes intervenant dans la dynamique de u sont : la stimulation externe i et la contribution des activités neuronales voisines. Remarquons toutefois que, contrairement au modèle d'Amari, le terme $-u(x, t)$, qui entraîne la décroissance exponentielle de u quand les autres termes de l'équation deviennent nuls, est absent. Cependant, la décroissance de l'activité u est implicitement assurée par la dynamique du terme $-g(i, v)$, qui devient actif si $v \approx 1$ (fait induit par une forte excitation locale), et $i \approx 0$ (ce qui implique une faible activité de stimulation externe).

Comme nous pouvons le constater en décrivant les aspects dynamiques énumérés ci-dessus, le champ, au tout début de son activité, fonctionne d'une manière similaire aux champs classiques. Afin de rendre le champ sélectif pour toutes les conditions de stimulation externe, il est donc nécessaire de paramétrer les équations du champ d'une telle manière que celui-ci soit capable de générer des bulles d'excitation même dans les contextes de faible stimulation externe.

Focalisons-nous ensuite sur les phénomènes dynamiques autour d'une bulle locale d'excitation, une fois qu'elle a émergé dans une zone quelconque du champ. Les unités générant cette réponse locale ont, par définition, une activité $u(x, t) \gg 0$. Toutes ces unités appartiennent aussi à un voisinage plus large d'unités pour lesquelles $\mathcal{E}(x, t) \gg 0$. Pour celles-ci, l'excitation \mathcal{E} est suffisamment forte pour entraîner la croissance de l'activité auxiliaire v . Tant que le champ continue d'être stimulé par une distribution i relativement forte, la bulle locale générée reste stable. Néanmoins, si cette bulle n'est pas issue d'une forte stimulation locale, ou entre temps, les stimuli externes ont cessé d'avoir une forte influence pour la maintenir (c'est-à-dire i est (re)devenue faible), $g(i, v)$ remplit toutes les conditions pour augmenter sa propre valeur. En effet, comme nous pouvons observer sur la figure 6.2, $v \approx 1$ et $i \approx 0$ implique $g(i, v) \approx 1$. Le facteur de pondération γ est choisi suffisamment grand pour qu'au moment où $g(i, v)$ s'approche de la valeur unitaire, le terme $-\gamma g(i, v)$ devienne prédominant dans l'équation 6.5. Finalement, ceci déclenche la forte décroissance de l'activité u pour toutes les unités pour lesquelles $v(x, t) \approx 1$,

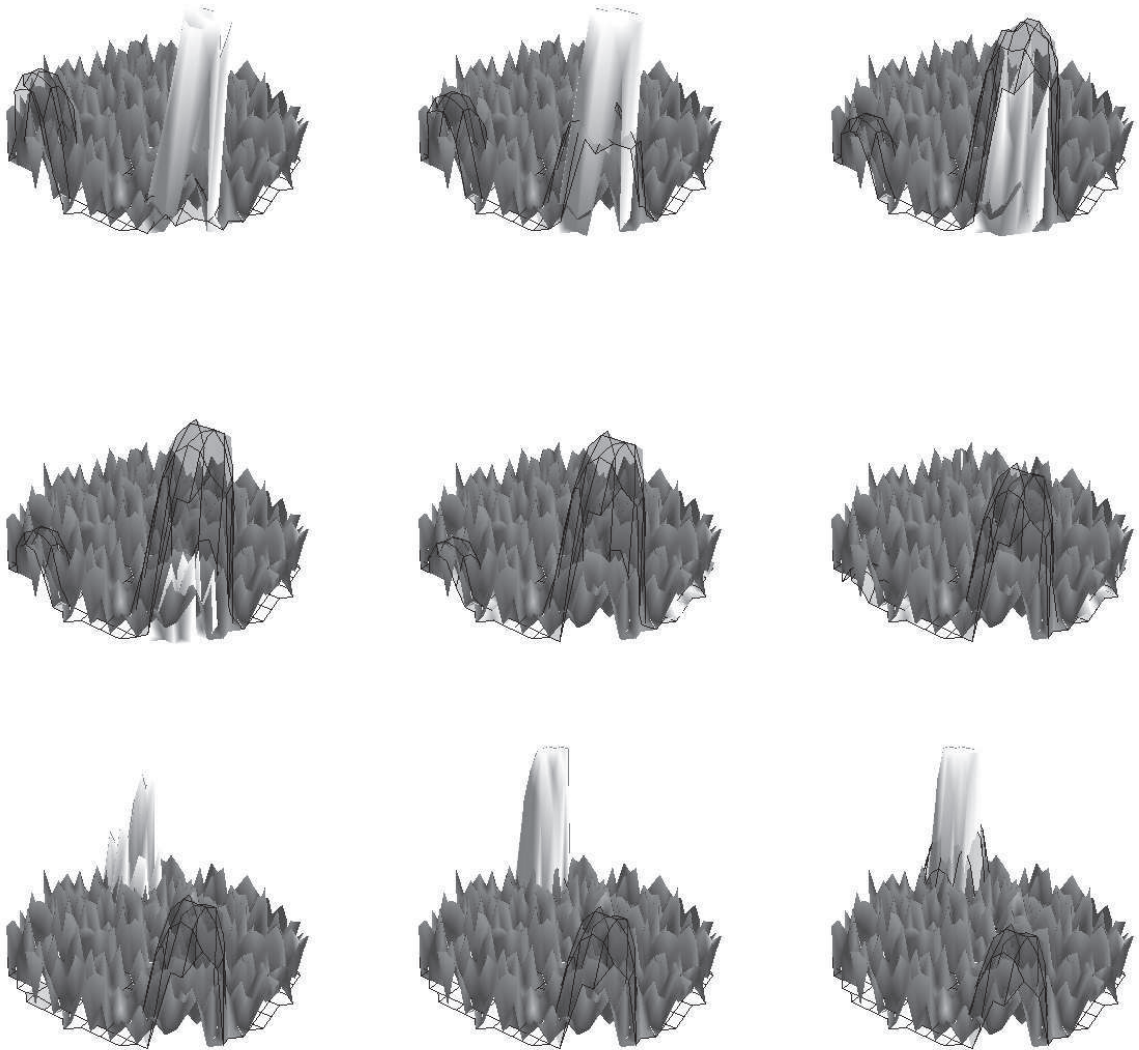


FIG. 6.3 – La formation d’une bulle d’activité neuronale u (en clair) en réponse à une faible entrée de stimulation i (en sombre) est accompagnée, après un certain délai, par une augmentation de l’activité v (en grille). Cette activité déclenche ensuite l’inhibition de l’activité u , afin de permettre l’émergence d’une nouvelle bulle d’activité neuronale u ailleurs dans le champ. Le déroulement de ces étapes est illustré par les graphiques de gauche à droite, et du haut vers le bas de cette figure.

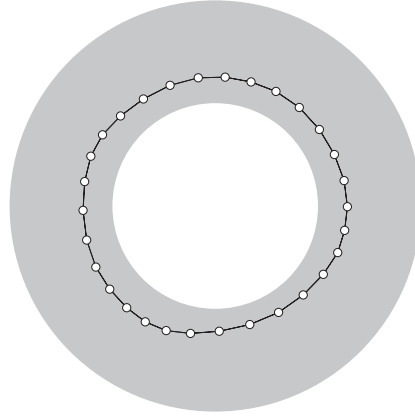


FIG. 6.4 – Expérience d'auto-organisation unimensionnelle, déroulée conformément au scénario décrit dans 5.3.4. Résultats obtenus en utilisant comme module de sélection de l'architecture d'auto-organisation d'inspiration corticale le nouveau modèle de champ, CNDRI.

ce qui se traduit par la réinhibition locale non seulement de la bulle d'excitation, mais aussi de son voisinage. En résumé, le mécanisme dynamique mis en œuvre par la contribution $g(i, v)$ a pour rôle de « détruire » une réponse locale de type bulle d'excitation qui n'est pas ou plus soutenue par une forte stimulation locale, et de « redistribuer » l'excitation locale dans le champ, afin de rendre possible la formation d'une nouvelle réponse sélective du champ, pertinente pour la zone du champ en question, qui serait cohérente avec les nouvelles conditions de stimulation externe. Étant donné qu'il s'agit d'un mécanisme dynamique à retardement, censé rétablir l'équilibre local dynamique du champ par une réinhibition locale, nous appelons les champs décrits par ces équations des *champs neuronaux dynamiques rétro-inhibés*, et référés dans la suite par l'acronyme CNDRI.

6.2 Résolution des tâches d'auto-organisation avec des champs neuronaux dynamiques

Le formalisme neuronal présenté dans le paragraphe précédent, conçu spécialement pour rendre sélectives et adaptatives les réponses d'un champ neuronal dans des conditions de forte stimulation externe, ainsi que dans celles de faible stimulation, est confronté dans cette section aux scénarios d'auto-organisation unidimensionnelle et bidimensionnelle, respectivement décrits dans les paragraphes 5.1.4 et 5.1.5.

6.2.1 Auto-organisation unidimensionnelle

La figure 6.4 présente l'état du réseau des prototypes, suite au déroulement du scénario unidimensionnel d'auto-organisation guidé par la dynamique d'un champ de type CNDRI.

Contrairement aux résultats enregistrés par les champs classiques (revoir la figure 5.17), l'utilisation d'un champ rétro-inhibé comme module de sélection de l'architecture neuronale d'auto-organisation mène aux résultats attendus. Comme nous pouvons observer, les prototypes sont organisés topologiquement, ce qui montre la réussite du processus de quantification vectorielle.

Comme nous allons le constater dans le paragraphe 6.2.3, le mécanisme dynamique de rétro-inhibition du nouveau mécanisme neuronal que nous avons proposé joue un rôle déterminant

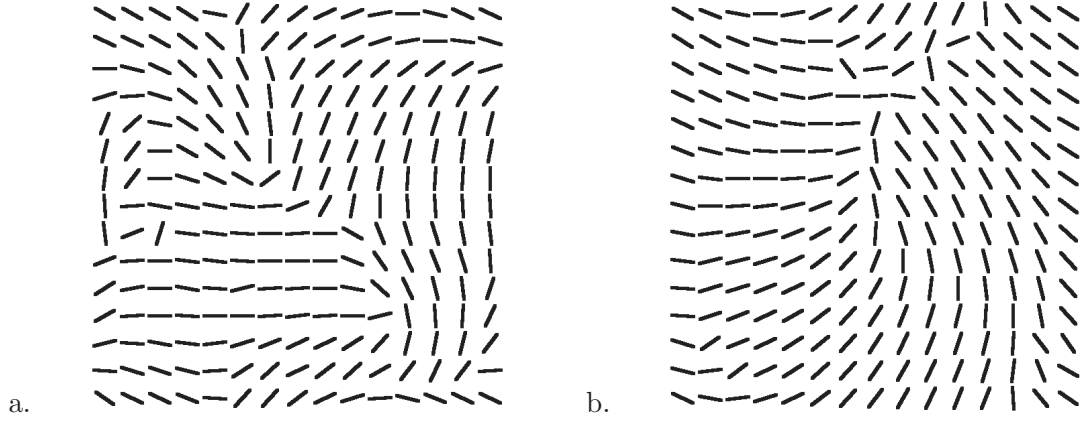


FIG. 6.5 – Expérience d'auto-organisation bidimensionnelle, déroulée conformément au scénario décrit dans 5.3.5, en considérant comme fonctions de reconnaissance des prototypes respectivement (a) la fonction f_2 définie par l'équation 5.2, et (b) la fonction f_3 définie par l'équation 5.3. Résultats obtenus utilisant comme module de sélection de l'architecture d'auto-organisation d'inspiration corticale le modèle neuronal CNDRI.

dans l'obtention de ces résultats.

6.2.2 Auto-organisation bidimensionnelle

La figure 6.5 présente les résultats expérimentaux du scénario d'auto-organisation bidimensionnelle, obtenus pour chacune des deux fonctions de reconnaissance des prototypes, f_2 et f_3 , respectivement définies par les équations 5.2 et 5.3.

Nous remarquons que, si le processus d'auto-organisation implémenté utilisant les champs classiques (CNDA et CNDPE) et la fonction de reconnaissance f_3 ne convergeait pas vers des résultats satisfaisants (revoir la figure 5.20), l'utilisation du modèle CNDRI réussit à conduire vers la réorganisation correcte de l'ensemble des prototypes, dans les deux cas de fonctions de reconnaissance considérés.

Afin d'expliquer ces résultats expérimentaux, nous présentons dans le paragraphe suivant une analyse détaillée du comportement de champ, et examinons comment celui-ci influence la dynamique globale du processus de réorganisation des prototypes.

6.2.3 Discussion

Analysons ici plus en détail les résultats obtenus dans les deux scénarios d'auto-organisation examinés ci-dessus.

La figure 6.6 montre plusieurs résultats intermédiaires de la topologie des prototypes, en parallèle avec la distribution d'entrée i (c'est-à-dire la réponse du module de reconnaissance de l'architecture d'auto-organisation à la présentation d'une même entrée ξ à toutes les unités, cf. la figure 2.9), obtenus au cours du déroulement de la simulation du processus d'auto-organisation. Une figure similaire peut être bien sûr obtenue également pour le scénario unidimensionnel d'auto-organisation, mais celle présentée pour le cas bidimensionnel illustre bien le comportement que nous nous proposons de décrire. Cette figure nous aide en effet à illustrer la correspondance entre la distribution i des réponses de reconnaissance des prototypes, et l'évolution du processus de réorganisation de ceux-ci.

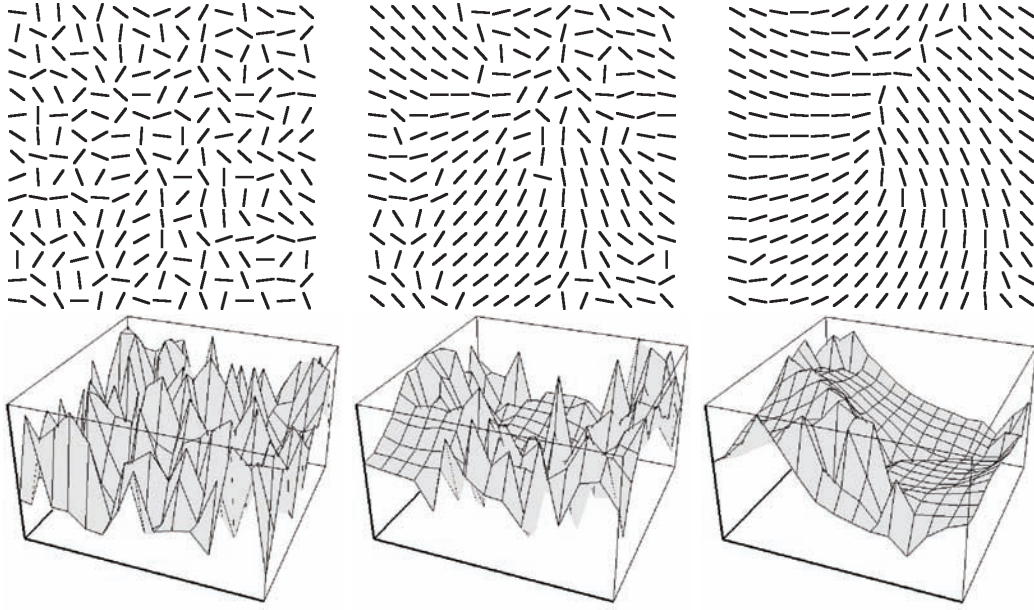


FIG. 6.6 – Résultats intermédiaires enregistrés pendant la simulation du scénario d'auto-organisation bidimensionnelle, utilisant le modèle de champ CNDRI et la fonction f_3 comme fonction de reconnaissance des prototypes. Les trois colonnes illustrées ici correspondent respectivement aux étapes suivantes : au début de l'expérience (gauche), après que 1500 échantillons ont été présentés (milieu), et après que 6500 échantillons ont été présentés. Le rang supérieur présente la topologie des prototypes, alors que celui inférieur présente la distribution i correspondante, calculée par le module de reconnaissance de l'architecture neuronale d'auto-organisation. Comme observé, le profil de la distribution i devient de plus en plus régulier avec le temps, grâce au fait que l'apprentissage des prototypes entraîne une réorganisation selon leurs sensibilités d'orientation.

Tel que nous l'avons décrit dans la section 2.6, le processus d'auto-organisation consiste à présenter à chaque pas de temps un nouvel échantillon ξ , tiré aléatoirement de l'ensemble Ξ considéré à être quantifié. Chaque prototype du module de reconnaissance calcule sa réponse vis-à-vis de ξ , formant ainsi la distribution i . Celle-ci stimule ensuite le champ neuronal du module de sélection. La présence du même échantillon ξ est maintenue durant plusieurs pas (100 dans nos simulations), afin de permettre la stabilisation de l'activité u du champ. Si pour les champs classiques (CNDA et CNDPE) examinés précédemment dans nos expériences, le profil de cette activité module l'ajustement des valeurs des prototypes, (voir la règle d'apprentissage définie par l'équation 2.18), nous avons choisi de considérer que pour le champ de type CNDRI, l'ajustement des valeurs des prototypes est effectué exclusivement par la partie excitatrice de l'activité du champ. Ceci revient à remplacer dans l'équation 2.18 le terme $u(x, t)$ par le terme $\mathcal{E}(x, t)$. Notons que ce changement est possible grâce à la séparation des contributions excitatrices et inhibitrices modélisées par les termes $\mathcal{E}(x, t)$ et $\mathcal{I}(x, t)$ de l'équation 6.5 du formalisme CNDRI. Considérer \mathcal{E} au lieu de u pour moduler l'apprentissage des prototypes a l'avantage que cette distribution a un profil plus lisse, ressemblant plus à une distribution gaussienne. De plus, la distribution \mathcal{E} couvre une région du champ plus large que u , qui s'étend autour de la bulle d'activité u . Cet aspect permet une différenciation plus graduelle de l'ajustement des valeurs des prototypes. Sur

le long terme, cet effet se traduit dans la spécialisation plus nette et plus rapide des prototypes dans le processus de quantification vectorielle.

Comme nous pouvons l'observer sur la colonne de gauche de la figure 6.6, au début du processus d'auto-organisation, la distribution i est caractérisée par une activité ayant de hautes fréquences spatiales, due au fait que tous les prototypes ont initialement des valeurs aléatoires. Les paramètres du modèle CNDRI sont fixés tels que le champ soit capable de générer une réponse locale de type bulle d'activité dans ces conditions. La formation d'une telle réponse entraîne implicitement l'augmentation de l'activité excitatrice \mathcal{E} . Étant donné que le champ est stimulé pendant plusieurs pas de simulation par la même distribution i (chaque échantillon ξ est présenté de manière répétée 100 pas), chaque prototype $\omega(x, t)$ ajuste sa valeur proportionnellement au niveau d'excitation $\mathcal{E}(x, t)$. Nous rappelons que pour toutes les unités x qui gardent un niveau élevé d'excitation \mathcal{E} pour suffisamment long temps, l'activité auxiliaire $v(x, t)$ devient aussi élevée après un certain délai (revoir l'équation 6.6).

Quand un nouvel échantillon est présenté au module de reconnaissance, la distribution i change son profil. Pour les unités du champ qui maintiennent leur niveau d'excitation $\mathcal{E}(x, t)$ élevé (et implicitement le niveau $v(x, t)$), une faible stimulation $i(x, t)$ déclenche le mécanisme de rétro-inhibition. Si suffisamment d'unités sont réinhibées par ce mécanisme, ce phénomène peut entraîner la forte inhibition de la bulle d'activité neuronale générée par le champ. Ceci permet l'émergence d'une nouvelle bulle d'excitation locale, en dehors de cette zone du champ qui vient d'être inhibée. Ainsi, grâce au mécanisme dynamique de régulation de l'inhibition locale, le champ est capable d'adapter sa réponse en fonction du profil de la stimulation externe i , gardant son caractère sélectif vis-à-vis de celle-ci. Par conséquent, la plupart des prototypes qui concordent le plus avec le nouvel échantillon (donc pour lesquels la réponse $i(x, t)$ est la plus forte) sont favorisés par la nouvelle bulle d'excitation locale à ajuster leurs valeurs, minimisant leur erreurs de quantification relatives au nouvel échantillon. En effet, comme nous pouvons constater sur la colonne du milieu de la figure 6.6, pour l'état intermédiaire illustré, le réseau des prototypes évolue vers une réorganisation topologique. En même temps, comme conséquence de celle-ci, la distribution i devient de plus en plus régulière. À la fin du scénario d'auto-organisation, (quand les prototypes ajustent seulement très peu leurs valeurs, suite à la présentation d'un nouvel échantillon), la topologie des prototypes devient très régulière, ce qui confirme le succès du processus d'auto-organisation (voir la dernière colonne de la figure 6.6). De plus, la distribution i montre également un profil très régulier, en forme de bulle locale, similaire au profil de l'activité excitatrice du champ.

6.3 Dynamiques émergentes de CNDRI

Si dans la section précédente nous avons pu constater le succès du modèle CNDRI dans la modélisation du processus d'auto-organisation, nous sommes également intéressé par l'étude du comportement de ce modèle dans d'autres contextes pratiques de modélisation.

Afin de mieux comparer les capacités dynamiques du nouveau modèle avec celles exhibées par les modèles classiques de champ, nous examinons dans ce paragraphe le comportement du modèle CNDRI dans les scénarios considérés dans la section 5.1, à l'aide de la plate-forme d'évaluation empirique des champs neuronaux, présentée dans la section 5.2.5.

6.3.1 Scénarios d'attention

La figure 6.7 présente les résultats expérimentaux des évaluations du modèle CNDRI dans des scénarios d'attention. Comparant ces résultats avec ceux obtenus par les champs neuronaux

classiques (revoir la figure 5.14), nous pouvons constater une grande ressemblance entre les comportements dynamiques des trois modèles de champ (CNDA, CNDPE et CNDRI).

Analysant ces résultats, nous soulignons le cas particulier du scénario (e) présenté dans la figure 6.7. Comme nous pouvons l'observer, dans ce contexte de stimulation, le champ génère une bulle d'activité neuronale au début du scénario, mais il est incapable d'adapter sa réponse vis-à-vis d'un nouveau stimulus, d'une intensité plus forte que celui déjà perçu. Ce phénomène résulte du fait que la génération d'une bulle d'activité induit l'augmentation de l'activité u et de l'excitation \mathcal{E} dans la zone où i est localement le plus fort, mais entraîne simultanément une augmentation de l'inhibition \mathcal{I} dans les autres endroits du champ. Par conséquent, tant qu'il est soutenu par une stimulation locale i suffisamment forte, le champ maintient au même endroit sa bulle d'activité, ainsi que son niveau d'inhibition ailleurs dans le champ. Ce comportement, qui contrevient aux propriétés $(P \wedge Q)$ (revoir pour illustration les figures 5.6 (a,g)), induit la courbe de qualité et l'histogramme de qualité, illustrées par les graphiques de la figure 6.7 (e). Cet exemple révèle ainsi le fait que le modèle CNDRI, comme les autres modèles de champ examinés (voir la figure 5.14 (e)), ne réussit pas non plus à respecter entièrement les propriétés dynamiques $(P \wedge Q)$, énoncées dans la section 5.2.1.

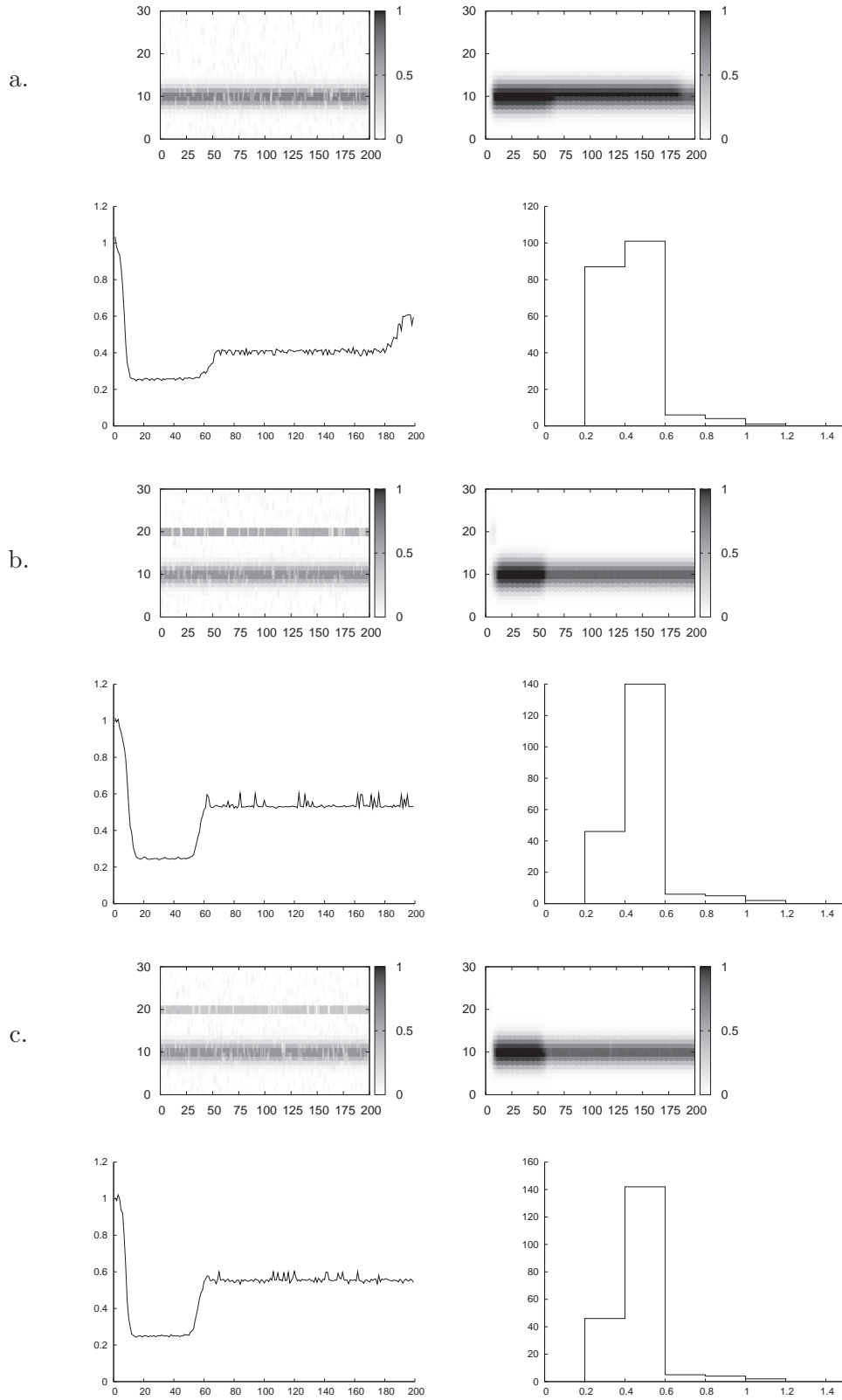
En outre, nous pouvons affirmer que ce comportement dynamique est intrinsèque aux formalismes de type champ neuronal, qui reposent (pour l'existence des bulles d'activité) sur le maintien d'un équilibre dynamique entre les contributions excitatrices et inhibitrices. Une méthode qui pourrait améliorer cet effet dynamique pourrait prendre en compte le déclenchement périodique du mécanisme de rétro-inhibition, indépendamment du niveau de stimulation i . Une autre méthode pour résoudre ce problème pourrait utiliser un mécanisme d'ajustement graduel du niveau de repos du champ, comme proposé dans les approches de modélisation de Erlhagen, Coombes et Owen, présentées dans la section 3.2.6. Nous prévoyons d'examiner plus en détail ces pistes de recherche dans nos travaux futurs, afin de rendre le modèle CNDRI plus versatile à l'égard des propriétés $(P \wedge Q)$, et des besoins pratiques d'utilisation.

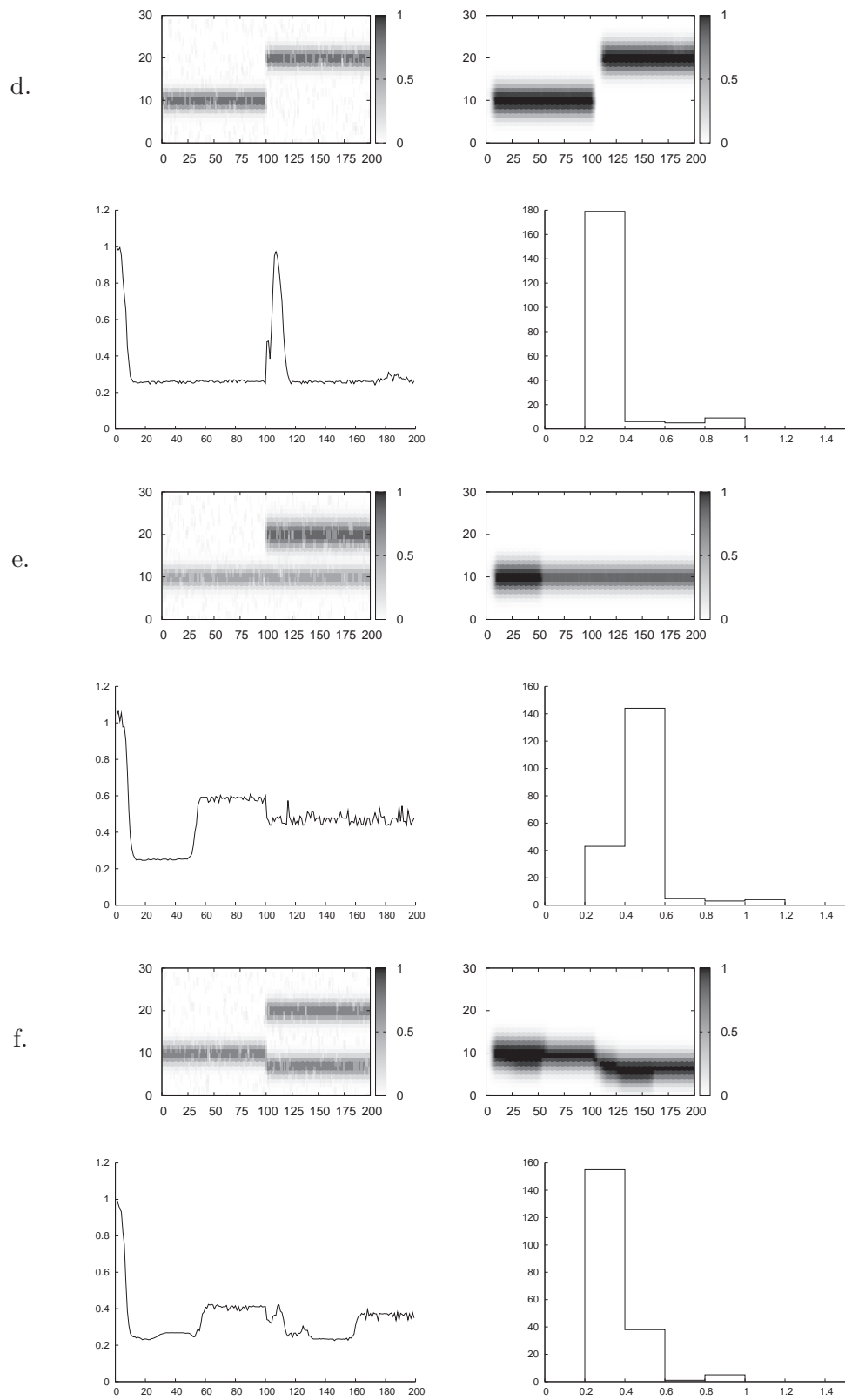
Cependant, comme nous allons le constater dans les paragraphes suivants, ce comportement dynamique n'affecte pas les performances du modèle CNDRI dans les scénarios d'auto-organisation.

6.3.2 Scénarios d'exploration

La figure 6.8 présente les résultats d'évaluation de la dynamique du modèle CNDRI dans les scénarios d'exploration. Dans les deux scénarios d'exploration examinés, le modèle montre un caractère sélectif. Toutefois, contrairement aux modèles classiques de champs que nous avons examinés (CNDA et CNDPE), qui n'exhibent aucune augmentation de leur activité dans des conditions de faible stimulation (revoir la figure 5.15 (b)), le modèle CNDRI réussit à générer une réponse locale de type bulle d'excitation dans ces conditions (voir la figure 6.8 (b)).

Si nous analysons en détail l'évolution du champ, nous pouvons observer qu'un certain temps après la formation de la bulle d'activité neuronale, la bulle réduit son niveau d'activation et une nouvelle bulle apparaît dans un autre endroit du champ. Ce changement dynamique est dû à l'augmentation de l'activité auxiliaire v (nous rappelons que v indique, avec un certain retard, un fort niveau d'excitation des unités). Ajoutant à cet aspect le fait que la bulle a été générée dans un contexte de faible stimulation externe, la forte réinhibition locale du champ est déclenchée par la contribution g de l'équation du champ (nous rappelons également que $v \approx 1$ et $i \approx 0$ implique $g(i, v) \approx 1$). Par conséquent, l'inhibition de la région du champ couvrant la bulle d'activité et son voisinage favorise l'émergence d'une nouvelle bulle d'activité ailleurs dans le champ. Grâce à ce mécanisme neuronal de réinhibition locale, le champ maintient son





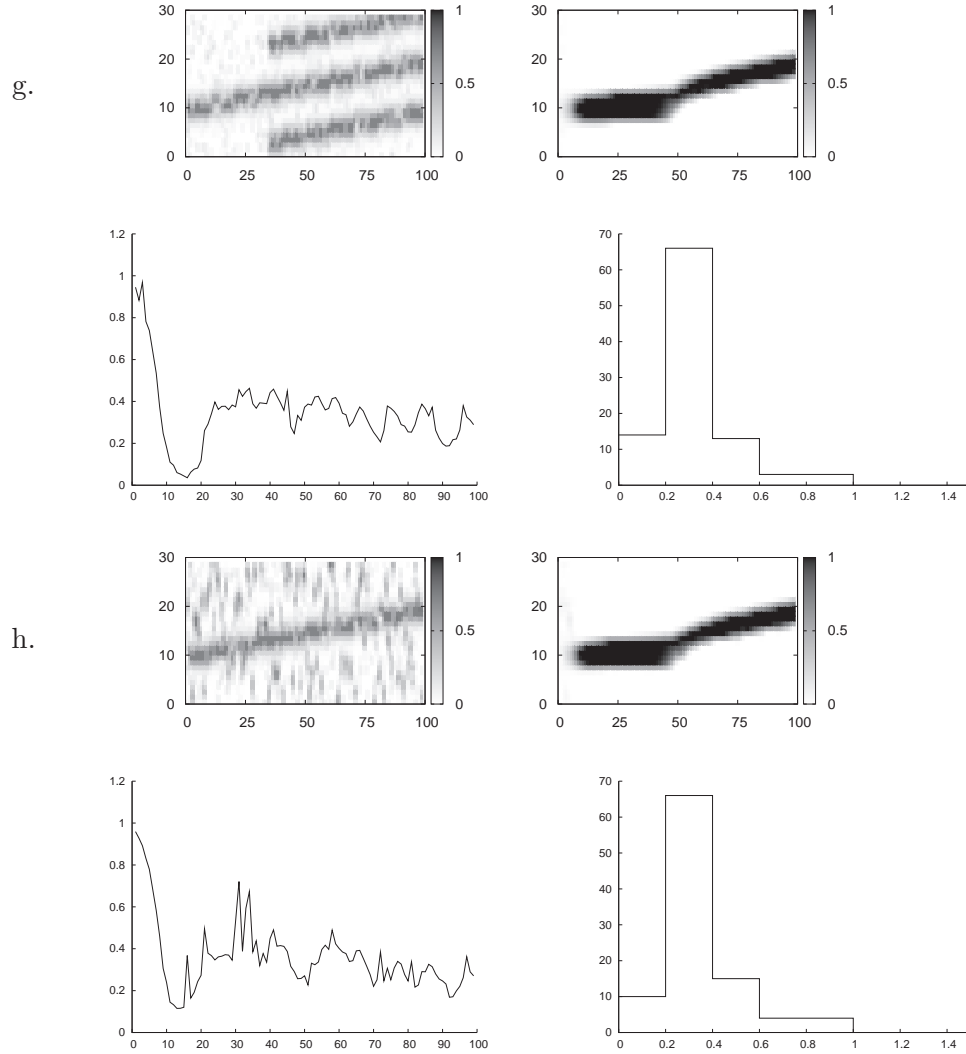


FIG. 6.7 – Résultats expérimentaux de l'évaluation du modèle CNDRI, obtenus dans les scénarios d'attention considérés dans la figure 5.1. Chaque expérience de simulation est illustrée par les quatre graphiques suivants : la distribution de stimulation i , l'activité neuronale u du champ, la courbe de qualité de la réponse du champ relative au scénario examiné, et l'histogramme de qualité calculée à partir de cette courbe de qualité.

caractère sélectif même dans les conditions de faible stimulation. Ce comportement aide également le champ à « explorer » le profil de stimulation externe, pour trouver les régions du champ les plus stimulées par la distribution i . Ainsi, une nouvelle bulle d'activité n'est jamais générée là où la bulle précédente avait émergée.

6.3.3 Discussion

Synthétisant les résultats d'évaluation dynamique présentés dans les paragraphes précédents, nous pouvons affirmer que, du point de vue qualitatif, le nouveau formalisme CNDRI peut être considéré comme étant une extension des modèles classiques des champs neuronaux. Dans les scénarios de forte stimulation, le nouveau modèle se comporte de manière similaire aux modèles classiques que nous avons examinés. La différence notable du modèle CNDRI est que ce nouveau modèle de champ réussit à maintenir son caractère sélectif dans les contextes de faible stimulation, adaptant sa réponse en permanence, vis-à-vis de la distribution de stimulation externe i . Comme nous avons pu constater dans la section 6.2, cet aspect du comportement dynamique du champ permet l'implantation avec succès du module de sélection de l'architecture neuronale qui modélise le processus cortical d'auto-organisation.

6.4 Les paramètres du modèle CNDRI

6.4.1 Procédure de paramétrage des équations du modèle CNDRI

L'analyse empirique des modèles classiques des champs neuronaux, présentée dans le chapitre précédent, avait mis en évidence la difficulté de paramétrer ces modèles pour les rendre capables d'exhiber les propriétés dynamiques ($P \wedge Q$).

Face à ces difficultés, nous avons proposé dans ce chapitre un nouveau modèle de champ, conçu pour satisfaire ($P \wedge Q$) dans le contexte de modélisation des processus neuronaux d'auto-organisation. Outre cet aspect, le modèle CNDRI a également l'avantage de permettre un paramétrage plus intuitif. Pour étayer cette affirmation, nous décrivons dans la suite une procédure pour établir les valeurs numériques des paramètres de ce modèle. Les étapes de cette procédure sont illustrées pour le cas d'un champ unidimensionnel, ayant vocation à ne produire qu'une seule bulle d'activité.

Ajuster les niveaux des contributions latérales excitatrices et inhibitrices

Considérons $b(r)$ la fonction qui décrit la forme souhaitée d'une bulle d'activité neuronale, où l'argument r désigne la distance entre une position quelconque $x \in X$ du champ et la position de cette bulle. Soient $exci(x)$ et $inhi(x)$ respectivement les contributions excitatrices et inhibitrices, calculées pour l'unité x :

$$exci(x) = \int_{-\infty}^{\infty} w^+(|x - x'|) b(|x - x'|) dx' \quad (6.7)$$

$$inhi(x) = \int_{-\infty}^{\infty} w^-(|x - x'|) b(|x - x'|) dx' \quad (6.8)$$

La forme des deux contributions est représentée graphiquement dans la figure 6.9. Soient E_{max} et I_{max} les valeurs maximales atteintes par chacune des deux contributions, $exci(x)$ et

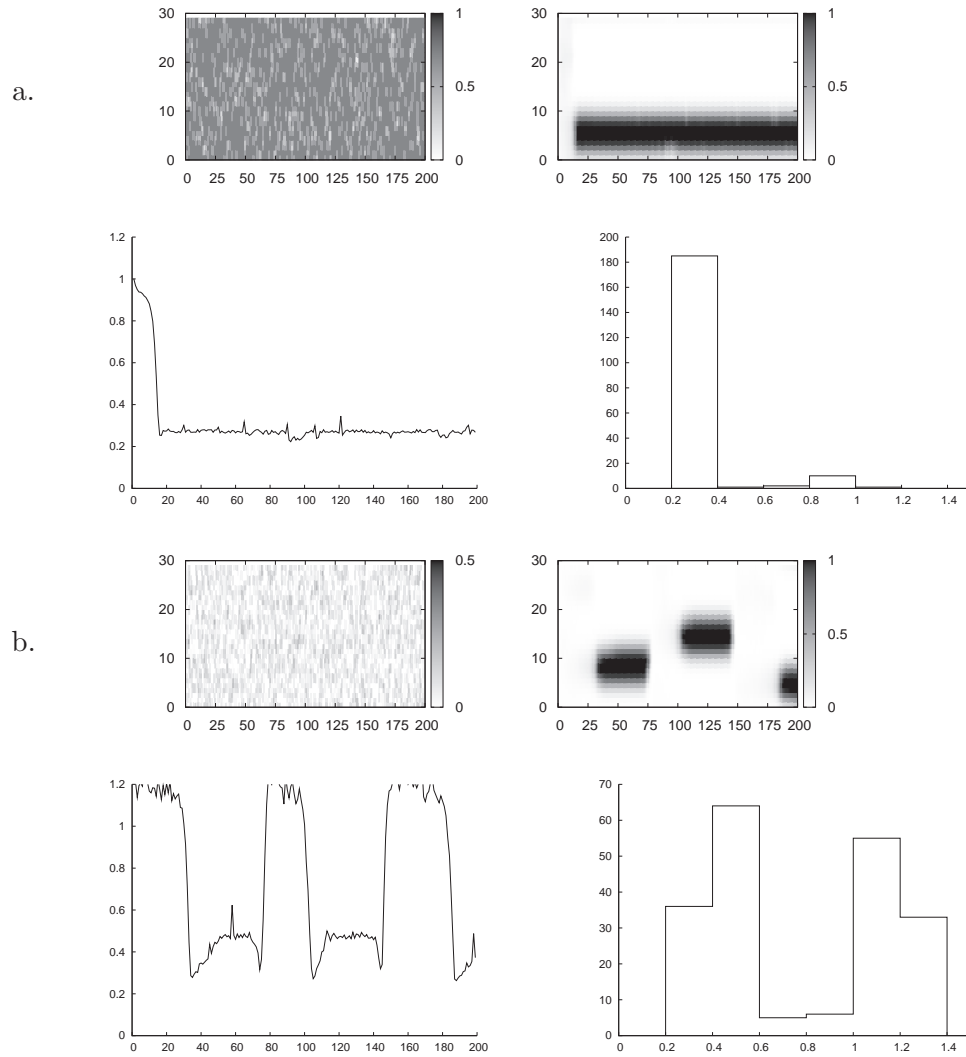


FIG. 6.8 – Résultats expérimentaux de l'évaluation du modèle CNDRI dans les scénarios d'exploration considérés dans la figure 5.2. La signification des graphiques pour chaque scénario est la même que pour les scénarios d'attention, présentés dans la figure 6.7.

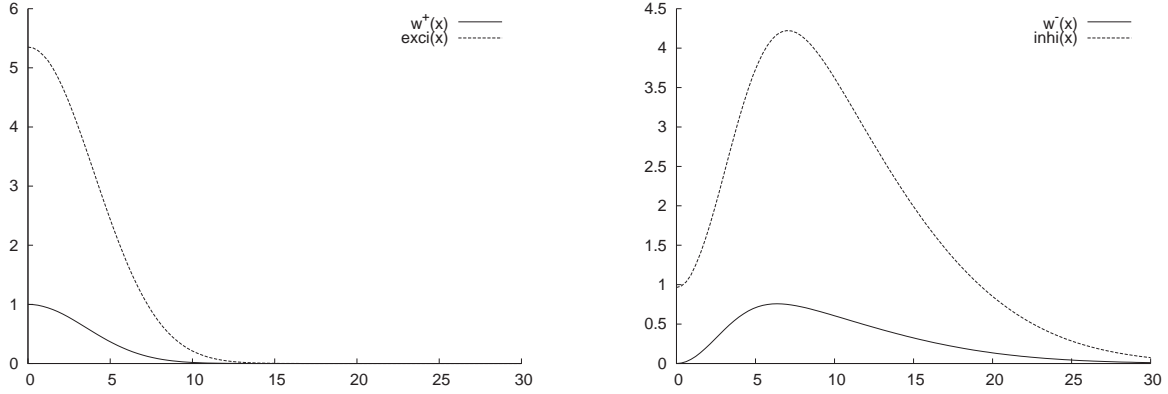


FIG. 6.9 – Le graphique des fonctions $exci(x)$ (gauche) et $inhi(x)$ (droite), pour une fonction $b(r) = b_2(r, 0)$ désignant une bulle d'activité de type créneau, avec un rayon $R = 3$ (voir l'équation 5.15). Les fonctions $w^+(x)$ et $w^-(x)$ sont également illustrées graphiquement ($a = 0.04$, $b = 0.005$ et $c = 0.07$).

$inhi(x)$. Soient ensuite les fonctions sigmoïdes σ^+ et σ^- , définies telles que :

$$\sigma^+(0) = \sigma^-(0) = 0 \quad (6.9)$$

$$\sigma^+(E_{max}) = \sigma^-(I_{max}) = 1 \quad (6.10)$$

Ces conditions impliquent le fait que le niveau maximal des contributions excitatrices ou inhibitrices voisines, ressenties par une unité quelconque du champ, soit atteint pleinement quand une seule bulle d'activité neuronale est présente dans le champ.

Notons que pour faciliter l'estimation des valeurs, le calcul des fonctions $exci(x)$ et $inhi(x)$, et les valeurs E_{max} et I_{max} , est illustré dans la figure 6.9 en considérant une forme de bulle de type $b(r) = b_2(r, 0)$. Toutefois, dans les simulations, la forme d'une bulle du champ est un créneau un peu lissé. Pour cette raison, en pratique, les valeurs E_{max} et I_{max} sont légèrement plus petites que celles estimées par ce calcul théorique.

Nous rappelons que, en ce qui concerne le comportement dynamique du modèle CNDRI, il est essentiel que le champ soit capable d'exhiber une réponse locale de type bulle d'activité dans des conditions de faible stimulation externe. Dans ce contexte ($i(x, t) \approx 0$) l'évolution de l'activité neuronale $u(x)$ d'une unité du champ dépend principalement des contributions excitatrices et inhibitrices des unités voisines, \mathcal{E} et \mathcal{I} . Pour que le champ puisse produire une bulle d'activité u , les paramètres α et β doivent être choisis tels que $\alpha exci(R) - \beta inhi(R) = 0$, où R est le rayon souhaité de la bulle. Cette condition est illustrée dans la figure 6.10.

La méthode présentée ici décrit en lignes générales une procédure pour fixer les valeurs des paramètres a, b, c, α et β , ainsi que les seuils des fonctions sigmoïdes σ^+ et σ^- , afin de rendre le champ capable de générer une bulle d'activité dans des conditions de faible stimulation externe.

Toutefois, cette méthode peut facilement s'étendre pour le cas des champs bidimensionnels, ou encore pour des champs pour lesquels nous souhaitons qu'ils exhibent simultanément plusieurs bulles d'activité neuronale. Pour ce dernier cas, considérons R_{inf} , le rayon de l'influence des contributions excitatrice et inhibitrices des connexions latérales, défini tel que :

$$R_{inf} = \inf_{R'} \{ |\alpha exci(R') - \beta inhi(R')| < \epsilon, \text{ tel que } |\alpha exci(x) - \beta inhi(x)| < \epsilon, \forall x > R' \}$$

pour un ϵ fixé, suffisamment petit.

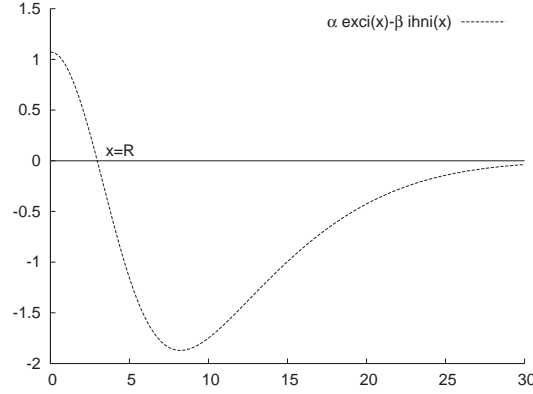


FIG. 6.10 – Illustration graphique de la fonction $\alpha \text{ exci}(x) - \beta \text{ inhi}(x)$, avec $\alpha = 1.6$ et $\beta = 2.0$.

Afin de respecter la contrainte de densité des bulles d'activité dans le champ, il suffit de réajuster les valeurs des paramètres ci-dessus mentionnés, d'une telle manière que R_{inf} respecte l'inégalité imposée par la propriété (Q) :

$$b_{min} < 2R_{inf} < b_{max}$$

Au final, nous rappelons que le taux de croissance de l'activité u est imposé par la constante de temps τ . De manière qualitative, ce paramètre influence seulement le temps nécessaire pour la saturation de l'activité du champ. La valeur τ utilisée dans toutes nos simulations est $\tau = 0.3^{-1}$.

Ajuster la dynamique de l'activité auxiliaire v

Nous avons déjà affirmé précédemment que l'activité $v(x, t)$ est une variante retardée de l'excitation $\mathcal{E}(x, t)$, sa dynamique étant modulée par la fonction sigmoïde h . La partie inférieure de cette fonction est négative, et définit le taux de décroissance de l'activité v vers son niveau de repos, alors que sa partie supérieure est positive, et détermine la vitesse de croissance de v par rapport à l'excitation. Les valeurs positives de h influencent, de manière proportionnelle, la latence avec laquelle le mécanisme de rétro-inhibition devient actif, pendant que ses valeurs négatives influencent aussi de manière proportionnelle le temps nécessaire à v pour revenir à son niveau de repos.

Dans nos simulations, h est une fonction monotone par morceaux, telle que $h(0) = -0.01$, $h(0.3) = 0$, et $h(1) = 0.1$, ou plus précisément,

$$h(v) = \begin{cases} -0.033v & \text{si } 0 \leq v < 0.3 \\ -0.066v - 0.02 & \text{si } 0.3 \leq v < 0.6 \\ -0.025v + 0.035 & \text{si } 0.6 \leq v \leq 1 \end{cases} \quad (6.11)$$

Ajuster la dynamique du mécanisme de rétro-inhibition

Remarquons tout d'abord qu'en absence du terme $-\gamma g(i, v)$, la dynamique de croissance de l'activité neuronale d'un champ décrit par les équations 6.5 et 6.6 est similaire à celle du modèle d'Amari (revoir l'équation 3.15).

Le but du mécanisme de rétro-inhibition, mis en place par la fonction g , est d'empêcher la permanence d'une bulle d'activité neuronale u issue d'une faible stimulation externe i . Ainsi,

la fonction g est conçue telle qu'elle puisse déclencher une réinhibition locale des unités de la bulle au bout d'un certain intervalle de temps après son apparition. Par conséquent, pour chaque unité du champ, g implémente la logique suivante : pour un niveau de stimulation locale i faible, correspondant à un niveau élevé de l'excitation retardée v , g doit être fort, alors que pour toute stimulation locale i forte, g doit rester faible. Ainsi, la fonction g est définie comme étant le produit de deux fonctions sigmoïdes, chacune ayant les valeurs comprises entre 0 et 1 : la première est une fonction sigmoïde décroissante de la stimulation i , et la deuxième fonction est une sigmoïde croissante dépendant de l'activité v . Dans nos simulations, la fonction $g(i, v) = g_1(i) g_2(v)$, où :

$$g_1(i) = \begin{cases} -0.33i + 1 & \text{si } 0 \leq i < 0.3 \\ -5.33i + 2.5 & \text{si } 0.3 \leq i < 0.45 \\ -0.18i + 0.18 & \text{si } 0.45 \leq i \leq 1 \end{cases} \quad (6.12)$$

$$g_2(v) = \begin{cases} 0.5v & \text{si } 0 \leq v < 0.4 \\ 1.75v - 0.5 & \text{si } 0.4 \leq v < 0.8 \\ 0.5v + 0.5 & \text{si } 0.9 \leq v \leq 1 \end{cases} \quad (6.13)$$

En dernier point, notons que pour devenir dominant dans la dynamique de l'activité neuronale u (c'est-à-dire pour pouvoir ré-inhiber cette activité), le terme $-\gamma g(i, v)$ doit être supérieur, en valeur absolue, à la partie positive de l'équation 6.5, $\alpha \mathcal{E}(x, t) + i(x, t)$. Pour ceci, le facteur γ doit être suffisamment grand. Dans nos simulations, $\gamma = 1.3$.

6.4.2 Influence des paramètres des équations du modèle CNDRI

Afin de mieux comprendre le comportement dynamique exhibé par le nouveau modèle de champ, nous examinons dans la suite l'influence de différents paramètres du modèle CNDRI dans la dynamique globale de l'activité neuronale du champ. Cette analyse est illustrée en étudiant le comportement d'un champ neuronal unidimensionnel dans plusieurs scénarios de stimulation. Les deux paragraphes suivants présentent en détail le cas d'un champ neuronal mono-bulle évoluant dans deux scénarios : un de compétition et un d'exploration.

Cas d'un champ unidimensionnel mono-bulle dans un scénario de compétition

Prenons d'abord l'exemple d'un champ paramétré comme dans la section 6.4.1. Comme illustré dans cette section, les valeurs de ses paramètres ont été choisies telles que le champ soit capable de générer une seule bulle d'activité à la fois. Dans la suite, la dynamique de ce champ est examinée en détail dans le scénario de compétition illustré dans la figure 6.11 (a). La figure 6.11 présente ensuite un ensemble de graphiques qui montrent plus en détail la dynamique de différents paramètres du champ.

Le graphique de la figure 6.11 (b) montre l'évolution de l'activité neuronale u du champ. Comme nous pouvons constater en comparant cette activité avec la stimulation externe i , illustrée dans le graphique (a) de la même figure, le champ génère une seule bulle d'activité, qui correspond toujours à une seule des deux zones de forte stimulation de i . Afin de mieux déceler ce comportement sélectif du champ, les graphiques (c-f) de la figure 6.11 présentent l'évolution de différents paramètres qui contribuent à la dynamique de l'activité neuronale du champ, u . La figure 6.12 propose également une vue détaillée sur la dynamique individuelle de trois unités du champ.

Au début de l'expérience, l'excitation externe du champ (induite par les deux régions de forte stimulation existantes dans la distribution i) entraîne la croissance de l'activité neuronale

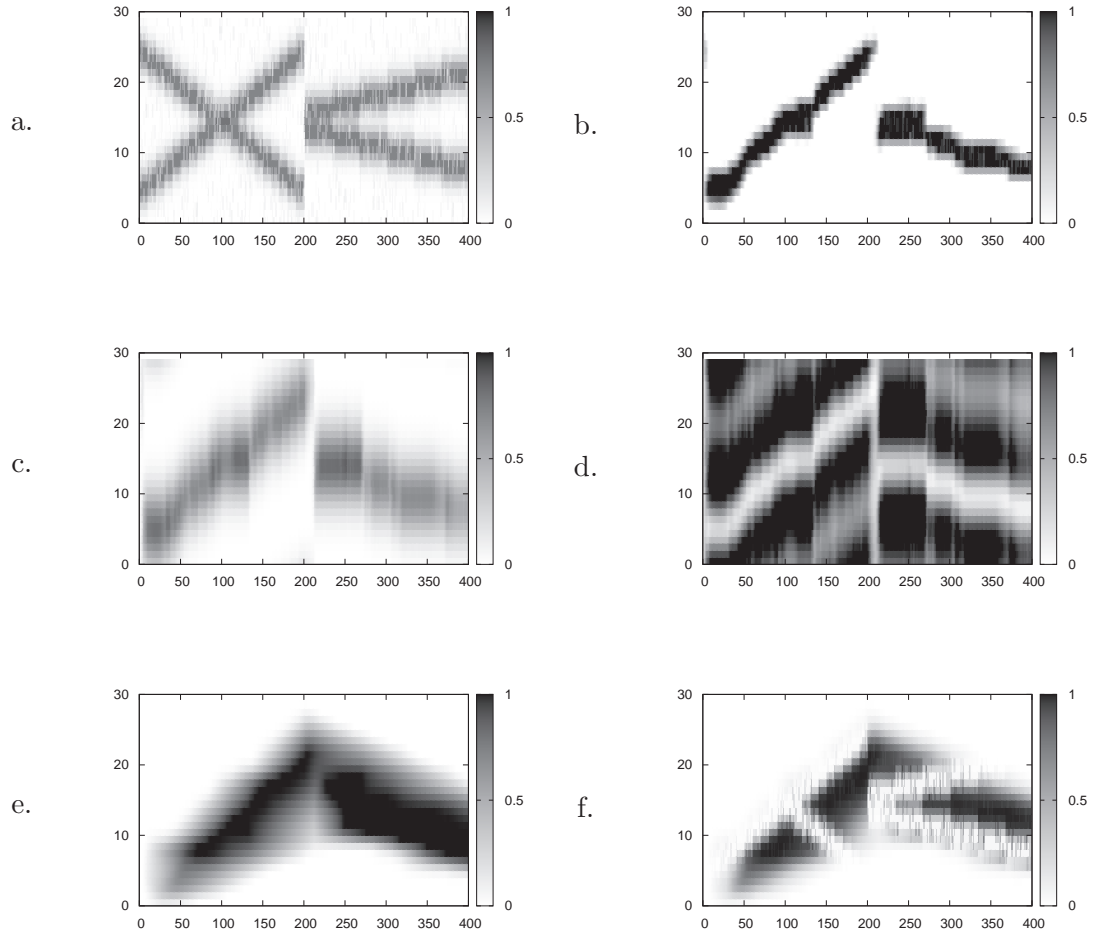


FIG. 6.11 – Illustration graphique de la dynamique de plusieurs paramètres du modèle CNDRI, pendant la simulation d'un scénario de compétition. Toutes les diagrammes sont tracées en niveaux de gris, la couleur noir symbolisant l'intensité la plus forte du paramètre considéré. (a) le scénario de stimulation externe i ; (b) l'activité neuronale u du champ; (c) la contribution latérale excitatrice \mathcal{E} ; (d) la contribution latérale inhibitrice \mathcal{I} ; (e) l'activité auxiliaire v ; (f) l'intensité g de la rétro-inhibition locale.

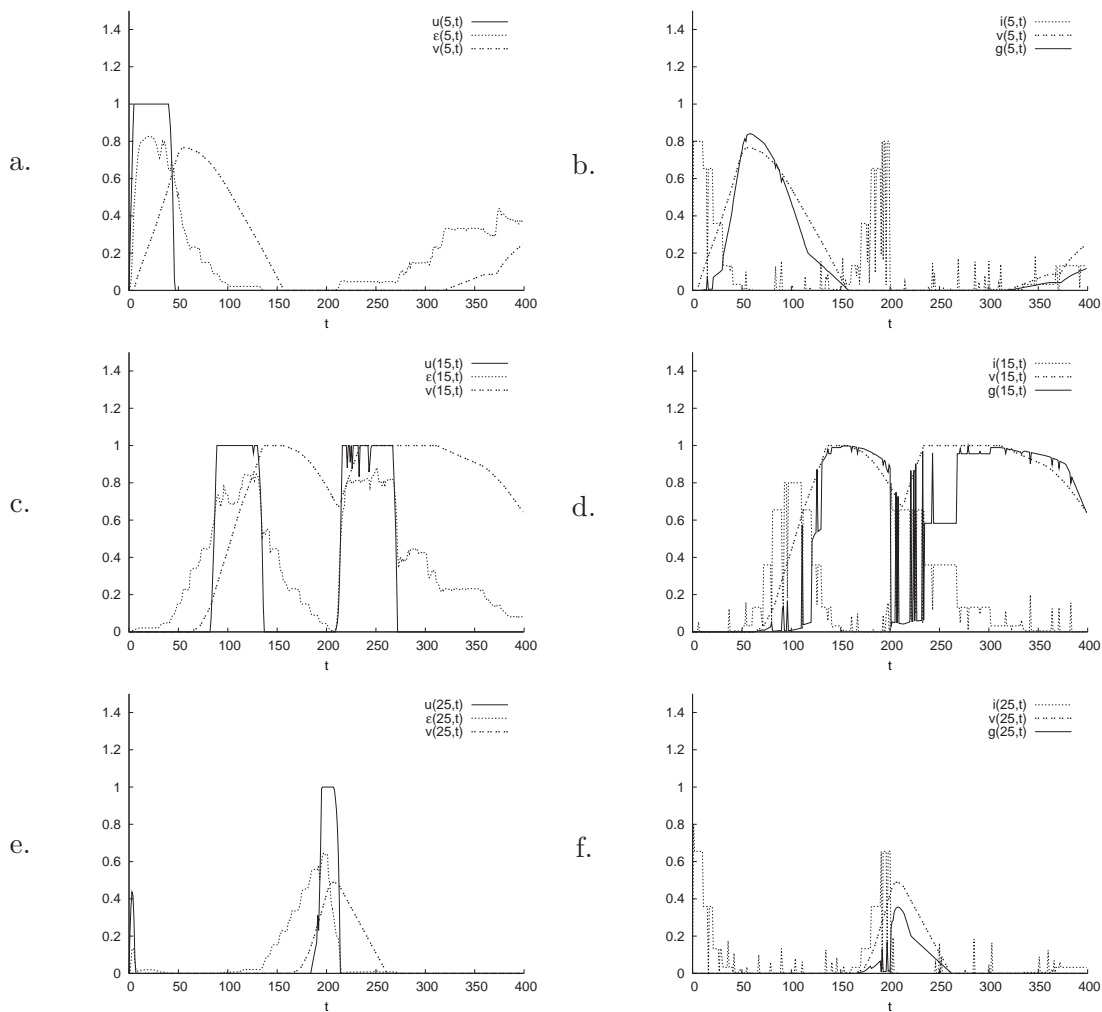


FIG. 6.12 – Détail sur le comportement du modèle CNDRI illustré graphiquement dans la figure 6.11. Analyse comparative de certains paramètres du champ, pour trois unités distinctes du champ, $x_1 = 5$ (première ligne), $x_2 = 15$ (deuxième ligne) et $x_3 = 25$ (troisième ligne). (a,c,e) dynamique des grandeurs u , ε et v ; (b,d,f) dynamique des grandeurs i , v et g . Voir le texte pour plus d'explications.

des unités correspondantes. Dans cette phase (très brève) de l'évolution de u , i est le seul terme dominant dictant sa dynamique de croissance (revoir l'équation 6.5). Cette croissance entraîne, à son tour, l'augmentation des contributions latérales excitatrices \mathcal{E} (voir la figure 6.11 (c)) et de l'inhibition \mathcal{I} (voir la figure 6.11 (d)) de ces unités et des unités voisines. Ainsi, les unités du champ entrent en compétition, via leurs connexions latérales (modélisées respectivement par les noyaux w^+ et w^-), chacune d'elle calculant son activité tenant compte de l'activité des unités voisines. Les premières unités à s'approcher d'un seuil d'excitation suffisamment élevé gagnent la compétition, le reste des unités s'inhibant (voir pour comparaison, dans la figure 6.12 (a,c,e), comment l'augmentation de l'excitation \mathcal{E} pour une unité du champ est corrélée avec la baisse de ce niveau pour d'autres unités du champ). Nous rappelons que, dans le contexte de paramétrage considéré, ce seuil de l'excitation correspond à l'excitation induite par une seule bulle d'activité dans le champ. Par conséquent, le processus de compétition permet la formation d'une unique bulle d'activité u . Ceci explique le fait que, si au tout début de l'expérience, deux zones du champ sont favorisées par la stimulation externe pour la génération d'une réponse de forte intensité, seulement l'une d'elles conduit à la formation d'une bulle d'activité.

La latence de génération d'une bulle d'activité dépend de plusieurs paramètres : τ (le taux de croissance de l'activité u), les valeurs numériques de α et β , ainsi que les pentes et les seuils des fonctions sigmoïdes σ^+ et σ^- (qui modulent les contributions latérales excitatrices et inhibitrices). Une fois la bulle d'activité formée, elle se stabilise (par saturation) et persiste dans le champ.

Afin de comprendre comment la bulle, une fois générée par le champ, suit un motif de stimulation, il est nécessaire de regarder plus en détail la dynamique d'autres paramètres influençant l'activité neuronale u . Ainsi, nous pouvons constater que les unités voisines d'une bulle ont également une activité excitatrice \mathcal{E} assez intense (voir la figure 6.11 (c)), même si elles ont une activité neuronale u nulle. Ceci fait qu'elles sont plus facilement excitées quand la stimulation externe i s'intensifie dans ces régions. Cet aspect représente le premier facteur qui favorise donc le déplacement léger de la bulle d'activité, dès que le motif de stimulation externe change dans son voisinage. Le deuxième aspect déterminant pour ce comportement adaptatif de la bulle est lié au mécanisme de rétro-inhibition : chaque unité ayant un niveau d'excitation élevé depuis suffisamment longtemps (fait indiqué par la variable « d'accumulation » v), mais n'étant plus suffisamment stimulée par i , est réinhibée par une contribution négative croissante, induite par le terme $-\gamma g$ de l'équation 6.5 du champ. Pour clarifier encore plus les cause de cette dynamique, détaillons ensuite le fonctionnement de ce mécanisme de rétro-inhibition.

Comme le montre la figure 6.11 (e), mais encore plus clairement les diagrammes (a,c,e) de la figure 6.12, l'activité auxiliaire v peut être considérée comme étant une variante (non-linéaire) retardée de l'excitation \mathcal{E} ressentie par le champ. Conformément à l'équation 6.6 du modèle CNDRI, un niveau bas de l'excitation \mathcal{E} maintient la fonction sigmoïde h négative, ce qui garde l'activité auxiliaire v à son niveau initial, nul. v commence à augmenter seulement après que \mathcal{E} dépasse un certain seuil. Si \mathcal{E} garde cette propriété pour suffisamment de temps, v continue à augmenter jusqu'à la saturation (comme par exemple dans la figure 6.11 (c)). Ainsi, comme le montrent les graphiques (a,c,d) de la même figure, $v(x)$ devient un indicateur de l'existence d'une forte et persistante excitation de l'unité x , pour toute unité x du champ.

Nous avons donc observé qu'une des deux conditions nécessaires à l'activation du mécanisme de rétro-inhibition est indiquée par un niveau élevé de l'activité auxiliaire v . La deuxième condition nécessaire est définie par un faible niveau de la stimulation externe i de l'unité considérée. Ainsi, en étudiant l'évolution des grandeurs i , v et g de manière globale respectivement dans les figures 6.11 (a), (e) et (f), ou de manière individuelle dans les figures 6.12 (b), (d) et (f), nous pouvons voir comment g impose son influence sur l'activité u du champ, déclenchant la

rétro-inhibition locale du champ dans les conditions de forte excitation et faible stimulation. Le cas le plus notable de ce comportement peut être observé après le pas 200 du scénario, marqué par un changement brusque du motif de stimulation du champ. Dans ces conditions, la fonction g de l'unité $x_3 = 25$ du champ remplit les deux conditions pour déclencher la rétro-inhibition locale du champ. Ce phénomène conduit à la redistribution de l'excitation dans le champ, et permet la génération d'une nouvelle réponse du champ, en concordance avec le nouveau motif de stimulation. Cet effet est visible sur les figures 6.11 (c), (d), qui mettent en évidence le fait que pour un court intervalle de temps après le pas 200, les niveaux des contributions excitatrices et inhibitrices ressenties par chaque unité du champ décroissent simultanément vers 0. Ce phénomène montre qu'un changement local dans l'activité de certaines unités du champ peut influencer la dynamique globale de ce champ.

Cas d'un champ unidimensionnel mono-bulle dans un scénario d'exploration

Dans un deuxième cas d'analyse, nous examinons plus en détail la dynamique du même champ neuronal dans le scénario d'exploration illustré dans la figure 6.13 (a), scénario clé pour notre discours d'analyse qui avait été décrit initialement dans la figure 5.2 (b). Nous mentionnons le fait que les valeurs numériques des paramètres du champ restent entièrement les mêmes que dans le cas d'analyse précédent.

Afin de comprendre le comportement sous-jacent du champ dans ce nouveau contexte de stimulation, les graphiques des figures 6.13 et figures 6.14 présentent respectivement de manière globale et individuelle l'évolution de plusieurs paramètres du champ.

Si dans le cas du premier scénario examiné au paragraphe précédent (scénario de compétition), l'existence d'au moins une zone de forte stimulation externe était le facteur déterminant pour l'émergence d'une réponse locale du champ de type bulle d'excitation, dans ce deuxième scénario, la faible stimulation du champ ne joue plus un rôle déterminant dans la dynamique de son activité neuronale. Cette fois-ci, la distribution i est quasi uniforme et a une basse amplitude pour toutes les unités du champ, ce qui entraîne une compétition directe entre les contributions latérales excitatrices et inhibitrices ressenties par chaque unité du champ. La domination des premières (voir par exemple dans la figure 6.14 (d) l'évolution comparative des deux contributions, le long du scénario, pour le unité $x = 7$ du champ) conduit à l'apparition d'une bulle d'activité neuronale dans la distribution u (voir le graphique (b) de la figure 6.13, pour l'activité globale du champ, et le graphique (a) de la figure 6.14). Ainsi, nous pouvons observer que le résultat final est qualitativement similaire à celui du premier scénario : le champ génère une seule bulle d'activité neuronale à la fois, maintenant son caractère sélectif même dans ces conditions de quasi absence de stimulation externe.

Comme observé dans le premier scénario, la dynamique de l'activité auxiliaire v suit celle de l'excitation \mathcal{E} , avec un certain délai (voir les figures 6.13 (e) et 6.14 (a)). La croissance lente de v n'influence pas l'évolution immédiate de l'activité neuronale : u se stabilise, gardant après cette phase son niveau d'activité, tant que v n'atteint pas un niveau suffisamment élevé. De ce point de vue, la durée de vie d'une bulle d'activité neuronale, générée à partir d'une faible stimulation externe, est inversement proportionnelle au taux de croissance de v . Si, dans nos simulations nous avons besoin d'une durée de vie plus élevée pour les bulles d'activité, il est suffisant de modifier les valeurs positives de la fonction h qui contrôlent la dynamique de l'activité auxiliaire v . Ainsi, comme le montre la figure 6.15, en diminuant légèrement les valeurs des seuils positifs de la fonction h considérée précédemment dans nos simulations, nous obtenons des bulles d'activité d'une durée de vie supérieure, tout en gardant le comportement dynamique global du champ.

Notons ensuite qu'une fois que v dépasse une certaine amplitude, i étant toujours faible

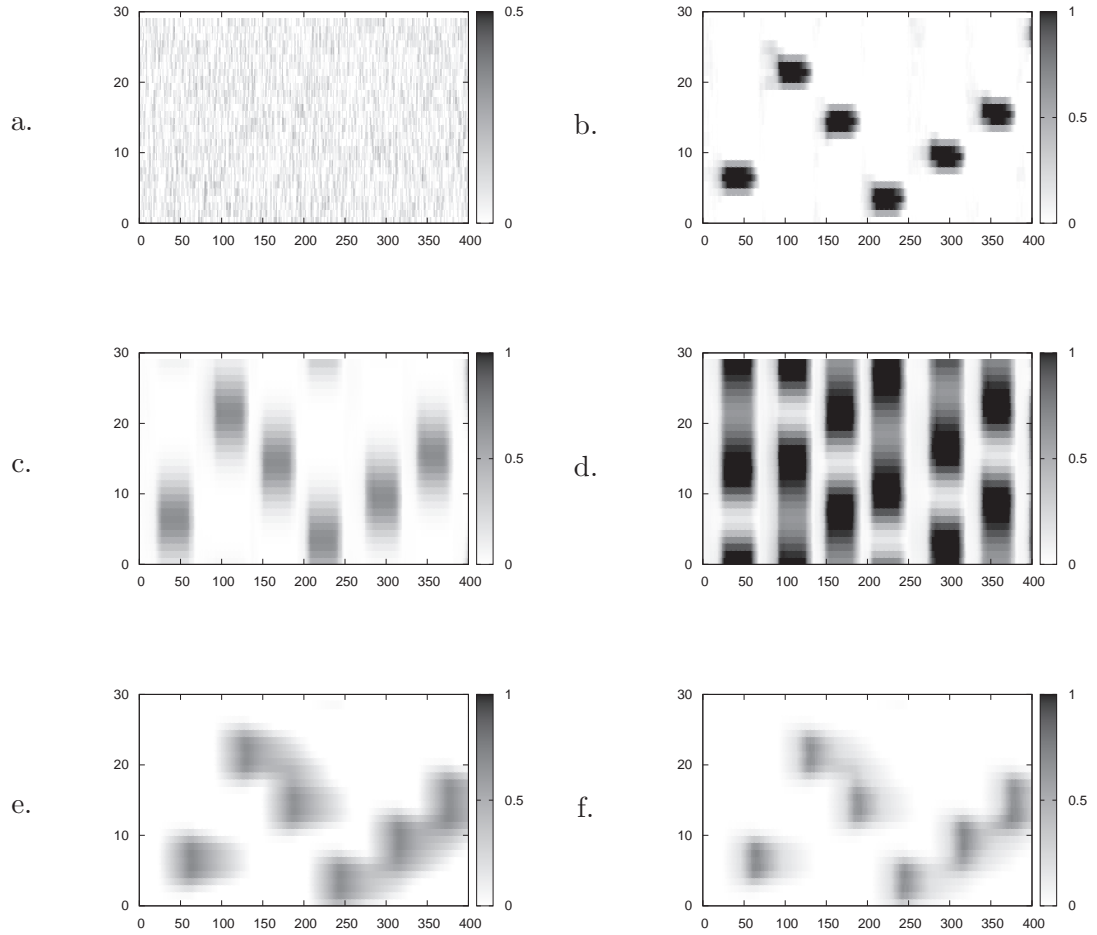


FIG. 6.13 – Illustration graphique de la dynamique de plusieurs paramètres du modèle CNDRI, pendant la simulation d'un scénario d'exploration. Toutes les diagrammes sont tracées en niveaux de gris, la couleur noir symbolisant l'intensité la plus forte du paramètre considéré. (a) le scénario de stimulation externe i ; (b) l'activité neuronale u du champ; (c) la contribution latérale excitatrice \mathcal{E} ; (d) la contribution latérale inhibitrice \mathcal{I} ; (e) l'activité auxiliaire v ; (f) l'intensité g de la rétro-inhibition locale.

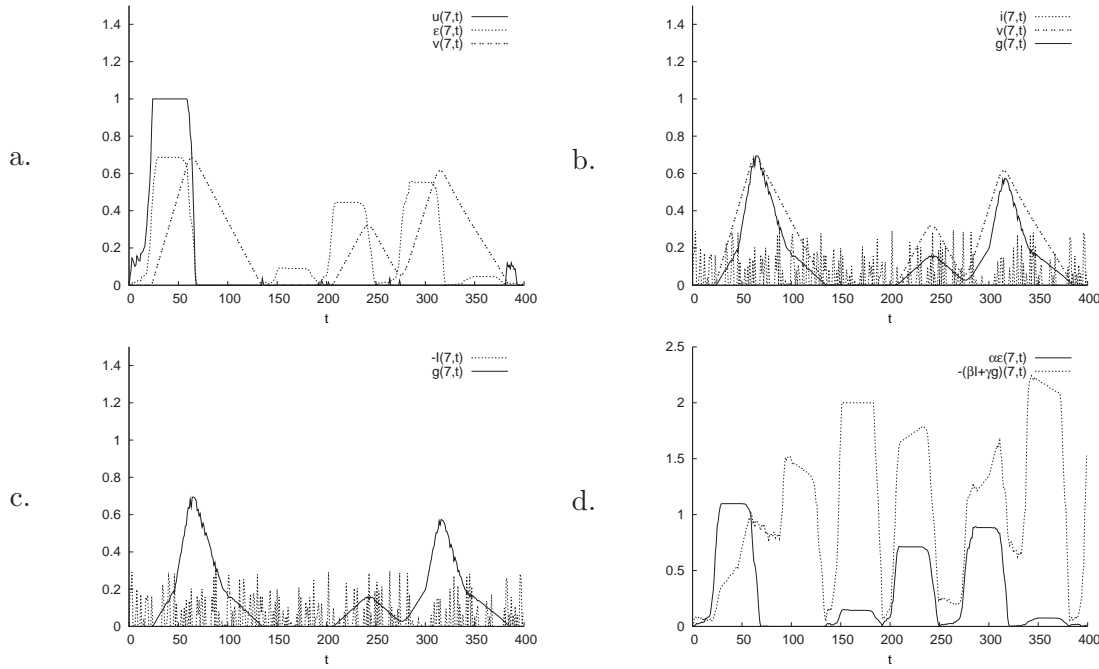


FIG. 6.14 – Détail sur le comportement du modèle CNDRI illustré graphiquement dans la figure 6.13. Analyse individuelle de certains paramètres du champ, uniquement pour l'unité $x = 7$ du champ. (a) dynamique des grandeurs u , \mathcal{E} et v ; (b) dynamique des grandeurs i , v et g . (c) dynamique des grandeurs $-\mathcal{I}$ et g . (d) dynamique des contributions cumulées excitatrices ($\alpha \mathcal{E}$) et inhibitrices ($-\beta \mathcal{I} - \gamma g$) du champ. Voir le texte pour plus d'explications.

(pendant toute la durée du scénario), g devient suffisamment élevée pour déclencher le mécanisme de rétro-inhibition locale du champ (voir les figures 6.13 (f) et 6.14 (b,c)). Par conséquent, l'influence induite par la dynamique de rétro-inhibition s'ajoute aux contributions latérales inhibitrices \mathcal{I} , renversant la domination des contributions excitatrices \mathcal{E} (voir le graphique (d) de la figure 6.14), et provoquant ainsi la remise à zéro de l'activité neuronale u .

Un dernier aspect qui peut être remarqué en étudiant la dynamique (corrélée) des activités u , \mathcal{E} et v (voir la figure 6.14 (a)) se réfère au fait que la baisse lente de l'activité auxiliaire v empêche, pour une certaine période de temps, la formation d'une nouvelle bulle d'activité u , au même endroit où la bulle qui vient d'être détruite par le mécanisme de rétro-inhibition était apparue.

Cas d'un champ unidimensionnel multi-bulles dans un scénario de compétition

Afin de permettre l'existence simultanée de plusieurs bulles d'activité neuronale u dans le champ, il est nécessaire de réajuster la distribution des contributions excitatrices et inhibitrices dans le champ. Comme celles-ci sont induites par les noyaux des poids des connexions latérales, modélisés par les fonctions w^+ et w^- , il est nécessaire de modifier les valeurs des paramètres a , b et c qui les définissent (voir les équations 6.1 et 6.2).

Afin de permettre au modèle du champ de générer deux bulles d'activité neuronale à la fois, nous avons choisi ces valeurs avec $a = 0.1$, $b = 0.03$ et $c = 0.15$, ajustant en conséquence les valeurs E_{max} et I_{max} des seuils des fonctions sigmoïdes σ^+ et σ^- .

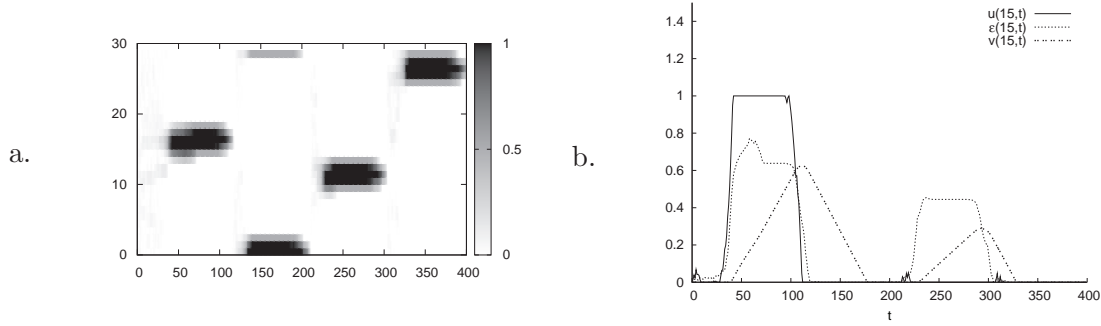


FIG. 6.15 – Étude de l'impact d'une légère modification de la fonction h (qui contrôle la dynamique de l'activité auxiliaire v) sur la réponse globale du champ. Analyse du même scénario d'exploration considéré précédemment dans la figure 6.13 (a). (a) la nouvelle distribution u de l'activité neuronale du champ ; (b) la dynamique des paramètres u , ε et v pour l'unité individuelle $x = 15$ du champ.

Comme nous allons le voir dans la suite du paragraphe, ainsi que dans le paragraphe suivant, aucune autre modification de paramètres n'est nécessaire. Ainsi, toutes les valeurs choisies antérieurement pour les autres paramètres du champ (α , β , γ , et les fonctions g et h), dans le contexte d'un champ unidimensionnel mono-bulle, restent identiques. D'après nous, cet aspect montre la simplicité de paramétrage du modèle CNDRI, et facilite son usage pratique dans la modélisation d'autres processus de traitement d'information.

La figure 6.16 montre l'évolution dynamique pour plusieurs variables du champ, dans le contexte du même scénario de compétition considéré dans le paragraphe 6.4.2. Comme nous pouvons l'observer à partir du graphique (b) de cette figure, le champ présente deux bulles d'activité neuronale, qui suivent le motif de stimulation externe illustré dans le graphique (a).

Comme attendu, dans cette nouvelle configuration des paramètres, il existe deux zones de dominance excitatrice dans le champ (comparer les graphiques (c) et (d) de la figure 6.16), fait qui induit l'existence de deux zones de forte intensité, à la fois pour l'activité auxiliaire v , et à la fois pour le mécanisme de rétro-inhibition g (voir les figures 6.16 (e) et (f)). La dynamique de tous ces paramètres suit la même évolution que celle observée et analysée dans le paragraphe 6.4.2, pour le cas du champ mono-bulle.

Examinant attentivement l'évolution de l'activité neuronale u (revoir la figure 6.16 (b)), nous pouvons remarquer également un effet de *collaboration locale*. Pendant les pas de simulation 100 et 130, et à nouveau pendant les pas 210 et 250 (approximativement), les deux bulles d'activité neuronale déjà existantes dans le champ se rassemblent, formant (pour une brève période transitoire), une seule zone d'excitation dominante dans le champ. Nous rappelons que ce phénomène de collaboration est un des comportements attendus pour la dynamique d'un champ neuronal (ceci étant une propriété souhaitable des champs neuronaux, comme illustré graphiquement dans la figure 5.6 (f)). Toutefois, comme nous pouvons l'observer, grâce au mécanisme de rétro-inhibition, si cette zone d'excitation n'est plus entretenue par une forte stimulation externe, elle est contrainte à s'inhiber de nouveau, après un certain délai. Enfin, nous remarquons que dans les périodes transitoires nécessaires pour rétablir la cohérence de l'activité du champ, la brève formation d'une bulle d'activité est possible (suivre l'activité du champ, dans le voisinage de l'unité $x = 0$, dans l'intervalle de simulation 250 – 300). Finalement, le surplus de stimulation apporté par la distribution i fait en sorte que l'activité du champ se dirige de nouveau vers cette

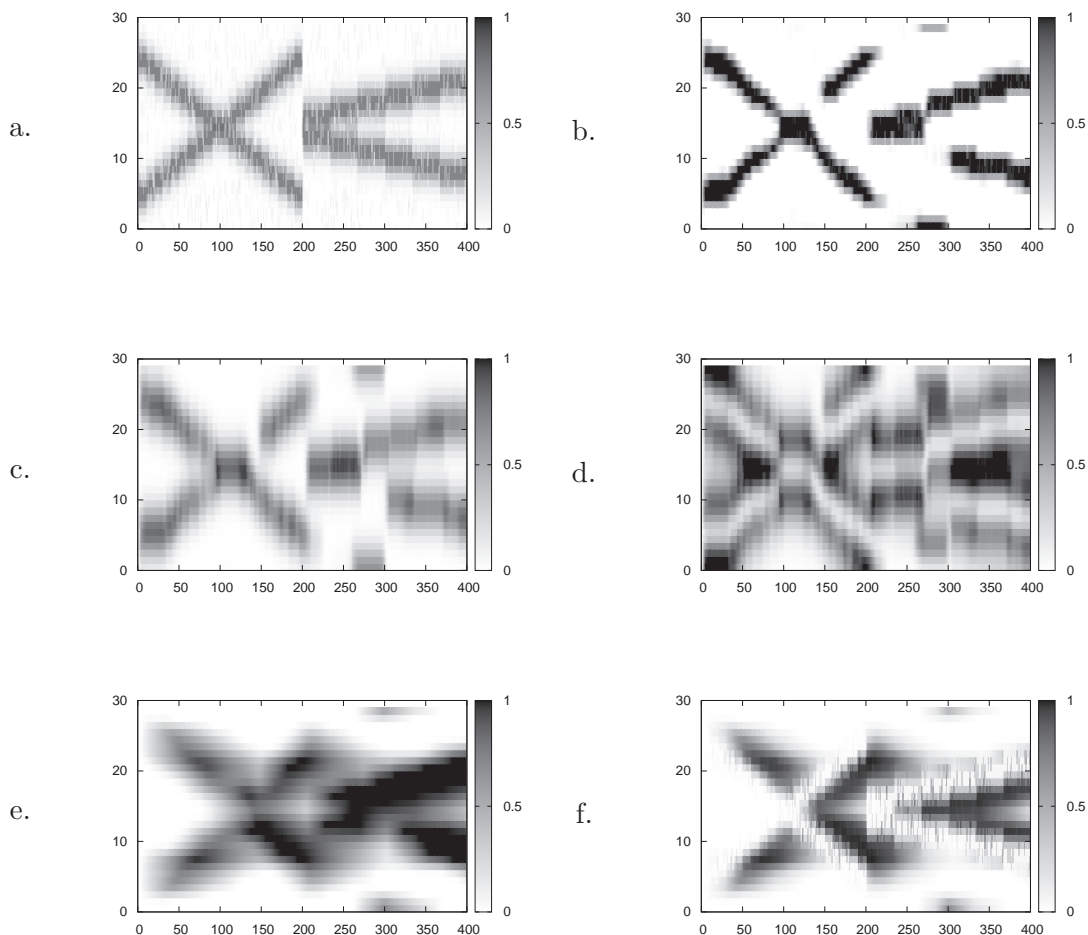


FIG. 6.16 – Illustration graphique de la dynamique de plusieurs paramètres du modèle CNDRI, pendant la simulation d'un scénario de compétition. Tous les diagrammes sont tracés en niveaux de gris, la couleur noir symbolisant l'intensité la plus forte du paramètre considéré. (a) le scénario de stimulation externe i ; (b) l'activité neuronale u du champ ; (c) la contribution latérale excitatrice \mathcal{E} ; (d) la contribution latérale inhibitrice \mathcal{I} ; (e) l'activité auxiliaire v ; (f) l'intensité g de la rétro-inhibition locale. Ici, les paramètres du champ ont été choisis tels qu'ils puissent rendre le champ capable de générer simultanément deux bulles d'activité neuronale u .

zone de stimulation, générant ainsi une nouvelle bulle d'activité qui la suit.

Cas d'un champ unidimensionnel multi-bulles dans un scénario d'exploration

Gardant les mêmes paramètres du modèle que dans le paragraphe précédent, nous examinons ici le comportement du champ, évoluant dans le même scénario d'exploration que celui considéré dans le paragraphe 6.4.2.

Comme le montre la figure 6.17, le champ est capable de générer simultanément deux bulles d'activité neuronale, dans ces conditions de faible stimulation externe. La dynamique illustrée ici pour les différents paramètres du champ est tout à fait similaire à celle tracée dans la figure 6.13 pour un champ mono-bulle. Ce fait expérimental confirme ainsi possibilité d'étendre ce comportement attendu pour un champ multi-bulles.

En conclusion, tous ces aspects comportementaux analysés dans tous ces derniers paragraphes montrent comment un champ mono-bulle ou multi-bulles peut garder son caractère sélectif et adaptatif, si nous suivons la procédure de paramétrage présentée dans la section 6.4.1.

6.5 Discussion

Dans ce chapitre, nous avons introduit un nouveau modèle de champ neuronal dynamique, conçu pour aider à la modélisation du processus d'auto-organisation corticale, tel qu'il a été décrit dans le chapitre 2. Le modèle, décrit formellement dans la section 6.1, s'inspire des approches de modélisation classiques, décrites dans le chapitre 3, et possède un nombre de particularités que nous résumons ci-dessous. Premièrement, le modèle CNDRI propose l'implantation d'un mécanisme de rétro-inhibition locale, qui a pour but la régulation de l'inhibition locale du champ, dans des conditions de forte activation et de faible stimulation locale. Ce mécanisme, qui peut être vu comme un mécanisme de feedback local, aide à déstabiliser les bulles d'activité neuronale qui ne sont pas soutenues par une forte stimulation externe, et crée les prémisses nécessaires pour l'émergence de nouvelles bulles d'activité dans d'autres endroits du champ. Deuxièmement, la séparation des contributions excitatrices et inhibitrices permet d'utiliser exclusivement l'influence excitatrice pour moduler l'apprentissage des prototypes pendant le processus d'auto-organisation. Cet aspect permet une différenciation plus graduelle de l'ajustement des valeurs des prototypes, et implique par conséquent une convergence plus rapide du processus d'auto-organisation. En troisième temps, nous soulignons que la formalisation théorique et l'implantation pratique du modèle CNDRI sont conformes aux principes de calcul neuronal énoncés dans la section 2.5. En particulier, tout le traitement de l'information est effectué de manière locale, distribuée et asynchrone.

Le modèle de champ, analysé en détail dans les paragraphes 6.2, 6.4 et 6.3, exhibe des caractéristiques dynamiques qui nous permettent de le considérer comme une extension des modèles classiques de champs neuronaux, présentés dans le chapitre 3. Ainsi, les résultats expérimentaux confirment à la fois la similarité entre les comportements dynamiques des modèles classiques (CNDA et CNDPE) et ceux du modèle que nous proposons (CNDRI), et la capacité du nouveau modèle d'étendre son spectre de sélectivité neuronale dans d'autres contextes de stimulation. Par conséquent, nous pouvons affirmer que le nouveau formalisme de champ élargit l'aire d'utilisation pratique du paradigme des champs neuronaux dans des contextes de modélisation neuronale peu explorés jusqu'à présent. De plus, nous considérons également que l'approche de modélisation proposée ici, basée entièrement sur des mécanismes locaux de calcul neuronal peut offrir à son tour une inspiration pour la conception d'autres modèles neuronaux.

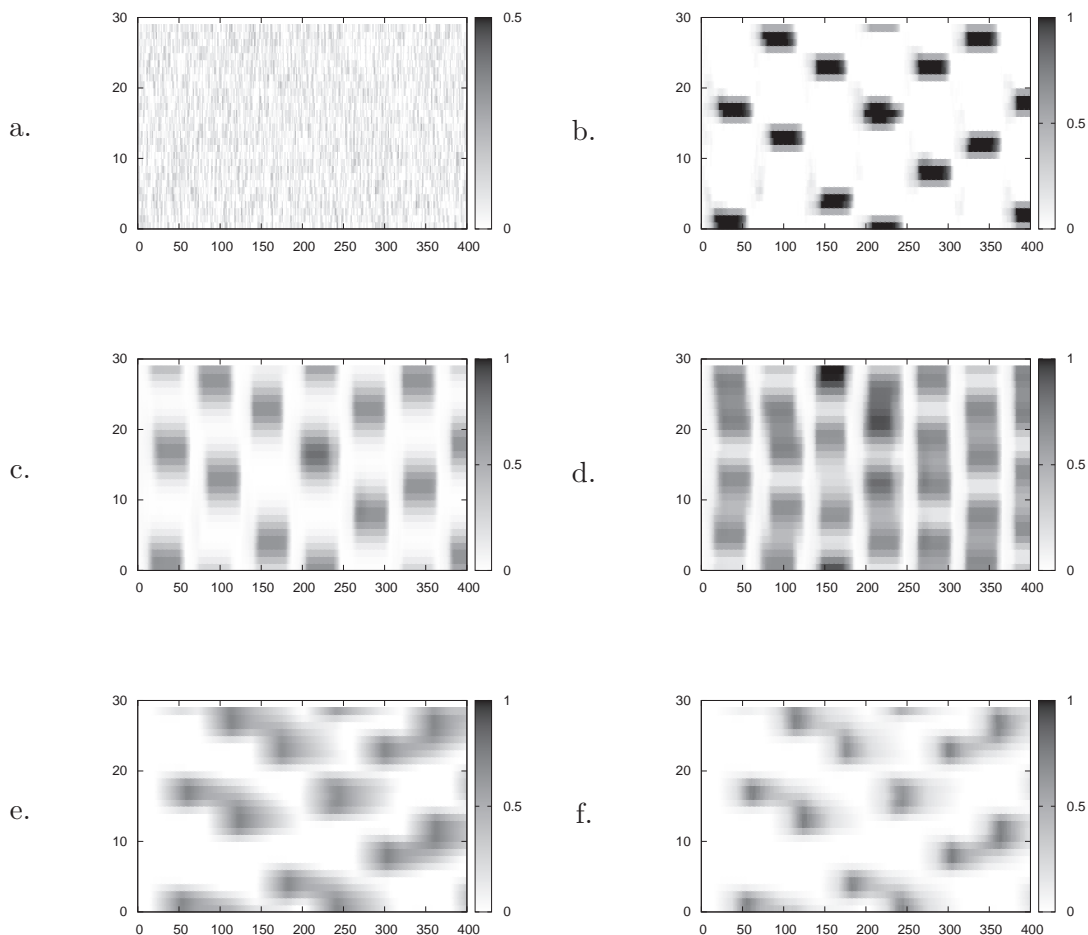


FIG. 6.17 – Illustration graphique de la dynamique de plusieurs variables du modèle CNDRI, pendant la simulation d'un scénario d'exploration. Tous les diagrammes sont tracés en niveaux de gris, la couleur noir symbolisant l'intensité la plus forte de la variable considérée. (a) le scénario de stimulation externe i ; (b) l'activité neuronale u du champ; (c) la contribution latérale excitatrice \mathcal{E} ; (d) la contribution latérale inhibitrice \mathcal{I} ; (e) l'activité auxiliaire v ; (f) l'intensité g de la rétro-inhibition locale. Ici, les paramètres du champ ont été choisis tels qu'ils puissent rendre le champ capable de générer simultanément deux bulles d'activité neuronale u .

En ce qui concerne les modèles neuronaux plus récents proposés dans la littérature, caractérisés par des phénomènes dynamiques de propagation d'ondes ou de fronts d'ondes d'activité neuronale, nous pouvons affirmer que le modèle CNDRI est complémentaire à ceux-ci. Bien que nous n'excluons pas la possibilité de modéliser un processus d'auto-organisation en utilisant ce type de dynamique neuronale, nous trouvons celle-ci difficilement contrôlable et peu appropriée pour assurer la stabilité de l'apprentissage en ligne des prototypes neuronaux. D'un autre point de vue, la ressemblance structurelle avec les formalismes des modèles classiques de champ nous a conduit de considérer la conservation d'un maximum de « compatibilité » avec les modèles classiques, préférant ainsi une extension de ceux-ci, plutôt qu'une reformulation radicale. Comme dernier argument, nous pouvons affirmer que nous avons veillé à garder la parcimonie des caractéristiques nécessaires à nos approches de modélisation, capables à la fois de répondre aux besoins réclamés par les contextes appliqués, et à la fois de permettre son extension future.

Afin de faciliter la compréhension et l'usage pratique du modèle CNDRI que nous avons proposé, dans la dernière partie de ce chapitre nous avons présenté une procédure détaillée pour paramétrer les équations du champ, ainsi qu'une analyse dynamique étendue de son comportement dans différents scénarios clé.

Nota : les travaux décrits dans ce chapitre ont fait l'objet des articles 3, 4 et 6, énumérés dans l'annexe A.

Chapitre 7

Conclusions

7.1 Résumé et analyse des résultats

Au début de ce mémoire, nous avons introduit notre objectif général de recherche, le plaçant dans le contexte informatique du domaine des sciences cognitives, et en particulier celui des neurosciences computationnelles. Nous réitérons ici que notre démarche vise à utiliser à la fois les résultats expérimentaux concernant la structure et le fonctionnement du système nerveux, et les paradigmes théoriques de modélisation neuronale, afin de concevoir des modèles de traitement d'information capables de s'intégrer dans le développement des solutions logicielles ou robotiques autonomes et adaptatives. Promouvant un « retour à la nature » pour s'inspirer des solutions que la nature a développé au cours de l'évolution, nous nous sommes focalisé sur les aspects anatomiques et fonctionnels de la structure corticale, afin de décortiquer certains mécanismes sous-jacents des processus cognitifs. Suite à cette analyse, nous nous sommes donné l'objectif de modélisation du processus d'auto-organisation des cartes corticales.

Par conséquent, dans le deuxième chapitre nous avons proposé une architecture neuronale générique, qui modélise ce processus d'auto-organisation de nature corticale. Abordant une approche de modélisation qui intègre les paradigmes computationnels, connexionnistes et dynamiques, nous avons proposé ensuite une implantation formelle de celle-ci. Dans la suite de ce mémoire, nous avons décrit notre quête pour trouver l'ensemble des éléments de modélisation, nécessaires pour accomplir cet objectif.

Ainsi, après avoir introduit le formalisme des champs neuronaux dynamiques, qui est un paradigme général de compétition neuronale, et qui vise la modélisation au niveau mésoscopique de l'activité corticale, nous avons présenté plusieurs modèles de champ proposés dans la littérature et nous avons analysé au fur et à mesure leurs caractéristiques structurales et fonctionnelles, dans un contexte à la fois théorique et appliqué.

Afin de mieux étudier les propriétés dynamiques des modèles de champs, dans le chapitre 5 nous avons décrit une plate-forme d'évaluation numérique, capable d'examiner la capacité d'un modèle quelconque de champ neuronal à satisfaire certains critères dynamiques, dans certaines conditions de stimulation. Appliquant cet outil d'évaluation à l'étude des modèles classiques des champs neuronaux (CNDA et CNDPE), nous avons pu mettre en évidence leur capacité à respecter les propriétés dynamiques ($P \wedge Q$), dans différents scénarios de stimulation. Ces résultats ont indiqué que ces modèles de champ sont appropriés à être utilisés dans des contextes de modélisation des processus de compétition, d'attention, ou de planification, fait confirmé par l'analyse des approches pratiques de modélisation réalisée précédemment. Notons que l'instrument d'évaluation proposé dans ce contexte nous a permis également de concevoir une procédure d'ajustement

des paramètres des modèles de champ, afin d'améliorer leurs comportements dynamiques dans certains scénarios de stimulation neuronale.

Cependant, grâce à la même analyse, nous avons pu constater qu'aucun des deux modèles classiques de champ examinés ne réussit à exhiber des résultats satisfaisants dans le cas des scénarios d'auto-organisation. Pour analyser les causes de cet échec, nous nous sommes rapporté aux autres scénarios individuels, examinés au cours de notre étude empirique du comportement des champs neuronaux. Dans ce contexte, nous avons observé que aucun des deux modèles de champ analysés n'est capable de générer une réponse sélective (une bulle d'activité) dans des conditions de faible stimulation externe, et adapter ensuite cette réponse en fonction des changements survenus dans la distribution des stimuli externes. Cette remarque nous avait conduit à affirmer que les résultats insatisfaisants des scénarios d'auto-organisation sont étroitement liés à l'incapacité de ces champs à maintenir leur caractère sélectif dans le cas des scénarios d'exploration de faible stimulation externe.

Pour tester cette affirmation, nous avons fait appel à l'instrument d'évaluation $\bar{\Delta}_i^{\mathcal{B}}(u)$, concevant cette fois-ci une activité de champ qui émule la dynamique d'un champ neuronal respectant entièrement les propriétés dynamiques $(P \wedge Q)$. Bien qu'elle recourt à une procédure combinatoire de calcul qui s'avère très coûteuse, cette démarche nous a permis de confirmer empiriquement qu'un champ neuronal capable de satisfaire entièrement $(P \wedge Q)$ (qui maintient donc son caractère sélectif et adaptatif dans toutes les conditions de stimulation), est aussi capable de conduire l'implantation du processus d'auto-organisation vers des résultats tout à fait satisfaisants.

Néanmoins, d'une part l'incapacité des modèles classiques de champs neuronaux à servir d'une manière appropriée à l'implantation du processus d'auto-organisation (cf. l'architecture proposée), et d'une autre part le coût d'implantation très élevé des champs de référence (les champs dont l'activité est calculée selon la procédure combinatoire évoquée ci-dessus) nous ont conduit à chercher la conception d'un modèle alternatif de champ neuronal. Cette démarche, décrite dans le chapitre 6 et concrétisée sous la forme du modèle CNDRI, repose sur un mécanisme de *rétro-inhibition*, capable de réguler localement le niveau de l'inhibition du champ. Ce mécanisme local, qui devient actif dans les conditions de faible stimulation externe, permet l'émergence des bulles d'activité dans ce contexte particulier de stimulation, et déstabilise ces bulles après un certain temps, si elles continuent à ne pas être soutenues par un niveau significatif de stimulation, permettant ainsi la génération d'autres réponses sélective (bulles d'activité neuronale), ailleurs dans le champ.

Grâce à ce mécanisme neuronal que nous avons proposé, le nouveau modèle de champ s'est avéré capable d'implanter avec succès le module de sélection de l'architecture neuronale d'auto-organisation, conduisant à chaque fois les scénarios d'auto-organisation vers des résultats satisfaisants. En outre, grâce à la plate-forme d'évaluation générique du comportement d'un champ neuronal, nous avons pu constater également que le comportement de ce nouveau modèle est tout à fait similaire aux comportements exhibés par les champs classiques, dans les autres types de scénarios examinés. Grâce à ces remarques, nous pouvons affirmer que le modèle CNDRI constitue une extension viable des modèles classiques de champ neuronal proposés dans le domaine, capable d'étendre le contexte appliqué de ce paradigme de modélisation corticale.

L'ensemble de ces résultats nous font conclure que le modèle neuronal que nous avons proposé répond à l'objectif de recherche défini au début de ce mémoire, notre approche théorique s'inscrivant dans le cadre d'une démarche informatique de modélisation neuronale, basée sur les paradigmes computationnels, connexionnistes et dynamiques, et inspirée des modèles cognitifs corticaux.

7.2 Perspectives

7.2.1 Perspectives envisagées pour nos approches

Les travaux effectués au cours de notre démarche de recherche et présentés dans ce mémoire nous permettent d'envisager un certain nombre de perspectives de développement théorique et appliqué des résultats et approches résumés dans le paragraphe précédent.

Sous l'aspect théorique, nous envisageons dans un premier temps d'approfondir l'étude du comportement dynamique du modèle CNDRI, afin de le rendre entièrement compatible avec les propriétés dynamiques $(P \wedge Q)$. Bien que le modèle CNDRI réussisse à conduire avec succès l'implantation du processus d'auto-organisation décrit par l'architecture d'inspiration corticale proposée dans le chapitre 2, nous rappelons qu'aucun des modèles de champs neuronaux (CNDA, CNDPE et CNDRI compris) ne réussit à satisfaire complètement ces propriétés. Ainsi, dans la section 6.3.1, nous avons pu mettre en évidence un comportement du modèle CNDRI qui ne satisfait pas les propriétés $(P \wedge Q)$, considérées comme souhaitables pour la dynamique d'un champ neuronal. Ce comportement est caractérisé par la capacité du champ à générer une bulle d'activité en condition de stimulation locale forte, mais aussi par une incapacité du champ à adapter sa réponse vis-à-vis d'un nouveau stimulus d'une intensité plus élevée, qui stimule une autre zone du champ. Ce comportement, qui caractérise aussi les modèles classiques de champs neuronaux (CNDA et CNDPE), est dû au fait que le maintien d'une bulle d'activité neuronale déjà générée nécessite l'inhibition de toutes les zones du champ voisines à cette bulle. Dans ces zones du champ, l'inhibition domine toute autre stimulation externe. Ce phénomène empêche le champ de modifier le profil de son activité, tant que la bulle existante continue à être soutenue par une source de stimulation suffisamment forte. Si nous souhaitons améliorer ce comportement dynamique du champ, nous devons recourir à modifier ce mécanisme. Dans nos travaux futurs, nous envisageons d'examiner plusieurs approches pour résoudre ce problème. Une première approche que nous visons d'explorer suggère une extension du mécanisme de rétro-inhibition, que nous avons proposé pour le modèle CNDRI, pour déclencher périodiquement l'inhibition locale du champ, indépendamment de l'intensité de stimulation perçue par le champ. Une deuxième approche que nous envisageons d'analyser s'inspire des travaux de Erlhagen, Coombes et Owen, présentés dans la section 3.2.6, qui avaient proposé une approche d'ajustement graduel du niveau de repos de l'activité du champ.

Dans un deuxième temps, nous souhaitons étendre l'exploration des capacités dynamiques des champs neuronaux, et du modèle CNDRI en particulier, pour des scénarios de stimulation plus complexes et plus variés que ceux déjà analysés, utilisant le même outil d'évaluation dynamique présenté dans ce manuscrit.

À un niveau plus abstrait de développement théorique, nous sommes intéressé à continuer d'explorer l'approche générale de *calcul dynamique* (c'est-à-dire implanter des mécanismes fonctionnels basés sur les effets computationnels émergents des systèmes dynamiques neuronaux), décrite dans la section 2.3.2. Dans ce contexte, nous rappelons que les bulles exhibées par le comportement d'un champ neuronal constituent les attracteurs du système dynamique implanté par l'architecture d'auto-organisation d'inspiration corticale, proposée dans la section 2.6. Suivant cette ligne de recherche, nous sommes intéressé à étudier comment utiliser les propriétés dynamiques des modèles de champs neuronaux pour implanter d'autres mécanismes génériques de traitement de l'information et de l'action, comme par exemple la sérialisation ou la spatialisation de l'aspect temporel de ces traitements, ou la séparation et la distribution optimale des tâches de traitement, etc.

Sous l'aspect pratique de la modélisation, nous visons d'abord à valoriser notre modèle de

calcul neuronal dans le développement des systèmes situés, capables d'exploiter les propriétés dynamiques que nous avons mises en évidence dans le chapitre 6. Plus spécifiquement, nous souhaitons aborder la problématique du traitement sensorimoteur multimodal, utilisant le modèle neuronal de *bijama*, décrit dans la section 4.4.2. Dans ce contexte, nous sommes également intéressé à étudier comment les connexions cortico-corticales à longue distance contribuent à la régulation et au maintien de la cohérence du fonctionnement dynamique d'un système neuronal complexe.

En outre, et lié à cet aspect de modélisation du traitement multimodal, nous souhaitons étudier comment intégrer le paradigme de l'apprentissage par renforcement dans la modélisation d'inspiration corticale basée sur la dynamique des champs neuronaux, continuant ainsi la ligne des applications qui abordent le problème cinématique inverse et le problème d'utilité, décrits et analysés dans la section 4.3.

7.2.2 Perspectives envisagées pour le domaine

L'intérêt croissant vis-à-vis du domaine des neurosciences, pour examiner des questions fondamentales concernant l'origine et la nature de la connaissance et des processus cognitifs, ainsi que l'effervescence de recherche dans le domaine de l'intelligence artificielle, ont conduit dans les dernières décennies à un développement accéléré des recherches, dont nous sommes capables à peine d'éprouver le potentiel des résultats.

À l'aube de cet horizon des recherches, envisager les perspectives sur le long terme du domaine peut s'avouer une entreprise assez difficile. Cependant, mises dans un contexte plus large, nous pouvons extrapoler sur l'avenir de directions générales de recherche visées par ce domaine.

Dans ce contexte, rappelons que l'approche computationnelle (et connexionniste en particulier) postule que tout processus cognitif peut être simulé exclusivement par des moyens de calcul. Notons qu'un des aspects centraux visés par cette approche concerne la modélisation des mécanismes neuronaux d'apprentissage, les travaux que nous avons décrits le long de ce mémoire s'inscrivant dans cette démarche générale de recherche. Étayés sur les résultats expérimentaux qui illustrent la plasticité neuronale d'une multitude de zones cérébrales (et particulièrement corticales), les chercheurs ont élaboré différents modèles d'apprentissage neuronal. Cependant, tout ce qui différencie ces modèles entre eux se réfère à ces trois aspects descriptifs (que nous allons noter SPL) : la structure topologique des connexions des unités neuronales, les poids synaptiques de ces connexions, et les lois qui régissent l'ajustement dynamique de ces poids synaptiques.

Les conséquences de cette perspective sur les processus d'apprentissage neuronal sont multiples. Premièrement, celle-ci met en valeur les vertus de la modélisation robotique et des systèmes situés : une fois qu'un robot (système) a réussi à apprendre comment résoudre une certaine tâche cognitive, il est suffisant d'enregistrer son triplet SPL et de l'installer (comme un produit logiciel quelconque) dans le support matériel de n'importe quel autre robot (système) compatible, pour que ce nouveau robot (système) bénéficie directement des compétences acquises par le premier robot, sans repasser par une phase initiale d'entraînement du processus d'apprentissage. Notons que cette vision de modélisation neuronale permet de passer de la phase de « câblage de connaissances » à une phase de « câblage de savoir-faire ».

Deuxièmement, cette perspective a conduit certains chercheurs comme O. Sporns ou P. Hagmann à envisager le projet ambitieux de réaliser la cartographie complète de toutes les connexions neuronales d'un cerveau humain. Si une démarche similaire a été accomplie récemment pour le nématode *Caenorhabditis elegans* [73, 122], la démarche pour le cerveau humain est en cours de déroulement, étant conduite à présent par le *Human Connectome Project* [1]. Notons que ce concept de *connectome*, comme son nom le suggère, peut s'inscrire dans la grande ligne d'évo-

lution de l'organisation de la matière. Ainsi, le biologiste G. J. op Akkerhuis postule dans [90] qu'une possible nouvelle étape dans l'évolution de la vie serait représentée par la transmission par reproduction non seulement des gènes (qui encodent les instructions nécessaires pour le développement structural et le fonctionnement physiologique de l'organisme), mais aussi des mèmes (qui encodent les connaissances acquises pendant la vie) de l'individu. Remarquons que la notion des mèmes n'est pas une notion abstraite de cette théorie, étant déjà présente dans nos vies, sous la forme des livres, des ressources multimédia, ou bien des sites Internet. Pourtant, dans cette perspective, le connectome pourrait très bien faire fonction de transmission directe des connaissances qui ne peuvent pas être acquises explicitement sans une expérience pratique (voir par exemple la capacité de parler une langue, ou celle de faire du vélo). Bien que pour le moment la technologie pour implanter cette idée soit au tout début de la phase de recherche et demande des ressources de calcul extrêmement puissantes et coûteuses, elle pourrait un jour être mise en œuvre.

Enfin, dans cette perspective à long terme sur les résultats des recherches des neurosciences, il est peu probable de supposer que l'Internet, l'outil le plus complexe créé par l'homme, ne subira pas des transformations radicales. Le web sémantique est juste une phase de développement longuement attendue, mais pour la vraie image d'un Internet bio-inspiré et neuro-inspiré il reste à l'avenir de nous la faire découvrir.

7.2.3 Défis de l'avenir

Dans ces derniers paragraphes nous avons esquissé une vision de l'avenir concernant à la fois nos propres démarches de recherche et certaines perspectives envisagées sur le long terme pour le domaine des neurosciences.

Sur le plan de la recherche théorique, les neurosciences ont un nombre significatif de problèmes fondamentaux à affronter.

Comme nous l'avons précisé dès l'introduction de ce mémoire, notre approche informatique repose sur une vision computationnelle et dynamique de la connaissance. Cependant, le débat sur la nature ontologique de la connaissance et des processus cognitifs reste ouvert. Seul l'avenir nous dira si les recherches futures peuvent trancher définitivement entre la vision computationnelle de Turing et al., et la vision non-computationnelle de Searle, Penrose et al.

Toutefois, peut-être la question la plus fascinante de toutes ces recherches fondamentales dans les neurosciences concerne la nature des processus mentaux de la conscience. Actuellement, une communauté extrêmement importante est consacrée à ce sujet. Des progrès significatifs ont été enregistrés grâce aux travaux de Baars [7], Crick et Koch [25] ou Edelman [32], mais comme dans le cas des processus cognitifs ou de l'intelligence d'un système cognitif, l'origine et la portée du phénomène de la conscience sont vivement débattues par les chercheurs et les philosophes. Nous rappelons à ce propos que dans la section 4.2.3, nous avons aussi analysé une démarche d'utilisation du paradigme des champs neuronaux dynamiques pour modéliser certains aspects liés aux processus de la conscience.

Cependant, à part ces défis d'ordre théorique du domaine des neurosciences, les conséquences pratiques d'utilisation de ses résultats rencontrent également de plus en plus visiblement d'autres défis. Si dès l'introduction de ce mémoire nous avons évoqué l'influence des débats philosophiques dans les fondations théoriques des sciences cognitives, il est attendu (comme par exemple dans le cas des technologies d'ingénierie génétique) que l'avancement de ce domaine provoque de nombreux débats éthiques et sociaux.

Pour l'instant, les conséquences du phénomène de « délocalisation » de l'intelligence humaine (le fait de confier aux instruments créés par l'homme la capacité à s'adapter et à devenir auto-

nomes, pour résoudre des problèmes complexes) ne sont que très peu débattues. Cependant, la *singularité technologique* (l'événement au-delà duquel l'homme ne pourra plus être capable de comprendre la complexité des instruments qu'il a créés, afin de les adapter ou de les contrôler) reste pour l'instant un concept et un sujet de débat, plutôt qu'une question de politique. Ironiquement, le concept même nous « assure » que, s'il s'avère vrai, il passera inaperçu avant que ses conséquences soient possibles d'être renversées.

Remarquons dans ce contexte que tous les instruments développés par l'homme (y compris l'ensemble des théories scientifiques) sont issus des nécessités spécifiques, mais aussi d'un incessant appétit de l'homme pour explorer l'inconnu. Ces nécessités sont celles qui formulent nos *problèmes d'utilité*. Si nous avons analysé ce type de problèmes dans la section 4.3.1, d'un point de vue de la modélisation neuronale, la vie même se pose ces problèmes à chaque instant. Nos cultures, nos sociétés, nos technologies, sont toutes des « boîtes à outils » nous offrant des solutions possibles pour traiter ces problèmes. Or, depuis les résultats de Gödel sur la non décidabilité de certaines affirmations logiques, la méthode scientifique (la réfutation logique des affirmations) est incapable de nous offrir une validation complète de nos théories et de nos résultats. De plus, d'une manière générale, le temps et l'effort nécessaires pour appréhender, contrôler et réadapter un objet quelconque (physique ou mental) augmente, de manière non-linéaire, avec la complexité de cet objet (remarquons à ce propos qu'il nous a fallu le travail d'une thèse pour comprendre et contrôler la dynamique des modèles de champ neuronal pour atteindre notre objectif initial de recherche). Ces aspects soulignent la difficulté mais aussi l'incertitude et l'inconnu qui demeurent à l'intérieur de tout objet de la réalité, qu'il soit découvert ou créé. Les systèmes dynamiques complexes de type économique ou social nous offrent un exemple éloquent d'objet créé par les interactions humaines, dont l'incertitude et le risque sous-jacents sont permanents et souvent impossibles à évaluer. De ce point de vue, il est attendu que la complexité croissante des systèmes intelligents résultant des recherches des sciences cognitives, couplée avec la complexité des systèmes dynamiques déjà existants (comme ceux sociaux ou économiques déjà mentionnés), sous-entendra des risques et des incertitudes accrus.

Les aspects énumérés ici nous montrent que toutes nos activités de recherche nous demandent un double acte de responsabilité : un pour valider la cohérence de nos modèles théoriques et un pour évaluer et assumer les incertitudes vis-à-vis des conséquences pratiques de ces modèles théoriques.

La nature est un système dynamique complexe dont l'homme fait partie. Grâce à sa curiosité et à son effort, l'homme a pu découvrir et étudier au cours de son histoire une multitude impressionnante d'aspects que la nature peut révéler. Plus que jamais, l'homme est témoin d'une ère d'une effervescence inouïe, en train pour la première fois dans son histoire d'être capable d'aborder et de répondre à des questions fondamentales concernant l'origine et la nature de la vie, la connaissance et la conscience de l'esprit. Devant le défi tout à fait remarquable posé par la compréhension de ces phénomènes, repose un défi encore plus grand pour l'homme et sa civilisation : celui de construire un équilibre dynamique capable d'intégrer à la fois lui-même, ses connaissances, et la nature. Les neurosciences et l'étude des solutions développées par la nature au cours de son évolution peuvent s'avérer des outils inestimables pour accomplir cette tâche.

Dans l'introduction de ce mémoire, nous avons affirmé que l'acte de connaissance est un acte créateur, un acte amené à donner du sens à un flot d'information. Dans l'univers des connaissances, la démarche scientifique se voit comme une méthode pour valider la cohérence de ces connaissances, par rapport à un ensemble d'éléments initiaux considérés irréductibles, et appelés hypothèses. Grâce aux recherches récentes, la démarche scientifique peut s'atteler maintenant aux problématiques concernant la création des connaissances, c'est-à-dire la compréhension et la modélisation des processus cognitifs. Les portes ouvertes par toutes ces connaissances, ainsi

que les chemins auxquels elles nous invitent, restent à être explorés soigneusement, afin de les *connaître* véritablement.

Annexe A

Liste de publications

Cette thèse a fait l'objet des publications suivantes :

1. L. Alecu, H. Frezza-Buet, "An Empirical Evaluation Framework for Qualifying Dynamic Neural Fields". In Proceedings of the second french conference on Computational Neuroscience, Neurocomp, Marseille, Laurent E. Perrinet and Emmanuel Dauce, ISBN : 978-2-9532965-0-1 :(4 pages), 2008.
2. L. Alecu, H. Frezza-Buet, "Are Neural Fields Suitable for Vector Quantization ?". In Proceedings of The Seventh International Conference on Machine Learning and Applications (ICMLA'08), IEEE :239-244, San Diego (CA, USA), 2008.
3. L. Alecu, H. Frezza-Buet, "Reconciling neural fields to self-organization". In European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning (ESANN), Bruges, Belgium, Michel Verleysen, ISBN : 2-930307-09-9 :571-576, 2009.
4. L. Alecu, H. Frezza-Buet, "A dynamic neural field mechanism for self-organization". In Supplement of the online journal BMC Neuroscience, The nineteenth Annual Computational Neuroscience Meeting, 10(Suppl 1) :P273, Berlin (Germany), 2009.
5. L. Alecu, H. Frezza-Buet, "Application-driven parameter tuning methodology for dynamic neural field equations". In Neural Information Processing, ICONIP'09 Proceedings, Part I, ISBN : 978-3-642-10676-7, Springer Berlin / Heidelberg, Lecture Notes in Computer Science, 5863/2009 :135-142, Bangkok (Thailand), 2009.
6. L. Alecu, H. Frezza-Buet, F. Alexandre, "Can self-organization emerge through dynamic neural fields computation ? ". In Connection Science, 23(1) :1-31, 2011.

Bibliographie

- [1] Human Connectome Project. Website, 2011. <http://humanconnectome.org>.
- [2] F. Alexandre. Cortical basis of communication : Local computation, coordination, attention. *Neural Networks*, 22(2) :126–133, 2009.
- [3] F. Alexandre, J. Fix, A. Hutt, N. P. Rougier, and T. Viéville. On practical neural field parameters adjustment. In *Deuxième conférence française de Neurosciences Computationnelles*, "Neurocomp08", Marseille France, 10 2008. ISBN : 978-2-9532965-0-1.
- [4] S.-I. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27 :77–87, 1977.
- [5] S.-I. Amari. Topographic organization of nerve fields. *Bulletin of Mathematical Biology*, 42 :339–364, 1980.
- [6] M. L. Andermann and C. I. Moore. A somatotopic map of vibrissa motion direction within a barrel. *Nat Neurosci*, 9(4) :543–551, 2006.
- [7] B. J. Baars. *A cognitive theory of consciousness*. Cambridge University Press, 1993.
- [8] R. Ben-Yishai, R. Lev Bar-Or, and H. Sompolinsky. Theory of orientation tuning in visual cortex. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 92, pages 3844–3848. National Academy of Sciences, 1995.
- [9] D. Bendor and X. Wang. The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7056) :1161–1165, 2005.
- [10] T. Bergener, C. Bruckhoff, P. Dahm, H. Janßen, F. Joublin, R. Menzner, and A. Steinhage. Complex behavior by means of dynamical systems for an anthropomorphic robot. *Neural Networks*, 12 :1087–1099, 1999.
- [11] J. Bower. GENESIS (GEneral NEural SIMulation System) software library. Website, 12 2010.
- [12] P. C. Bressloff, J. D. Cowan, M. Golubitsky, P. J. Thomas, and M. C. Wiener. What geometric visual hallucinations tell us about the visual cortex. *Neural Computation*, 14 :473–491, 2002.
- [13] P. C. Bressloff and S. E. Folias. Front bifurcations in an excitatory neural network. *SIAM Journal of Applied Mathematics*, 65(1) :131–151, 2004.
- [14] J. Bullier. Integrated model of visual processing. *Brain Research Reviews*, 36(2-3) :96–107, 2001.
- [15] D. V. Buonomano and W. Maass. State-dependent computations : spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10(2) :113–125, 2009.
- [16] Y. Burnod. *An adaptive neural network : The cerebral cortex*. Masson, 1989.

- [17] S. R. Cajal. *Les nouvelles idées sur la structure du système nerveux chez l'homme et chez les vertébrés*. C. Reinwald, 1894.
- [18] S. R. Cajal. *Comparative study of the sensory areas of the human cortex*. 1899.
- [19] S. Camazine, N. R. Franks, J. Sneyd, E. Bonabeau, J.-L. Deneubourg, and G. Theraula. *Self-Organization in Biological Systems*. Princeton University Press, 2001.
- [20] A. Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2) :345–363, 1936.
- [21] P. Churchland and T. J. Sejnowski. *The computational brain*. The MIT Press, 1994.
- [22] S. Coombes. Waves, bumps, and patterns in neural field theories. *Biological Cybernetics*, 93(2) :91–108, 2005.
- [23] S. Coombes and M. R. Owen. Exotic dynamics in a firing rate model of neural tissue with threshold accommodation. *Contemporary Mathematics*, (440) :123–144, 2007.
- [24] M. Cottrell, J. C. Fort, and G. Pagès. Theoretical aspects of the som algorithm. *Neuro-computing*, 21 :119–138, 1998.
- [25] F. Crick and C. Koch. Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 1990.
- [26] N. Cuperlier, M. Quoy, P. Laroque, and P. Gaussier. Transition cells and neural fields for navigation and planning. In José Mira and José Alvarez, editors, *Mechanisms, Symbols, and Models Underlying Cognition*, volume 3561 of *Lecture Notes in Computer Science*, pages 147–152. Springer Berlin / Heidelberg, 2005.
- [27] H. Dale. Pharmacology of nerve-endings. In *Proceedings of Royal Society of Medicine*, volume 28, pages 319–332, 1934.
- [28] P. Dayan and L. F. Abbot. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001. 978-0262041997.
- [29] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston. The dynamic brain : from spiking neurons to neural masses and cortical fields. *PLoS Computational Biology*, 4(8) :e1000092, 08 2008.
- [30] J. Eccles. From electrical to chemical transmission in the central nervous system. *Notes and Records of the Royal Society of London*, 1976.
- [31] J. C. Eccles. Evolution of consciousness. *Proceedings of the National Academy of Sciences Usa*, 89 :7320–7324, 1992.
- [32] G. Edelman. *The remembered present : a biological theory of consciousness*. Basic Books, 1990.
- [33] T. Elbert and B. Rockstroh. Reorganization of human cerebral cortex : the range of changes following use and injury. *The Neuroscientist*, 10(2) :129–141, 2004.
- [34] T. Elbert, A. Sterr, B. Rockstroh, C. Pantev, M. M. Müller, and E. Taub. Expansion of the tonotopic area in the auditory cortex of the blind. *Journal of Neuroscience*, 22(22) :9941–9944, 2002.
- [35] W. Erlhagen. Internal models for visual perception. *Biological Cybernetics*, 88(5) :409–417, 2003.
- [36] W. Erlhagen and E. Bicho. The dynamic neural field approach to cognitive robotics. *Journal of Neural Engineering*, 3 :36–54, 2006.

-
- [37] B. Ermentrout. XPPAUT (X-Windows Phase Plane plus Auto) software library. Website, 12 2010. <http://www.math.pitt.edu/~bard/xpp/xpp.html>.
 - [38] J. Fix, N. Rougier, and F. Alexandre. A top-down attentional system scanning multiple targets with saccades. In *From Computational Cognitive Neuroscience to Computer Vision : CCNCV 2007*, 2007.
 - [39] S. E. Folias and P. C. Bressloff. Breathing pulses in an excitatory neural network. *SIAM Journal of Applied Dynamical Systems*, 3(3) :378–407, 2004.
 - [40] S. E. Folias and P. C. Bressloff. Breathers in two-dimensional neural media. *Physical Review Letters*, 95, 2005.
 - [41] H. Frezza-Buet. BIJAMA (Biologically Inspired Joint Associative MAPs) software library. Website, 2010. <http://ims.metz.supelec.fr/spip.php?rubrique149>.
 - [42] H. Frezza-Buet. Intercell project. Website, 2011. <http://intercell.metz.supelec.fr>.
 - [43] B. Fritzke. A growing neural gas network learns topologies. In *Advances in Neural Information Processing Systems 7*, pages 625–632. MIT Press, 1995.
 - [44] P. Fuhr, L. G. Cohen, N. Dang, T. W. Findley, S. Haghighi, J. Oro, and M. Hallett. Physiological analysis of motor reorganization following lower limb amputation. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 85(1) :53–60, 1992.
 - [45] J. M. Fuster. *The prefrontal cortex*. Academic Press, 2008.
 - [46] S. B. Nelson G. G. Turrigiano. Homeostatic plasticity in the developing nervous system. *Nat Rev Neurosci*, 2004.
 - [47] M. Gardner. The fantastic combinations of John Conway’s new solitaire game “life”. *Scientific American*, 223 :120–123, 1970.
 - [48] C. Gläser, F. Joublin, and C. Goerick. Enhancing topology preservation during neural field development via wiring length minimization. In *ICANN ’08 : Proceedings of the 18th International Conference on Artificial Neural Networks, Part I*, pages 593–602. Springer-Verlag, 2008.
 - [49] C. Gläser, F. Joublin, and C. Goerick. Homeostatic development of dynamic neural fields. In *Proc. International Conference on Development and Learning 2008*, pages 121–126, 2008.
 - [50] D. Golomb and Y. Amitai. Propagating neuronal discharges in neocortical slices : Computational and experimental study. *J Neurophysiol*, 78(3) :1199–1211, 1997.
 - [51] C. Gros. *Complex and adaptive dynamical systems, a primer*. Complexity Series. Springer, 2008. 978-3540718734.
 - [52] C. Gros. Cognitive computation with autonomously active neural networks : an emerging field. *Cognitive Computation*, 1 :77–90, 2009.
 - [53] H.-M. Gross, V. Stephan, and M. Krabbes. A neural field approach to topological reinforcement learning in continuous action spaces. In *In Proc. of WCCI-IJCNN’98, Anchorage*, pages 1992–1997. IEEE Press, 1998.
 - [54] S. Grossberg. Adaptive pattern classification and universal recoding. i. : parallel development and coding of neural feature detectors. pages 243–258, 1988.
 - [55] T. Hansen, W. Sepp, and H. Neumann. Recurrent long-range interactions in early vision. In *Emergent Neural Computational Architectures based on Neuroscience*, pages 127–138. Springer-Verlag, 2001.

- [56] I. Harvey. Cognition is not computation ; Evolution is not optimisation. In *Proceedings of the 7th International Conference on Artificial Neural Networks, ICANN '97*, pages 685–690. Springer-Verlag, 1997.
- [57] D. O. Hebb. *The organization of behavior : A neuropsychological theory*. Wiley, 1949.
- [58] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the theory of neural computation*. Addison-Wesley, 1991.
- [59] S. D. Van Hooser, J. A. F. Heimel, S. Chung, S. B. Nelson, and L.J. Toth. Orientation selectivity without orientation maps in visual cortex of a highly visual mammal. *J. Neurosci.*, 25(1) :19–28, 2005.
- [60] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8) :2554–2558, 1982.
- [61] C. Horta and W. Erlhagen. Robust persistent activity in neural fields with asymmetric connectivity. *Neurocomputing*, 69 :1141–1145, 2006.
- [62] J. C. Horton and D. L. Adams. The cortical column : a structure without function. *Philosophical Transactions of the Royal Society of London - Series B : Biological Sciences*, 360(1456) :837–862, 2005.
- [63] D. H. Hubel. Single unit activity in striate cortex of unrestrained cats. *The Journal of Physiology*, 147(2) :226–238, 1959.
- [64] D. H. Hubel. *Eye, brain and vision*. W. H. Freeman & Co, 1988.
- [65] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat’s striate cortex. *The Journal of Physiology*, 148(3) :574–591, 1959.
- [66] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction, and functional architecture of cat striate cortex. *Journal of Physiology*, 160(1) :106–154, 1962.
- [67] A. Hutt, M. Bestehorn, and T. Wennekers. Pattern formation in intracortical neuronal fields. *Network : Computation in Neural Systems*, 14 :351–368, 2003.
- [68] F. Jacob. Evolution and tinkering. *Science*, 196(4295) :1161–1166, 1977.
- [69] H. Jaeger. The ”echo state” approach to analysing and training recurrent neural networks. GMD Report 148, GMD - German National Research Institute for Computer Science, 2001.
- [70] J. S. Johnson, J. P. Spencer, and G. Schöner. Moving to higher ground : The dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology*, 26(2) :227 – 251, 2008.
- [71] J. F. Kihlstrom. Letter to the Editor. *Trends Cognitive Sci.*, 6(502), 2002.
- [72] T. Kohonen. *Self-organization and associative memory*, volume 8 of *Springer Series in Information Sciences*. Springer-Verlag, 1989.
- [73] R. A. Kosinski and M. Zaremba. Dynamics of the model of the caenorhabditis elegans neural network. *Acta Physica Polonica B*, 38 :2201–2210, jun 2007.
- [74] I. Lagzi, S. Soh, P.J. Wesson, K.P. Browne, and B.A. Grzybowski. Maze solving by chemotactic droplets. *Journal of the American Chemical Society*, 132(4) :1198–1199, 2010.
- [75] W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states : a new framework for neural computation based on perturbations. *Neural Computation*, 14(11) :2531–2560, 2002.

-
- [76] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.
- [77] V. Markounikau, C. Igel, A. Grinvald, and D. Jancke. A dynamic neural field model of mesoscopic cortical activity captured with voltage-sensitive dye imaging. *PLoS Comput Biol*, 6(9) :e1000919, 09 2010.
- [78] H. Markram. The blue brain project. *Nature Reviews Neuroscience*, 7(2) :153–160, 2006. <http://bluebrain.epfl.ch>.
- [79] O. Ménard and H. Frezza-Buet. Model of multi-modal cortical processing : Coherent learning in self-organizing modules. *Neural Networks*, 18(5-6) :646–655, 2005.
- [80] R. Menzner, A. Steinhage, and W. Erhlagen. Generating interactive robot behavior : A mathematical approach. In *From Animals to Animats 5 : Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 135–144, 2000.
- [81] J. Meyer. *Modelling primary visual cortex dynamics with a dynamic neural field based on voltage sensitive dyes*. PhD thesis, Ruhr-University Bochum, 2007.
- [82] R. C. Miall and D. M. Wolpert. Forward models for physiological motor control. *Neural Networks*, 9(8) :1265–1279, 1996.
- [83] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh. Self-organization, plasticity, and low-level visual phenomena in a laterally connected map model of the primary visual cortex. In R. L. Goldstone, P. G. Schyns, and D. L. Medin, editors, *Perceptual Learning, Psychology of Learning and Motivation*, pages 257–308. 1997.
- [84] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh. *Computational maps in the visual cortex*. Springer, 2005.
- [85] I. Mikhailova and C. Goerick. Conditions of activity bubble uniqueness in dynamic neural fields. *Biological Cybernetics*, 92(2) :82–91, 2005.
- [86] V. B. Mountcastle. Modality and topographic properties of single neurons of cat’s somatic sensory cortex. *J. Neurophysiol.*, 20(4) :408–434, 1957.
- [87] V. B. Mountcastle. The columnar organization of the neocortex. *Brain*, 120 :701–722, 1997.
- [88] C. L. Nehaniv. Self-reproduction in asynchronous cellular automata. In *NASA/DoD Conference on Evolvable Hardware*, pages 201–209. MIT Press, 2002.
- [89] E. Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15 :267–273, 1982.
- [90] G. J. op Akkerhuis. *The operator hierarchy, a chain of closures linking matter, life and artificial intelligence*. PhD thesis, Radboud Universiteit Nijmegen, 2010.
- [91] K. R. Pawelzik, U. Ernst, F. Wolf, and T. Geisel. Orientation contrast sensitivity from long-range interactions in visual cortex. In *Neural Information Processing Systems*. MIT Press, 1996.
- [92] D. J. Pinto and G. B. Ermentrout. Spatially structured activity in synaptically coupled neuronal networks : I. traveling fronts and pulses. *SIAM J. Appl. Math*, 62 :206–225, 2001.
- [93] D. J. Pinto and G. B. Ermentrout. Spatially structured activity in synaptically coupled neuronal networks : Ii. lateral inhibition and standing pulses. *SIAM J. Appl. Math*, 62 :226–243, 2001.

- [94] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, J. O. McNamara, and L. E. White. *Neuroscience*. Sinauer Associates, 4th. edition, 2007.
- [95] S. R. Quartz, T. J. Sejnowski, and H. Hughes. The neural basis of cognitive development : A constructivist manifesto. *Behavioral and Brain Sciences*, 20 :537–596, 1997.
- [96] G. H. Recanzone, M. M. Merzenich, W. M. Jenkins, K. A. Grajski, and H. R. Dinse. Topographic reorganization of the hand representation in cortical area 3b owl monkeys trained in a frequency-discrimination task. *J Neurophysiol*, 67(5) :1031–1056, 1992.
- [97] G. Rizzolatti, L. Fogassi, and V. Gallese. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci*, 2(9) :661–670, 2001.
- [98] F. Robert. *Les systèmes dynamiques discrets*. Springer, 1995.
- [99] N. P. Rougier and J. Vitay. Emergence of attention within a neural population. *Neural Networks*, 5(19) :573–581, 2006.
- [100] A. Roxin, N. Brunel, and D. Hansel. Role of delays in shaping spatiotemporal dynamics of neuronal activity in large networks. *Physical Review Letters*, 94(23) :238103, 2005.
- [101] L. C. Rutherford, A. DeWan, H. M. Lauer, and G. G. Turrigiano. Brain-derived neurotrophic factor mediates the activity-dependent regulation of inhibition in neocortical cultures. *J. Neurosci.*, 17(12) :4527–4535, 1997.
- [102] R. Schmidt and T. Lee. *Motor control and learning : A behavioral emphasis, (fourth edition)*. Human Kinetics, 2005.
- [103] S. Schneider and W. Erlhagen. A neural field model for saccade planning in the superior colliculus : speed-accuracy tradeoff in the double-target paradigm. *Neurocomputing*, 44-46 :623–628, 2002.
- [104] J. Sima. Energy-based computation with symmetric Hopfield nets. *Limitations and Future Trends in Neural Computation*, pages 45–69, 2003.
- [105] V. R. Simmering, A. R. Schutte, and J. P. Spencer. Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research*, 2007.
- [106] J. Sirosh and R. Miikkulainen. Self-organizing feature maps with lateral connections : Modeling ocular dominance. In *Proceedings of the 1993 Connectionist Models Summer School*. Erlbaum, 1994.
- [107] M. L. Stavrinou, S. D. Penna, V. Pizzella, K. Torquati, F. Cianflone, R. Franciotti, A. Bezerianos, G. L. Romani, and P. M. Rossini. Temporal dynamics of plastic changes in human primary somatosensory cortex after finger webbing. *Cerebral Cortex*, 17(9) :2134–2142, 2007.
- [108] G. Stephens, W. Ryu, and W. Bialek. The emergence of stereotyped behaviors in c. elegans. *APS Meeting Abstracts*, pages 10007–+, 2010.
- [109] M. Steriade, D. A. McCormick, and T. J. Sejnowski. Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262(5134) :679–685, 1993.
- [110] S. M. Stringer and I. E. T. de Araujo E. T. Rolls, T. P. Trappenberg. Self-organizing continuous attractor networks and motor function. *Neural Networks*, 16(2) :161–182, 2003.
- [111] T. Sugimoto. A theoretical analysis of formation flight as a nonlinear self-organizing phenomenon. *Journal of Applied Mathematics*, 68 :441–470, 2003.
- [112] R. S. Sutton and A. G. Barto. *Reinforcement learning : an introduction*. The MIT Press, 1998.

-
- [113] J. G. Taylor. Neural networks for consciousness. *Neural Networks*, 10(7) :1207–1225, 1997.
 - [114] J. G. Taylor. Neural ‘bubble’ dynamics in two dimensions : foundations. *Biological Cybernetics*, 80(6) :393–409, 1999.
 - [115] M. Toussaint. Learning a world model and planning with a self-organizing dynamic neural system. In *NIPS*, 2003.
 - [116] D. Y. Ts’o, M. Zarella, and G. Burkitt. Whither the hypercolumn? *The Journal of Physiology*, 587(12) :2791–2805, 2009.
 - [117] A. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1) :230–265, 1937.
 - [118] A. Turing. Computing machinery and intelligence. *Mind*, 59(236) :433–460, 1950.
 - [119] A. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London - Series B*, 237 :37–72, 1952.
 - [120] T. van Gelder. What might cognition be, if not computation? *The Journal of Philosophy*, 92(7) :345–381, 1995.
 - [121] N. A. Venkov. *Dynamics of neural field models*. PhD thesis, University of Nottingham, 2009.
 - [122] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393 :440–442, jun 1998.
 - [123] H. Werner and T. Richter. Circular stationary solutions in two-dimensional neural fields. *Biological Cybernetics*, (85) :211–217, 2001.
 - [124] C. Wilimzig and G. Schöner. The emergence of stimulus-response associations from neural activation fields : dynamic neural theory. pages 2359–2364, 2004.
 - [125] H. R. Wilson and J. D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13 :55–80, 1973.
 - [126] S. P. Wilson, J. L. Law, B. Mitchinson, T. J. Prescott, and J. A. Bednar. Modeling the emergence of whisker direction maps in rat barrel cortex. *PLoS ONE*, 5(1) :e8778, 2010.
 - [127] D. L. Woods, G. C. Stecker, T. Rinne, T. J. Herron, A. D. Cate, E. W. Yund, I. Liao, and X. Kang. Functional maps of human auditory cortex : Effects of acoustic features and attention. *PLoS ONE*, 4(4) :e5183, 04 2009.
 - [128] J. Wyller, P. Blomquist, and G. T. Einevoll. Turing instability and pattern formation in a two-population neuronal network model. *Physica D : Nonlinear Phenomena*, 225(1) :75 – 93, 2007.
 - [129] J. y. Wu, L. Guan, and Y. Tsau. Propagating activation during oscillations and evoked responses in neocortical slices. *J. Neurosci.*, 19(12) :5005–5015, 1999.

Résumé

L'objectif général de cette thèse s'inscrit dans le contexte des sciences cognitives, avec une démarche informatique de modélisation bio-inspirée. Cette démarche vise la conception de modèles de calcul neuronal qui soient capables de participer au développement des solutions logicielles et des systèmes situés autonomes et adaptatifs.

Grâce à son rôle déterminant dans le traitement de l'information au niveau d'une multitude de processus cognitifs, notre source principale d'inspiration dans cette démarche est la structure cérébrale corticale. L'étude des résultats expérimentaux concernant la structure anatomique et le fonctionnement du cortex cérébral nous permettent de mettre en évidence un certain nombre de principes structuraux et fonctionnels généraux que nous employons dans la modélisation des mécanismes et des processus neuronaux. Basé sur ces principes, nous proposons une architecture neuronale d'inspiration corticale, capable de modéliser le processus émergent de la formation des cartes corticales d'auto-organisation.

Afin d'implémenter cette architecture neuronale, nous faisons recours au modèle de champs neuronaux dynamiques, un formalisme mathématique générique conçu pour modéliser la compétition des activités neuronales au niveau cortical mésoscopique. Pour analyser en détail les propriétés et les caractéristiques comportementales des modèles de référence de ce formalisme existant dans la littérature, nous proposons une plate-forme d'évaluation, capable d'examiner et de quantifier le comportement dynamique d'un champ neuronal quelconque dans différents contextes de stimulation. Si cet instrument nous permet de mettre en évidence les avantages pratiques de ces modèles, il nous révèle aussi l'incapacité de ces modèles à conduire l'implantation des processus d'auto-organisation (tel que nous l'avons décrit dans l'architecture neuronale d'inspiration corticale) vers des résultats satisfaisants.

Ces résultats nous amènent à proposer une alternative aux modèles classiques des champs neuronaux dynamiques, basée sur un mécanisme de rétro-inhibition, qui implémente un processus local de régulation neuronale. Grâce à ce mécanisme, le nouveau modèle de champ réussit à implémenter avec succès le processus d'auto-organisation décrit par l'architecture proposée d'inspiration corticale. Une analyse étendue du comportement de ce nouveau modèle, effectuée utilisant l'instrument d'évaluation introduit précédemment, confirme que ce formalisme garde les caractéristiques dynamiques exhibées par les modèles classiques de champs neuronaux, ce qui nous autorise à considérer ce modèle comme une extension des modèles classiques de champs neuronaux dynamiques.

Les travaux décrits dans cette thèse ouvrent à la fois la perspective d'une étude étendue des comportements dynamiques des champs neuronaux, et la possibilité d'implantation des architectures de calcul neuronal de traitement d'information multimodal pour la conception des solutions logicielles ou robotiques bio-inspirées.

Mots-clés: champs neuronaux dynamiques, auto-organisation, modélisation corticale, cartes corticales, systèmes situés.

Abstract

The general goal of this thesis is included in the broader context of cognitive science, and uses a bio-inspired computational approach which aims at designing neuro-computational models that can be used in the development of adaptive and autonomous software solutions and embedded cognitive systems.

Thanks to its essential role in several information processing neural mechanisms, the main biological inspiration for our models comes from the cortical structure. An extensive study of the anatomic and functional properties of the cortical tissue and the cortical activity entitles us to proclaim a number of structural and functional principles that one should employ when designing neural mechanisms and processes. Based on these principles, we propose a cortically inspired neural architecture capable of modeling the emergent process of self-organizing cortical maps.

In order to implement this neural architecture, we focus our interest towards the dynamic neural fields paradigm, a generic mathematical formalism aimed at modeling the competition between the neural activities at a mesoscopic level of the cortical structure. In order to examine in detail the behavioral properties and features of classical models proposed in the last decades, we design an evaluation framework, capable of analysing and quantifying the dynamic behavior of the any neural field, in specific contexts of stimulation. While this instrument highlights the practical advantages of the usage of such models, it also reveals the inability of these models to help implementing the self-organization process (as it is described by the cortically inspired neural architecture) with satisfactory results.

These results lead us to suggest an alternative to the classical neural field models, based on a back-inhibition model which implements a local process of neural activity regulation. Thanks to this mechanism, the new neural field model is capable of achieving successful results in the implementation of the self-organization process described by our cortically inspired neural architecture. Using the neural field evaluation framework previously introduced, we highlight through an extended analysis of the dynamic behavior of the new field model that the new field preserves the dynamic features and properties shown by the classical field models. As a consequence, this result entitles us to consider our field model as an extension of the classical ones.

The research described in this thesis opens the perspectives of both an extensive study of the dynamic behavior of neural fields and the ability of implementing new neuro-computational architectures of multimodal information processing, capable of being used in the development of software solutions or biologically-inspired robot applications.

Keywords: dynamic neural fields, self-organization, cortical model, cortical maps, embedded cognitive systems.