



HAL
open science

Apprentissage de représentations sensori-motrices pour la reconnaissance d'objet en robotique

Nicolas Do Huu

► **To cite this version:**

Nicolas Do Huu. Apprentissage de représentations sensori-motrices pour la reconnaissance d'objet en robotique. Automatique / Robotique. Université Paul Sabatier - Toulouse III, 2007. Français. NNT : . tel-00283073

HAL Id: tel-00283073

<https://theses.hal.science/tel-00283073>

Submitted on 29 May 2008

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

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

Thèse préparée au

Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS

en vue de l'obtention du

Doctorat de l'Université de Toulouse

délivré par l'Université Paul Sabatier

Soutenue le 4 décembre 2007

Spécialité : Intelligence Artificielle

par

Nicolas Do Huu

***APPRENTISSAGE DE REPRÉSENTATIONS
SENSORI-MOTRICES POUR LA RECONNAISSANCE
D'OBJET EN ROBOTIQUE***

Soutenance devant le jury :

Raja Chatila	LAAS, Toulouse	Directeur de thèse
Philippe Gaussier	ETIS, Cergy-Pontoise	Rapporteur
Frédéric Alexandre	LORIA, Vandoeuvre-les-Nancy	Rapporteur
Rachid Alami	LAAS, Toulouse	Examineur
Michel Courdesses	Université Paul Sabatier, Toulouse	Examineur
Olivier Sigaud	ISIR, Paris	Examineur

LAAS-CNRS
7, Avenue du Colonel Roche
31077 Toulouse Cedex 4

Pense à celui qui a planté l'arbre dont tu manges les fruits

REMERCIEMENTS

La rédaction de ce mémoire n'aurait pas été possible sans le concours des personnes que je tiens ici à remercier sincèrement :

Mon directeur de thèse M. Raja Chatila, pour m'avoir prodigué ses conseils, pour avoir su me guider dans les méandres de la recherche et me transmettre sa passion pour la robotique.

Monsieur Philippe Gaussier et Monsieur Frédéric Alexandre pour m'avoir fait l'honneur d'être rapporteur de cette thèse.

Williams Paquier qui m'a aidé plus particulièrement dans toutes mes recherches sur les réseaux de neurones artificiels dont il est un éminent spécialiste.

Mes camarades et amis du LAAS, Léonard, Maxime, Sébastien, Alexandre, Tony, Manu, Fred pour tous les bons moments passés ensemble.

Je remercie également mes parents pour le soutien sans faille qu'ils m'ont toujours apporté. Je suis, à tous, très reconnaissant, sans eux ce mémoire n'aurait pu dépasser le stade de projet. Les indications qu'ils m'ont données m'ont été très précieuses.

Je remercie enfin ma femme, pour son sa patience, son aide et son amour.

TABLE DES MATIÈRES

Remerciements	5
Avant-propos	15
1 Introduction	17
1.1 Considérations générales	17
I. Apprentissage, mémoire et subjectivité	17
II. Représentations distribuées	18
1.2 Présentation générale du modèle	19
I. La boucle sensori-motrice	19
II. Apprentissage des représentations	20
III. Représentations initiales et récompense globale	21
IV. Associations et choix de l'action	22
1.3 Organisation du manuscrit	22
Références	23
2 Représentations et actions	25
2.1 La représentation des objets	26
I. Approche objet-centrée	26
II. Approche vue-centrée	27
III. L'approche écologique	28
IV. Positionnement de notre approche	29
2.2 La représentation de l'espace	29
I. Les représentations spatiales en robotique	29
II. Les représentations spatiales chez les animaux : l'hippocampe	32
III. Les modélisations des cartes cognitives	34
IV. Positionnement de notre approche	36
2.3 La sélection de l'action par renforcement	36

I.	Origines	36
II.	Principes généraux	37
III.	Méthodes d'évaluation des politiques	39
IV.	Positionnement de notre approche	41
	Références	42
3	Apprendre et reconnaître	47
3.1	Problématique	48
I.	Apprentissage en ligne	48
II.	Apprentissage non-supervisé	49
3.2	Neurone et modèles	50
I.	Le neurone biologique	51
II.	Les modèles du neurone, décrire c'est choisir	53
III.	Les modèles biophysique	54
IV.	L'approche fréquentielle	55
V.	L'approche impulsionnelle	57
VI.	L'approche stochastique	57
VII.	Positionnement de notre approche	58
3.3	Neurones, cartes et connectivité	59
I.	Cartes neuronales, invariance à la translation	59
II.	Hierarchie et compétition	60
3.4	Le modèle "intègre et tire"	60
I.	Intégration	60
II.	Tir	62
3.5	Apprentissage et compétition	65
I.	La règle de Hebb	65
II.	La compétition par inhibitions latérales	65
III.	Compétition à l'activation	67
IV.	Loi d'apprentissage	68
V.	Matrices de poids adaptatives	69
3.6	Les 3 voies de la perception	71
3.7	La représentation des objets	72
I.	Représentation par vues, structure du réseau	72
II.	Association temporelle et spatiale de vues, invariance à l'objet	74
3.8	La représentation des positions et des actions	75
I.	Cartes temporisées	75
II.	La représentation égocentrique de l'espace	79
III.	La représentation proprioceptive des actions	80
IV.	L'apprentissage incrémental de motifs de position et d'action	80
3.9	Récompense perçue, récompense induite	81
3.10	Conclusion	83
	Références	83

4	Associer	87
4.1	De la statique à la dynamique	87
4.2	Structure de la mémoire associative	88
	I. Vue générale	88
	II. Cartes référentes et groupes associatifs	89
4.3	Association sensori-motrice de motifs de position : <i>déplacements</i>	89
4.4	Association de vues d'objets	90
4.5	Actions, préconditions et <i>réflexes</i>	95
4.6	Associations et récompense : évaluation des gains	95
4.7	Complexité du réseau, élagage	97
4.8	Conclusion	97
5	Prévoir et Agir	99
5.1	Neurones et connexions pour la prévision	100
	I. Cartes temporisées étendues	100
	II. Stabilité perceptive et diffusion de la prévision	100
	III. Connexions de maintien d'activation	102
	IV. Cohérence de l'activation	104
5.2	Propagation des gains et valeur d'utilité des association <i>réflexes</i>	106
5.3	Propagation de la prévision : représentation perspective	107
5.4	Le réseau associatif résultat de l'apprentissage	109
5.5	Choix de l'action, cohérence de la prévision	110
5.6	Résolution d'une action, validité de la prévision	110
5.7	Aspects temporels	114
5.8	Conclusion	114
6	Conclusion et perspectives	117
	Références bibliographiques	124
	Assemblée : création et monitoring de réseaux	125
	NeuSter : simulation de réseaux de neurones et calculs distribués	127
	Résumé	129

TABLE DES FIGURES

1.1	Composantes objectives et subjectives	18
1.2	Vue globale du système	20
2.1	Décomposition d'objets simples en géons	26
2.2	Vue canonique d'un objet	27
2.3	L'approche vue-centrée pour la reconnaissance d'objets	28
2.4	Grille d'occupation obtenue à l'aide de capteurs à ultra-sons	30
2.5	Carte topologique dans un environnement structuré	31
2.6	Représentation hybride de l'espace	31
2.7	Anatomie de l'hippocampe	32
2.8	Champ de lieu chez le rat	33
2.9	Activations des cellules de vue chez le primate	34
2.10	La boîte de Skinner	36
2.11	Interaction entre agent et environnement dans l'apprentissage par renforcement. Traduit de [Sutton 98].	37
3.1	Evolution du nombre de transistors dans les processeurs Intel	50
3.2	Schéma d'un neurone	52
3.3	Schéma d'une synapse	52
3.4	Evolution au cours du temps du potentiel d'action et des modifications de la perméabilité au sodium et au potassium	53
3.5	Schéma du modèle de Lapicque	54
3.6	Schéma du modèle de Hodgkin et Huxley	55
3.7	Fonctions logiques réalisées par le perceptron	56
3.8	Fonctions d'activation courantes	56
3.9	Schéma d'une carte neuronale	59
3.10	Hierarchie de cartes neuronales	61
3.11	Compétition entre cartes neuronales	62

3.12	Distribution d'échantillonnage des niveaux de décharge d'un neurone	63
3.13	Distribution cumulée des niveaux de décharge	64
3.14	Compétition par inhibitions latérales	66
3.15	Compétition à l'activation	67
3.16	Expansion des matrices de poids	70
3.17	Rétraction des matrices de poids	70
3.18	Plateforme expérimentale	71
3.19	Les 3 voies de la perception	72
3.20	Schéma d'un réseau utilisé pour l'apprentissage de vues d'objet	73
3.21	Schéma des champs de vue d'un objet	75
3.22	Diagramme d'activation des cartes de vues	76
3.23	Matrices de poids des cartes de vues obtenues après apprentissage	76
3.24	Propagation des activations dans une carte de conversion temporisée	77
3.25	Schéma de connexions d'une carte de conversion temporisée vers deux cartes en compétition	77
3.26	Schéma de connexions simplifié d'une carte de conversion temporisée vers deux cartes en compétition	78
3.27	Mesure de disparité et carte de conversion de profondeur	78
3.28	Schéma d'un réseau utilisé pour l'apprentissage des motifs de positions	79
3.29	Schéma des champs de position	80
3.30	Schéma d'un réseau utilisé pour l'apprentissage des motifs d'actions	81
3.31	Schéma du cycle de la récompense	83
4.1	Structure générale de la mémoire associative	88
4.2	<i>Déplacements</i> et champs de lieu	90
4.3	Schéma de connexion d'une association de <i>déplacement</i>	91
4.4	Conversion d'une carte de vue rétinotopique vers une carte de vue temporisée	91
4.5	Filtre à hystérésis	92
4.6	<i>Transformations</i> et champs de vue	93
4.7	Schéma de connexion d'une association de <i>transformation</i>	94
4.8	Schéma de connexion d'une association <i>réflexe</i>	95
4.9	L'apprentissage des valeurs de gain par les associations	96
5.1	Cartes temporisées étendues	100
5.2	Représentation plane des connexions pour la prévision d'un <i>déplacement</i>	101
5.3	Représentation plane des connexions pour la prévision d'une <i>transformation</i>	102
5.4	Représentation plane des connexions pour la prévision d'un <i>réflexe</i>	103
5.5	Connexions de maintien d'activation	103
5.6	Représentation perspective des cartes temporisées étendues	108
5.7	Réseau associatif complet	111
5.8	Propagation de la prévision dans la mémoire associative	112
5.9	Propagation de la prévision dans la mémoire associative (suite)	113

1	Assemblée : création et monitoring de réseaux	126
2	NeuSter : simulation de réseaux de neurones et calculs distribués . . .	128

AVANT-PROPOS

Les travaux qui vont être présentés dans cette thèse constituent une suite directe à l'analyse proposée par Williams Paquier dans sa thèse intitulée "Apprentissage ouvert de représentations et de fonctionnalités en robotique : analyse, modèle et implémentation" [Paquier 04]. Dans ce manuscrit est décrit un modèle d'architecture de contrôle permettant à un robot d'extraire ses propres représentations de l'environnement et d'activer ses effecteurs afin d'agir sur celui-ci et d'atteindre un état subjectif sans cesse amélioré. Nos travaux de thèse entrent dans le cadre de cette architecture tout en proposant une approche plus systématique et détaillée afin de l'appliquer à l'apprentissage ouvert d'objets incluant la reconnaissance visuelle et l'apprentissage sensori-moteur.

INTRODUCTION

1.1	Considérations générales	17
	I. Apprentissage, mémoire et subjectivité	17
	II. Représentations distribuées	18
1.2	Présentation générale du modèle	19
	I. La boucle sensori-motrice	19
	II. Apprentissage des représentations	20
	III. Représentations initiales et récompense globale	21
	IV. Associations et choix de l'action	22
1.3	Organisation du manuscrit	22
	Références	23

1.1 Considérations générales

I. Apprentissage, mémoire et subjectivité

Le monde est structuré et dynamique. Son évolution est régie par des lois que l'intelligence humaine est capable d'identifier et d'exploiter pour atteindre ses objectifs dans un processus d'interaction. Les propriétés physiques de l'environnement ne sont, chez l'homme, pas connues a priori, bien que l'évolution de l'espèce ait permis dans une large mesure l'adaptation au milieu. L'apprentissage phylogénétique¹, peut dans certains cas suffire à certaines espèces évoluant dans des milieux simples et se contentant d'actes réflexes mais à l'échelle qui nous intéresse, celle de l'homme, le caractère imprévisible de l'environnement nécessite des capacités d'adaptation qui, dans le cadre de la robotique, constitueront notre sujet d'étude. L'apprentissage est donc un moyen

¹apprentissage de l'espèce, opposé à l'apprentissage épigénétique ou ontogénétique qui est celui de l'individu nécessitant une expérience personnelle.

d'acquérir des connaissances et permet donc d'adapter le comportement de chaque individu au milieu dans lequel il évolue. La mémoire, dont l'existence est indispensable à toute forme d'apprentissage, nous permet d'une part de stocker l'information pour une utilisation ultérieure mais elle permet d'autre part de prévoir des événements et de créer des relations entre concepts. C'est donc d'une mémoire active dont il s'agit, active par opposition à la mémoire morte d'un ordinateur. D'autre part, la mémoire n'est pas uniquement chez l'homme, un modèle de la dynamique de l'environnement, c'est la mémorisation des conséquences des actions qu'il est capable de produire sur celui-ci ou encore d'une fonctionnalité associée à un but. Or, tandis que l'état du monde tel qu'il apparaît à nos sens est perçu d'une façon comparable d'un individu à l'autre, les buts sont propres à chaque individu. Il existe donc une composante subjective (lié au sujet) évoluant durant l'expérience qui nous permet d'évaluer les situations et d'exprimer la motivation d'agir.

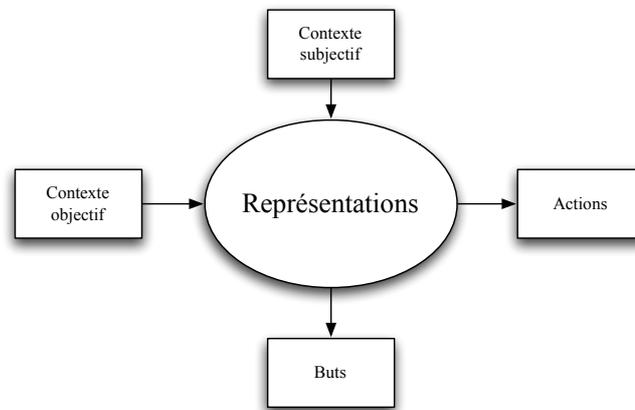


FIG. 1.1 – Le système de représentation est au croisement des composantes objectives et subjectives. Dans le cadre d'un apprentissage ouvert, les représentations construites par le système sont génératrices des actions et des buts.

Nous considérerons dans cette thèse un système qui devra explicitement prendre en compte ces deux composantes objectives et subjectives. On trouvera dans [Paquier 04] une analyse complète et une justification d'un tel système. Nous présenterons ici ses caractéristiques et son fonctionnement. Toutefois, lorsque nécessaire, nous reprendrons l'analyse en certains points.

II. Représentations distribuées

Notre système de représentations est constitué d'un réseau de neurones formels à impulsions. Le modèle de neurone tel qu'il sera décrit dans le Chapitre 3 hérite de diverses caractéristiques issues du domaine des neurosciences computationnelles telles que l'utilisation d'un modèle de neurone de type "intègre et tire" [Lapicque 07, Tuckwell 88], l'emploi d'une règle locale d'apprentissage s'inspirant de la règle de Hebb [Hebb 49] et d'un modèle de propagation de l'information de type "event-driven" per-

mettant d'économiser le temps de calcul. Notre système de représentation est distribué et hiérarchique. À la différence des approches symboliques, les objets du monde ne sont pas représentés par des symboles mais par un ensemble d'unités de traitement actives au même instant. C'est un code, une signature binaire dont le support est l'ensemble des unités de traitement du système. Plusieurs objets peuvent par conséquent partager une partie de leur signature et des relations d'inclusion peuvent exister entre différentes représentations. Ce code binaire peut bien-sûr représenter un nombre, donc en quelque sorte un symbole. Cependant, la capacité expressive du codage distribué permet d'appréhender un nombre illimité de concepts sans en modifier ses règles internes. En effet, dans un système purement symbolique, chaque symbole étant unique et atomique, apprendre un nouveau concept revient à créer un nouveau symbole et explicitement définir ses relations avec les autres. Cette tâche serait donc dévolue au concepteur et non, comme nous le souhaitons, au système lui-même. Nous utilisons un système de codage par position où "l'infinité des sens est ... rejetée dans l'infinité des lieux" [Paquier 04]. Notre système de représentation dispose par ailleurs de liens réunissant les unités de traitement en graphe. C'est grâce à ce graphe qu'il est possible de représenter les liens de dépendance qui relient les représentations entre elles. La représentation d'un objet en un ensemble de composantes est par exemple un moyen efficace de décomposition permettant 1) le partage de composantes entre différentes représentations et 2) une reconnaissance robuste grâce à un niveau d'abstraction accru [Biederman 87].

1.2 Présentation générale du modèle

Cette section est consacrée à une présentation générale du modèle proposé dans [Paquier 04]. Ce bref panorama sera suivi dans les prochains chapitres d'une présentation plus détaillée des diverses composantes fonctionnelles du système ainsi que des contributions qui ont été apportées par nos travaux. La Figure 1.2 représentant une vue simplifiée de la boucle sensori-motrice globale permettra au lecteur d'accéder directement aux chapitres présentant chaque sous-ensemble fonctionnel.

I. La boucle sensori-motrice

Comme le montre la Figure 1.2, nous représentons l'activité globale du système au sein d'une boucle perception-action très classique en robotique. Le robot est en constante interaction avec l'environnement par l'intermédiaire de capteurs et d'effecteurs de natures diverses. Caméras, capteurs à ultra-sons ou même laser sont les outils classiques permettant de recueillir l'information distante tandis que des détecteurs de contact ou des écrans tactiles sont capables de collecter des informations issues d'interactions plus directes. Il est nécessaire d'ajouter à ce bestiaire de capteurs un flux entrant supplémentaire : les capteurs proprioceptifs. Ils permettront d'extraire les informations reflétant la configuration du robot. Ces capteurs particuliers procèdent d'une mise à disposition de valeurs (variables) reflétant la configuration des différents degrés de liberté du système robotique. Les effecteurs ou actionneurs participent à la modi-

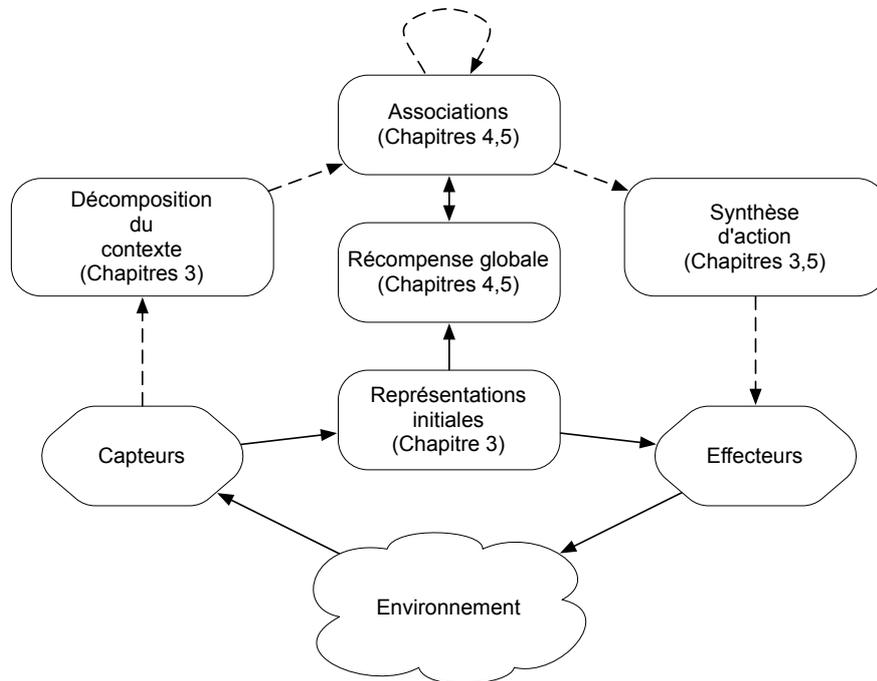


FIG. 1.2 – Vue globale du système sous la forme d’une boucle sensori-motrice. Les associations sensori-motrices jouent un rôle central : elles sont au carrefour des représentations activées par les capteurs (représentations objectives) et des signaux de récompenses (représentations subjectives) afin d’offrir une prise de décision pour la synthèse d’actions.

fication de la configuration du robot dans son environnement et constituent le moyen de mettre en œuvre les actions disponibles du robot. Les mécanismes d’apprentissage sont représentés dans la Figure 1.2 par des flèches en pointillés.

II. Apprentissage des représentations

Dans une première phase, le robot doit être capable d’apprendre des représentations pertinentes de son environnement ou de ses propres configurations par l’exploitation des données de ses capteurs. Ce mécanisme appelé la *décomposition du contexte* est basé sur la recherche d’invariants perceptifs spatiaux ou temporels. D’une façon symétrique, le robot doit apprendre les représentations des actions qu’il est capable de produire, ces données n’étant pas disponibles a priori. La décomposition du contexte et la synthèse d’action sont mis en œuvre par le biais de réseaux de neurones à impulsions dont l’apprentissage est non-supervisé grâce notamment à une compétition entre neurones et à un apprentissage hebbien. Une large part de nos travaux de recherche a été dédiée à l’amélioration du modèle de neurones afin de produire des représentations plus stables à partir de données visuelles. Nous avons par exemple proposé une méthode dynamique de redimensionnement des matrices de poids synaptiques dont les

dimensions était jusqu'à présent figées. Cette méthode a montré son efficacité pour l'apprentissage de vue d'objets en temps réel et de façon non-supervisée. La deuxième étape de ce travail fut de permettre au système d'apprentissage d'associer différentes vues d'un même objet par l'utilisation d'une mémoire à court terme mettant à profit la cohérence spatio-temporelle reliant ces différentes vues. Ces différents apports se sont accompagnés d'une révision profonde du modèle de neurone : le passage à un modèle stochastique et à une fonction de transfert adaptative a permis d'une part la diminution du temps nécessaire à l'apprentissage et d'autre part une plus grande robustesse de la reconnaissance.

L'apprentissage des représentations de positions et d'actions nous a permis d'introduire un nouveau type d'organisation neuronale sous la forme de cartes neuronales temporisées. Ces cartes ont pour but de conserver en mémoire l'état d'activation d'un ensemble de neurones et de permettre ainsi l'apprentissage de séquences d'activations. Les représentations de positions et d'action n'étant pas connues a priori, elles sont apprises par l'observation des données proprioceptives sous la forme de cartes temporisées et séparées en deux canaux distincts : la position des caméras (pan et tilt) ainsi que la distance de l'objet obtenue par une mesure de stéréo-disparité définissent les entrées d'apprentissage des cartes de positions tandis que la vitesse de chaque moteur du robot constituent les entrées pour d'apprentissage des représentations d'actions.

La dissociation de ces trois types de représentations : représentations visuelles (Quoi), représentations de positions (Où) et représentations d'actions (Comment) ainsi que l'introduction explicite de la notion de temps par les cartes temporisées sont à l'origine d'une simplification importante des mécanismes associatifs menant à la prise de décision. Cette dissociation a en effet permis d'éviter de recourir à une interconnexion complète et systématique de l'ensemble des représentations apprises telles que proposé dans le modèle originel de mémoire associative.

III. Représentations initiales et récompense globale

Au centre de la Figure 1.2 se situe la composante subjective du système. Des représentations initiales permettent d'associer, avant tout apprentissage, des récompenses positives ou négatives (punitions) aux données capteurs. Le robot peut par exemple recevoir une punition lors de l'activation d'un capteur de contact ou une récompense si ses capteurs proprioceptifs mesurent le rechargement de sa batterie. Ces différentes données sont par la suite utilisées afin de générer une valeur globale de récompense symbolisant l'état subjectif courant du système. Les représentations apprises lors de la phase de décomposition du contexte sont dans un second temps associées au niveau de récompense globale perçu au moment de leur détection. Toute représentation apprise par le système sera donc, à terme, évaluée par un niveau de récompense reflétant la valeur de récompense globale mesurée en moyenne lors de son activation. C'est par ce biais que le système est capable d'accroître l'ensemble des représentations subjectives à partir de l'ensemble initial donné lors de la conception. A la différence d'un apprentissage par renforcement classique, notre système est donc en mesure d'évaluer le contexte courant sans la présence explicite de signaux de récompense offerts par l'environnement.

IV. Associations et choix de l'action

L'association d'une valeur de récompense à une représentation apprise qu'elle soit visuelle, de position ou d'action ne suffit pas à la prise de décision. Il n'existe en effet à un instant donné qu'un sous-ensemble d'états atteignables et d'actions envisageables. De plus, étant donné le contexte courant, atteindre un état à fort niveau de récompense peut mener en des états intermédiaires subjectivement moins intéressants, pouvant même entraîner un solde de récompense négatif pour la série d'actions envisagée. C'est pourquoi, nous proposons d'apprendre grâce aux cartes neuronales temporisées citées plus haut, les ensembles de transitions d'états (pour les représentations visuelles et de positions) ainsi que les préconditions d'actions indispensables à la prise de décision et à la création de l'horizon des états atteignables. A l'image des représentations, ces associations sont donc de trois types :

- les transformations : qui à une vue, une position et une action, associent une nouvelle vue,
- les déplacements : qui à une position et une action associent une nouvelle action,
- les réflexes : qui à une vue et une position associent une action.

Ainsi, grâce à ces trois types d'associations obtenue incrémentalement par apprentissage et par l'interconnexion des cartes neuronales temporisées les constituant nous proposons de construire un réseaux associatif directement exploitable pour la prise de décision. Nous montrerons en effet comment l'emploi systématique de schémas de connexions simples permet de construire un réseau associatif au sein duquel la diffusion passive des informations capteurs permet la mise à jour des valeurs d'utilité exploitées par l'apprentissage par renforcement. Nous souhaitons ainsi favoriser des algorithmes de calcul locaux afin d'exploiter au maximum le parallélisme des calculs et des flux d'informations.

1.3 Organisation du manuscrit

Suite à cette discussion très générale qui nous a permis d'introduire les éléments clés de notre approche nous proposons dans le prochain chapitre un panorama des principes et méthodes utilisés couramment dans le domaine de la robotique suivant trois axes distincts. Tout d'abord nous présenterons trois types d'approches qui ont encadré le domaine de la reconnaissance d'objet : l'approche vue-centrée, l'approche objet-centrée et l'approche écologique. Dans un second temps nous examinerons les techniques de représentations spatiales du point de vue de la robotique, de la biologie ainsi que les modèles biologiquement inspirés. La troisième et dernière partie de ce chapitre sera consacrée aux méthodes de sélection d'action par renforcement dont notre modèle est fortement inspiré. Cet état de l'art nous permettra de positionner notre approche avant de poursuivre par une présentation détaillée de nos travaux.

Le Chapitre 3 aura pour objet la présentation du modèle de neurone qui constitue le coeur des capacités d'apprentissage de notre système. Nous présenterons également dans ce chapitre les trois types de représentations mentionnées plus haut : les représentations visuelles d'objet, les représentations de positions et les représentations d'actions.

L'acquisition de représentations permettant la décomposition du contexte perceptif n'a de sens qu'à l'issue de la création d'un réseau associatif exploitable. Nous montrerons dans le chapitre 4 comment notre système est en mesure d'acquérir un ensemble d'associations sensori-motrices en s'appuyant sur les représentations de vues, de position et d'actions apprises en parallèle. Nous préciserons également le mécanisme d'apprentissage subjectif reliant la valeur de récompense globale aux éléments de la mémoire associative.

Le Chapitre 5 plus richement illustré, permettra de présenter la propagation des activations au sein des associations et plus globalement, la construction de la prévision pour la décision. Nous introduirons notamment la notion de carte temporisée étendue qui constitue le support neuronal de la prévision. Chaque type d'association (transformation, déplacement et réflexe) sera associé à un schéma de connexion particulier pouvant être reproduit de façon incrémentale et systématique en garantissant la cohérence de l'ensemble des états prédits.

Le Chapitre 6 sera dédié aux conclusions et perspectives.

Références

- [Biederman 87] I. Biederman. *Recognition-by-components : a theory of human image understanding*. Psychological Review, vol. 94, pages 115–147, 1987.
- [Hebb 49] Donald O. Hebb. *The organization of behavior : A neuropsychological theory*. Wiley, New York, June 1949.
- [Lapicque 07] L. Lapicque. *Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation*. J. Physiol. Pathol. Gen, no. 9, pages 620–635, 1907.
- [Paquier 04] W. Paquier. *Apprentissage ouvert de représentations et de fonctionnalités en robotique : analyse, modèles et implémentation*. Thèse de doctorat, Université Paul Sabatier de Toulouse, 2004.
- [Tuckwell 88] H. Tuckwell. *Introduction to theoretical neurobiology*, volume 1, 2. Cambridge University Press, 1988.

REPRÉSENTATIONS ET ACTIONS

2.1	La représentation des objets	26
I.	Approche objet-centrée	26
II.	Approche vue-centrée	27
III.	L'approche écologique	28
IV.	Positionnement de notre approche	29
2.2	La représentation de l'espace	29
I.	Les représentations spatiales en robotique	29
A.	Représentations métriques	29
B.	Représentations topologiques	30
C.	Représentations hybrides	31
II.	Les représentations spatiales chez les animaux : l'hippocampe	32
A.	Anatomie	32
B.	L'hippocampe du rat, les cellules de lieu	33
C.	L'hippocampe des primates, les cellules de vue	33
III.	Les modélisations des cartes cognitives	34
A.	Modèle simulés	34
B.	Modèles intégrés	35
IV.	Positionnement de notre approche	36
2.3	La sélection de l'action par renforcement	36
I.	Origines	36
II.	Principes généraux	37
III.	Méthodes d'évaluation des politiques	39
A.	La Programmation Dynamique	39
B.	Méthode de Monte-Carlo	41
C.	Méthode des différences temporelles	41
IV.	Positionnement de notre approche	41
	Références	42

Ce chapitre qui est essentiellement bibliographique, a pour objectif de positionner notre approche selon trois points de vue distincts : la représentation des objets, la représentation de l'espace et la sélection de l'action. Pour chacun de ces points, nous mettons à la disposition du lecteur une brève présentation de différentes techniques existant dans la littérature pour offrir un éclairage suffisant lors de la description détaillée de notre système qui aura lieu dans les prochains chapitres.

2.1 La représentation des objets

La question du type de représentation utilisé par notre système visuel pour la reconnaissance des objets a constitué une large polémique au sein de la communauté des sciences cognitives. Deux approches se sont frontalement opposées. Nous proposons au lecteur un bref rappel des points clés de ce débat afin de pouvoir situer plus précisément notre approche.

I. Approche objet-centrée

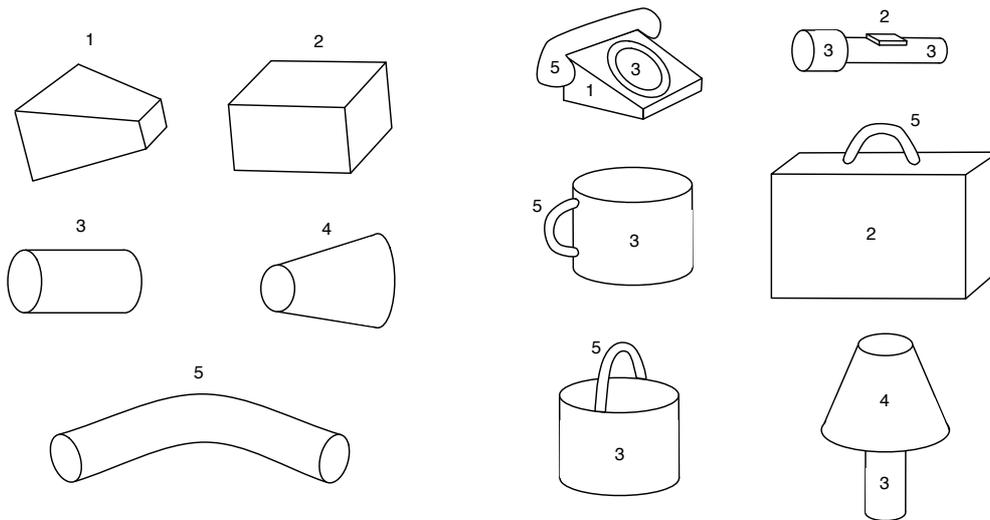


FIG. 2.1 – Décomposition d'objets simples en géons (Inspiré de [Biederman 87]) Selon l'approche par composantes, les objets (à droite) peuvent être décomposés en un ensemble de primitives géométriques simples : les *géons* (à gauche).

Tout d'abord l'approche objet-centrée, soutenue par Irving Biederman suppose l'existence de primitives géométriques (ou géons) permettant aux représentations de décrire la structure tri-dimensionnelle des objets [Biederman 87, Biederman 93, Biederman 00]. Ces représentations nécessitent donc d'extraire des images les primitives géométriques ainsi que leurs configurations dans l'espace (Figure 2.1). La Reconnaissance par Composante (en anglais RBC : Recognition-by-components) est soutenue par

des expériences montrant la capacité de notre système visuel à reconnaître les objets quelle que soit leur orientation dans l'espace (reconnaissance vue-invariante). Ce type de représentations, très utile en robotique pour la planification de mouvement et la manipulation d'objets, demeure complexe à mettre en œuvre pour un apprentissage en temps réel basé sur les seules données visuelles.

II. Approche vue-centrée

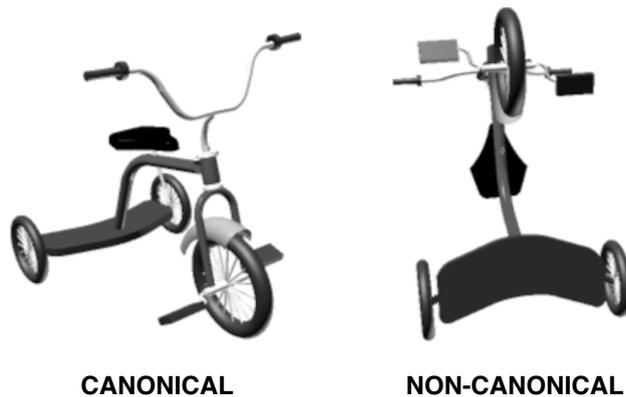


FIG. 2.2 – Des expériences de reconnaissance d'objet menées chez l'homme mettent en évidence des temps de reconnaissance différents selon la vue de l'objet présentée au sujet. La vue produisant un temps de reconnaissance le plus faible est qualifiée de vue canonique [Bülthoff 94].

Un deuxième type de représentations défendu par Heinrich H. Bulthoff et Michael J. Tarr notamment, inspiré des techniques de reconnaissance de formes, propose d'encoder les objets tels qu'ils apparaissent selon différents points de vues (approche vue-centrée). Cette approche s'appuie également sur des données expérimentales montrant l'influence du point de vue sur le temps de reconnaissance et de discrimination des objets [Bülthoff 94, Tarr 95] et introduit la notion de vue canonique produisant la reconnaissance la plus rapide (Figure 2.2). Cette influence du point de vue sur le temps de reconnaissance a été un argument d'importance pour les défenseurs de l'approche vue-centrée. Cette approche doit néanmoins faire face à de nombreuses limitations telles que la quantité importante de mémoire nécessaire à l'apprentissage des vues, la sensibilité de la reconnaissance ou le problème de l'apprentissage de classes d'objets par généralisation. Nous renvoyons le lecteur à [Tarr 98] pour une discussion plus approfondie. Bien que l'approche vue-centrée soit désormais privilégiée pour la modélisation de la reconnaissance des objets chez l'homme, il est aujourd'hui également admis qu'un tel type de représentations ne saurait se priver d'une composante vue-invariante de l'objet. Le travaux de M. Riesenhuber, T. Poggio et G. Wallis ont tenté de réconcilier ces deux approches tout en proposant de réduire le nombre de vues nécessaires à l'apprentissage par l'utilisation de réseaux de fonction à base radiale (Radial Basis Function Networks) [Poggio 90, Riesenhuber 99, Riesenhuber 00].

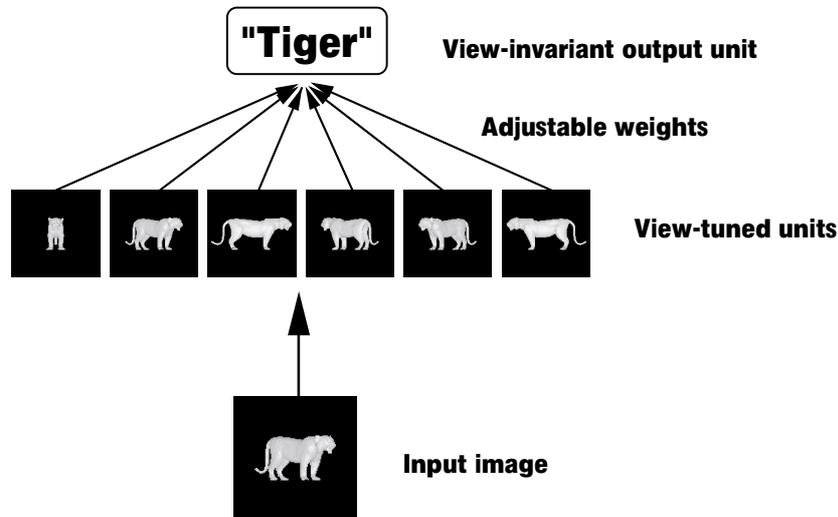


FIG. 2.3 – L’approche vue-centrée proposée par M. Riesenhuber et T. Poggio [Riesenhuber 00] propose d’utiliser un ensemble de détecteurs de vue associés à un module de sortie dont l’activité est invariante à la vue présentée en entrée. Le signal de détection sortant est binaire. Celui-ci vaut 1 si la somme pondérée de l’activité des détecteurs de vue dépasse un seuil prédéterminé. La valeur des poids associés aux connexions est ajustée durant l’apprentissage pour obtenir un taux de reconnaissance satisfaisant.

III. L’approche écologique

Les approches présentées précédemment supposent une séparation stricte entre la perception et l’action. La représentation des objets est en effet construite dans le but de fournir une reconnaissance stable (invariante à l’objet) afin de mettre en œuvre par la suite un mécanisme décisionnel symbolique. Cette approche cognitive qui semble très "naturelle" du point de vue du roboticien n’est pas pour autant universellement admise. James J. Gibson, psychologue américain, qui a joué un rôle majeur dans le domaine de la perception visuelle, a proposé une approche radicalement opposée, qualifiée d’*approche écologique* [Gibson 79]. Il suggère de replacer l’animal au centre de son environnement et de considérer la perception du point de vue de leurs interactions. Cette théorie s’articule selon deux concepts principaux : le champ optique ambiant et les "affordances". Le champ optique ambiant, à ne pas confondre avec le champ de vision est constitué de l’ensemble des rayons lumineux qui parviennent à la rétine. Pour Gibson, c’est la structure du champ optique ambiant qui fournit l’information perceptive, par la détection d’invariants et non la simple lumière. La notion d’affordance, néologisme issu de l’anglais *to afford* (permettre, donner la possibilité de), traduit les interactions possibles entre l’animal et son environnement. Pour Gibson, les affordances ne sont pas des abstractions d’un processus mental de haut niveau mais sont au contraire perçues directement. Ces affordances existent d’ailleurs d’elles-mêmes qu’elles soient perçues ou non. L’état, les possibilités physiques de l’animal vis-à-vis de l’environnement les

définissent de fait, d'où le nom d'*approche écologique*.

IV. Positionnement de notre approche

Notre choix s'est rapidement porté sur une représentation d'objets formée par un ensemble de détecteurs de vues (vue-centrée). Ce type de représentation est en effet bien adapté à l'apprentissage par réseau de neurones et l'utilisation d'une décomposition hiérarchique de l'image d'entrée par des détecteurs locaux de contrastes et d'orientations. Nous montrerons cependant qu'un détecteur invariant au changement de point de vue (objet-centrée) a également été introduit pour faciliter la construction incrémentale de la collection de détecteurs de vues de chaque objet. Notre approche est donc très similaire aux travaux de M. Riesenhuber et T. Poggio [Riesenhuber 00] dont l'architecture est illustrée par la figure 2.3. De plus, nous montrerons comment l'apprentissage d'associations permet d'étendre la représentation par vues en reliant des vues proches et les actions permettant de transiter entre ces vues.

2.2 La représentation de l'espace

I. Les représentations spatiales en robotique

Dans le cadre de la navigation robotique autonome, de nombreuses méthodes ont été proposées pour représenter l'espace de manière efficace et sous une forme la plus compacte possible. Des représentations métriques basées sur les données brutes des capteurs, aux représentations topologiques mettant en jeu un symbolisme spatial, ces approches sont le résultat d'un intérêt toujours grandissant pour la localisation et la navigation autonome ou SLAM (Simultaneous Localization and Mapping). Sebastian Thrun, qui est un actif contributeur du domaine, énonce dans [Thrun 98] une liste de facteurs limitant la capacité du robot à apprendre un modèle de l'environnement :

- **Les capteurs** ne mesurent pas directement l'information utile à la localisation.
- **Les perceptions sont limitées** autour du robot, celui-ci doit explorer son environnement.
- **Les capteurs sont bruités** et la distribution de ce bruit est souvent inconnue.
- **Le glissement du robot** sur le sol rend les mesures odométriques génératrices d'erreurs importantes.
- **L'environnement est complexe et dynamique**, il est donc impossible de maintenir un modèle exact et difficile de prédire de façon efficace.
- **Les contraintes "temps-réel"** nécessitent un modèle simple et facilement accessible.

Nous proposons ici un bref rappel des différentes approches et les représentations spatiales sous-jacentes.

A. REPRÉSENTATIONS MÉTRIQUES

Les représentations métriques sont basées sur une représentation de l'espace sous la forme d'une grille d'occupation [Moravec 88, Elfes 89, Elfes 87, Moravec 85], de

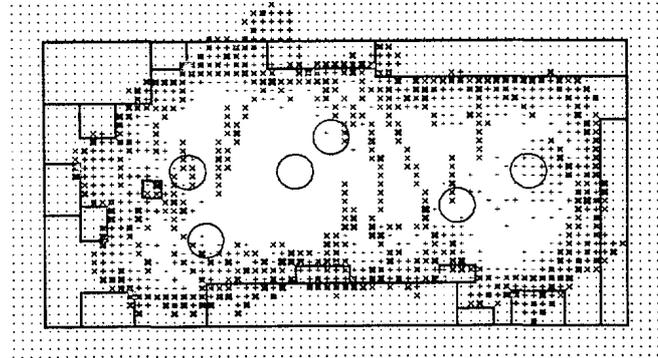


FIG. 2.4 – Grille d’occupation obtenu à l’aide de capteur à ultra-sons [Moravec 85]. Les positions du robot lors des mesures sont représentées par des cercles et les obstacles et murs de l’environnement par des traits pleins. Chaque case de la grille peut recevoir les symboles suivant : (vide) forte certitude d’espace libre, (+) incertitude sur la nature de la zone, (x) zone occupée, (.) zone inconnue.

points, de segments ou de plans extraits des données des capteurs (laser, vision...). Dans le cas des grilles d’occupation, chaque cellule de la grille représente la présence (ou la probabilité de présence) d’un obstacle (Figure 2.4). Ce type de représentation s’avère très facile à mettre en œuvre notamment par l’utilisation de capteurs laser ou infrarouge et rend la localisation aisée car elle dépend uniquement de la géométrie de l’environnement immédiat du robot. Le principal inconvénient de cette méthode repose sur la taille de la grille qui, pour obtenir un niveau de détail suffisant, peut atteindre une taille très grande. La taille de cette grille, qui est indépendante de la complexité de l’environnement, a un impact direct sur le temps nécessaire à la planification. Les représentations métriques peuvent également être basées sur l’extraction de primitives géométriques telles que des droites, segments ou des plans.

B. REPRÉSENTATIONS TOPOLOGIQUES

Les approches topologiques, représentent l’environnement sous la forme d’un graphe dont les cellules correspondent à des lieux et les arcs représentent l’existence de chemins entre ces lieux [Kuipers 78, Kuipers 91, Kortenkamp 94, Yamauchi 96]. Les lieux sont identifiés par le biais d’amers perceptifs, chaque cellule de lieu est donc associée à un modèle permettant la localisation du robot. Il se pose alors le problème de la discrimination des lieux ayant des caractéristiques perceptives proches pouvant être faussement identifiées : l’*aliasing perceptif*. La représentation graphique est cependant bien plus compacte que les grilles d’occupation et sa complexité varie avec celle de l’environnement. Les représentations topologiques sont directement exploitables par les systèmes symboliques de planification et plus accessibles pour l’utilisateur humain.

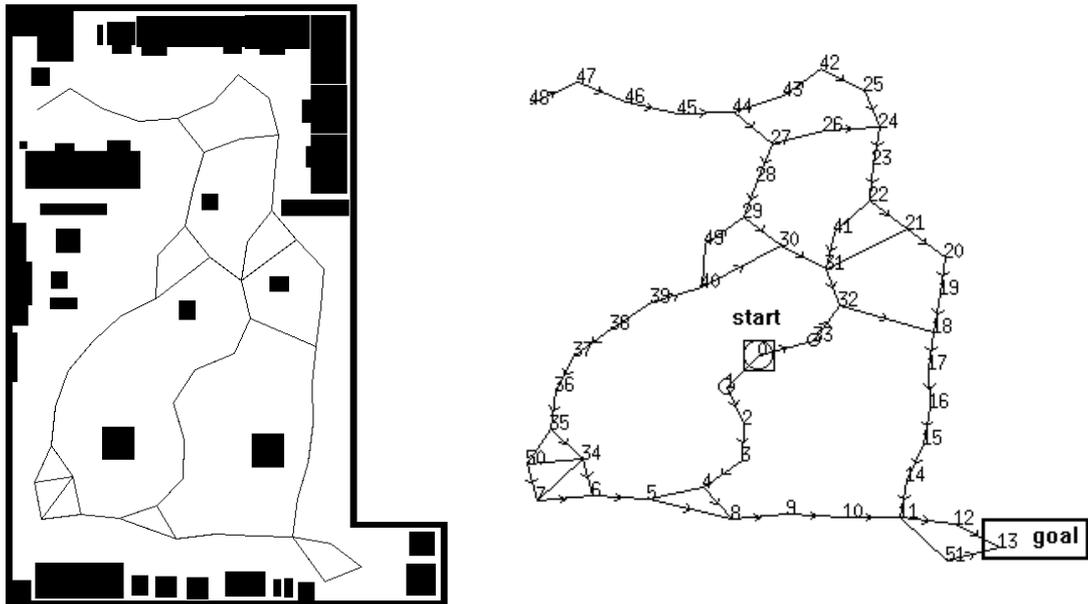


FIG. 2.5 – Carte topologique dans un environnement structuré [Yamauchi 96]. Ce type de carte est particulièrement utile à la navigation cependant, la simplification du graphe par la sélection de noeuds n'est pas une tâche triviale à accomplir.

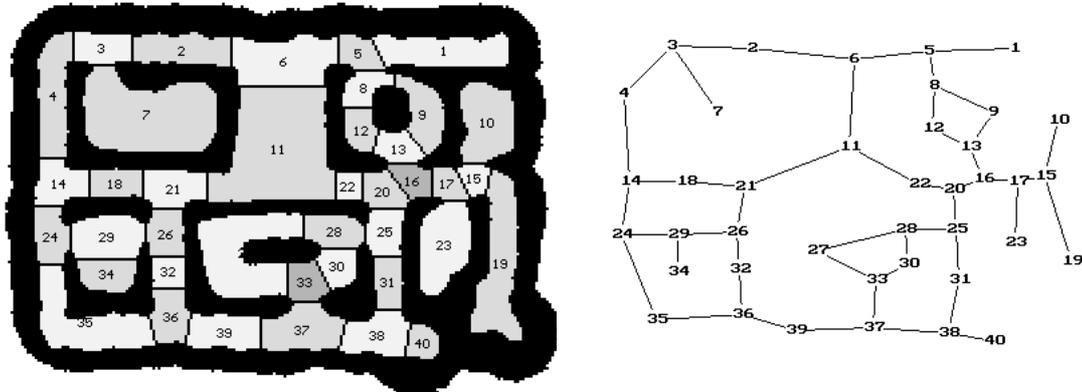


FIG. 2.6 – Représentation hybride de l'espace [Thrun 98]. Un résultat majeur du SLAM accompli grâce à l'utilisation simultanée des représentations métriques et topologiques.

C. REPRÉSENTATIONS HYBRIDES

Les représentations métriques et topologiques de l'espace ont des caractéristiques très différentes. Leurs avantages et inconvénients respectifs, qu'ils soient liés au temps d'apprentissage, à la facilité d'exploitation ou à la compacité des représentations, les rendent très complémentaires lorsque elles sont réunies en des représentations hybrides de l'espace [Chatila 85, Thrun 98, Arleo 99]. Un des problèmes majeurs du SLAM

consistant à contrecarrer les erreurs odométriques lors de parcours dans des environnements circulaires a été résolu par ce type de méthodes (Figure 2.6).

II. Les représentations spatiales chez les animaux : l'hippocampe

A. ANATOMIE

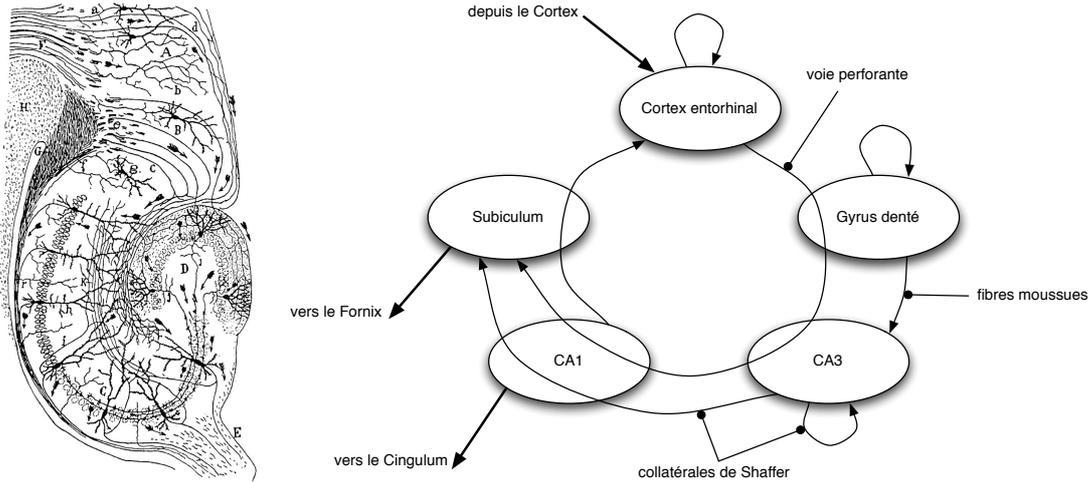


FIG. 2.7 – Anatomie de l'hippocampe [y Cajal 11]. L'information y est transmise au sein d'une boucle unidirectionnelle.

Il est communément admis que l'hippocampe joue un rôle majeur dans la consolidation de la mémoire, la mémoire spatiale et la navigation. L'hippocampe est une structure du cerveau bilatéralisée et symétrique qui fait partie du système limbique. C'est une formation corticale appartenant à l'archicortex : l'hippocampe est une structure ancienne qui se retrouve chez les mammifères mais aussi chez les espèces primitives telles que les cyclostomes (lamproie). La stabilité des structures neuroanatomiques de l'hippocampe parmi les différentes espèces suggère que cette structure a un rôle majeur dans le fonctionnement du système nerveux central. La formation hippocampique est divisée en trois parties : le gyrus denté, la corne d'Ammon (à cause de sa ressemblance avec la forme des cornes de bélier du dieu Ammon de la mythologie) elle-même subdivisée en 3 sous-parties (CA1, CA2, CA3) et le subiculum. L'hippocampe est caractérisé par un circuit neuronal tri-synaptique (Figure 2.7). L'entrée de l'information dans cette boucle unidirectionnelle se fait par les axones du cortex entorhinal nommés fibres perforantes qui forment les premières connexions avec le gyrus denté. Les cellules granulaires du gyrus denté projettent massivement leurs axones (fibres moussues) vers les cellules pyramidales de CA3 (deuxième connexion). Ces cellules se projettent alors vers les dendrites des cellules pyramidales de CA1, formant la troisième partie de la boucle : les collatérales de Schaffer. Celles-ci projettent à leur tour leurs axones vers le subiculum et produisent une voie sortante vers le cingulum. Le subiculum constitue l'étape finale du circuit hippocampique et produit une voie sortante vers le Fornix.

B. L'HIPPOCAMPE DU RAT, LES CELLULES DE LIEU

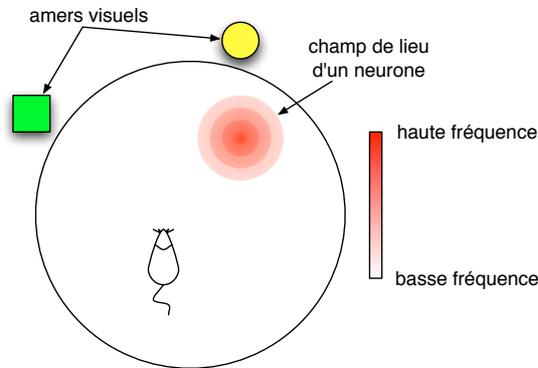


FIG. 2.8 – Champ de lieu chez le rat : l'activation des cellules de lieu chez le rat est étroitement liée à sa position par rapport à celles d'indices visuels présents lors d'une phase d'apprentissage. L'enregistrement des fréquences d'activation de cellules prises individuellement met en évidence des zones de l'environnement pour lesquelles ces cellules sont actives, ces zones définissent les champs de lieu des cellules.

L'identification des cellules de lieu dans l'hippocampe du rat [O'Keefe 71] a eu une incidence fondamentale dans la compréhension de la formation des représentations de l'espace chez les animaux et chez l'homme. Les cellules de lieu sont des neurones dont l'activation est en étroite corrélation avec la position de l'animal dans son environnement, chaque cellule de lieu ne s'active que lorsque l'animal est situé dans une zone spécifique de l'environnement [O'Keefe 96]. Ce type de représentations est allocentrique : elle est relative à l'environnement et non à l'animal. Cette découverte est à l'origine de l'hypothèse selon laquelle le rôle de l'hippocampe serait de constituer une carte cognitive de l'environnement [O'Keefe 78]. La position des champs de lieu, c'est à dire la zone produisant l'activation d'une cellule particulière n'est néanmoins pas absolue mais relative à la détection d'indices visuels. Les cellules de lieu peuvent aussi montrer une activation dépendante de l'orientation de l'animal dans certains environnements ou lorsque la direction est un facteur générant la récompense lors de l'apprentissage. Bien que la vision ait une influence majeure sur la localisation, d'autres modalités telles que l'olfaction ou la proprioception ont montré leur influence sur l'apprentissage ou l'activation des champs de lieu. L'intégration de chemin à l'aide de données proprioceptives permettrait en outre la discrimination de lieux visuellement semblables et la navigation dans l'obscurité.

C. L'HIPPOCAMPE DES PRIMATES, LES CELLULES DE VUE

Les expériences qui ont permis la découverte des cellules de lieu chez le rat ont naturellement encouragé la recherche de telles cellules chez les espèces plus proches de l'homme, les primates. A la différence du rat, les expériences menées sur les primates ont montré l'existence de cellules dépendantes de la zone de l'espace observée par l'animal et non celle où il se situe [Rolls 95] : les cellules de vue. Cette différence notable est

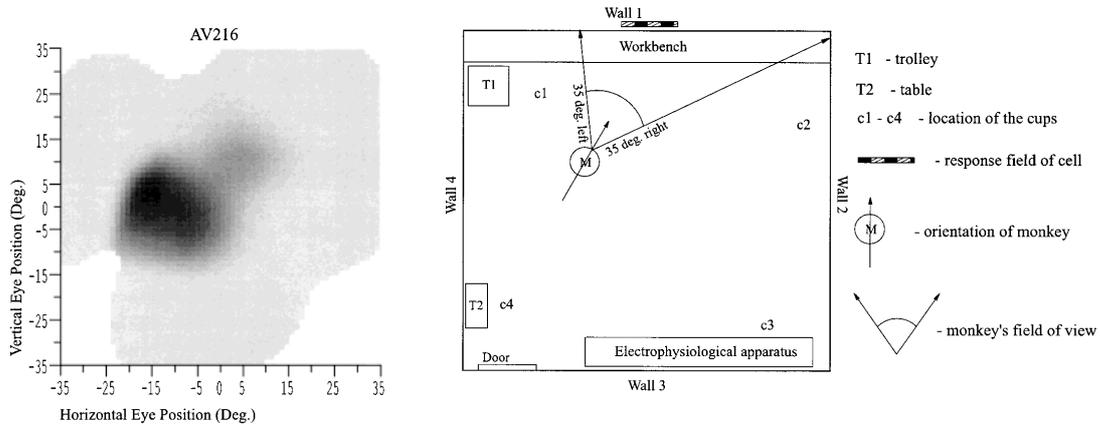


FIG. 2.9 – Activations des cellules de vue chez le primate [Rolls 99]. L'enregistrement des fréquences d'activation de certaines cellules de l'hippocampe montre l'existence de cellules dont l'activation est associée à la position dans un repère global de l'objet observé, indépendamment de la position et de l'orientation du sujet.

illustrée par la Figure 2.9. A l'image des cellules de lieu, les cellules de vue forment une présentation allocentrique de l'espace dans la mesure où ni la position des yeux ni l'orientation de l'animal n'ont d'influence sur leur activation [Georges-François 99]. De récentes publications montrent en outre l'existence dans l'hippocampe de neurones s'activant lors de la reconnaissance d'objets indépendamment de leurs positions et de neurones s'activant lors de la combinaison d'un objet et d'une position particulière [Rolls 04, Rolls 05, Rolls 06]. Ces constatations offrent une meilleure compréhension du rôle de l'hippocampe dans la formation de la mémoire spatiale.

III. Les modélisations des cartes cognitives

Depuis la mise en évidence des cellules de lieu de l'hippocampe, différents modèles ont été proposés afin d'en reproduire les mécanismes et vérifier les hypothèses concernant l'apprentissage et l'exploitation des représentations spatiales pour des tâches de localisation et de navigation. Nous allons présenter différents modèles en distinguant les modèles simulés, utilisant un système visuel abstrait, des modèles intégrés à des systèmes robotiques, ces derniers étant bien sûr d'un intérêt particulier pour nos travaux. La liste des modèles présentés n'est pas exhaustive, leur présentation ayant pour but de définir quelques principes clés dans ces domaines.

A. MODÈLE SIMULÉS

Parmi ces modèles citons tout d'abord celui proposé par Sharp [Sharp 91] où une architecture neuronale à trois étages permet l'apprentissage de cellules de lieu par l'exploitation d'indices visuels, leur distance à l'agent ainsi que l'orientation de l'agent vis-à-vis de l'indice observé. L'apprentissage des neurones est de type hebbien (voir Chapitre 3) et des connexions latérales inhibitrices supportent un mécanisme de type

"Winner-take-All". Le système visuel est ici abstrait, les expériences étant menées en simulation.

Le modèle proposé par Wan, Redishand et Touretzky [Wan 94, Touretzky 96, Redish 97a, Redish 97b] est composé de quatre ensembles fonctionnels : les représentations de vues locales, les représentations des directions de la tête, l'intégrateur de chemin et les représentations de lieux. Le système visuel fournit les informations de détection d'indices visuels, leurs distances à l'agent et l'angle selon lequel ils sont détectés. Grâce à ces informations et en accord avec la direction courante de la tête, une valeur globale d'orientation de l'agent peut être calculée. Cette valeur d'orientation est maintenue cohérente avec la dynamique de l'agent, sa vitesse angulaire fournit par le système vestibulaire et son action courante. L'intégrateur de chemin a pour but de maintenir une estimée de position en sommant les actions que produit l'agent. Cette sommation produisant inévitablement une dérive, le quatrième ensemble est chargé d'extraire une représentation stable de position à l'aide de l'intégrateur et des vues locales. Dans cet ensemble le recrutement de nouvelles cellules de lieu est incrémental. Il n'existe pas d'implantation neuronale de ce modèle.

B. MODÈLES INTÉGRÉS

Le modèle proposé par Gaussier [Leprêtre 00, Gaussier 02, Cuperlier 06] s'appuie sur la détection d'indices visuels au sein d'images panoramiques recomposées. La représentation de l'espace est basée sur l'apprentissage de transitions entre différentes positions du robot. La reconnaissance de la position locale du robot est basée sur la configuration des amers visuels. Les cellules de lieu sont apprises par un mécanisme de compétition de type Winner-Take-All, et leur activité lors du déplacement du robot définit un champ de lieu pour chaque cellule. Le recrutement d'une nouvelle cellule de lieu est amorcé lorsque l'activité des cellules de lieu courantes ne dépasse pas un seuil de reconnaissance. Des cellules de transition permettent d'associer l'action générant les activités consécutives de deux cellules de lieu distinctes et sont obtenues par apprentissage hebbien. Le choix de la prochaine action est le résultat d'un double mécanisme de propagation des transitions avant et arrière. La propagation arrière des transitions à partir d'une cellule but (goal planning) permet d'obtenir le plus court chemin vers le but lorsque cette propagation rencontre les cellules représentant les prochaines transitions accessibles depuis le lieu courant, biaisant ainsi le mécanisme de compétition et générant l'action correspondante.

Le modèle proposé par Arleo et Gerstner [Arleo 00a, Arleo 00b, Arleo 01, Arleo 04] présente l'intérêt d'associer l'utilisation d'informations visuelles d'une part, et la mise en œuvre d'un intégrateur de chemin pour la création des cellules de lieu d'autre part. L'apprentissage de ces représentations spatiales est non supervisé et basé sur un apprentissage hebbien durant une phase d'exploration de l'environnement. La scène visuelle est traitée par un ensemble de filtres de Gabor pour en extraire une représentation stable. Les représentations allothétiques (issues des entrées visuelles) et idiothétiques (issues de la proprioception) se combinent par un nouvel apprentissage hebbien à un niveau supérieur de l'architecture modélisant les cellules de lieu de l'hippocampe. C'est la combinaison de ces deux modalités qui permet de répondre au problème de l'aliasing

perceptif.

IV. Positionnement de notre approche

Nos travaux se sont focalisés sur l'apprentissage des positions d'objets dans un repère centré sur le robot. De ce fait, il n'existe pas, dans notre système de représentation allocentrique de l'espace. Notre approche est cependant à relier à la modélisation des cellules de vues chez les primates : l'activité de ces neurones sera dépendante de l'orientation du robot. Nous définirons ainsi des détecteurs neuronaux sensibles à des positions particulières d'un objet par rapport au robot, ainsi que les champs de position associés. L'apprentissage associatif permettra également d'apprendre des transitions entre différentes positions par l'intermédiaire des actions produites par le système.

2.3 La sélection de l'action par renforcement

I. Origines

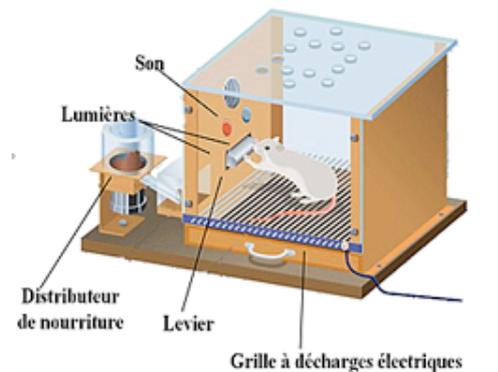


FIG. 2.10 – La boîte de Skinner est un dispositif expérimental qui fut inventé par Burrhus Frederic Skinner (1904-1990) pour tester la capacité des rongeurs à subir un conditionnement opérant. Le comportement de l'animal est renforcé par deux types de stimuli : la prise de nourriture pour un renforcement positif, un choc électrique pour un renforcement négatif.

L'objectif de l'apprentissage par renforcement est d'apprendre les actions que doit produire un agent dans son environnement, étant donné un état perçu, afin de maximiser un signal de gain [Sutton 98]. Le problème de la représentation de l'environnement est donc écarté dans la mesure où les capacités perceptives sont supposées acquises. L'apprentissage par renforcement est un domaine dont les fondements sont multiples. Tout d'abord, les théories behavioristes de l'apprentissage et plus précisément le *conditionnement opérant* ou *apprentissage skinnerien* est basé sur la loi de l'effet de Thorndike : un comportement suivi d'une récompense sera associé à la situation qui

l'a déclenché. Le conditionnement opérant tente d'expliquer les comportements appris lors de l'ontogénèse¹ de l'organisme. Il repose sur deux éléments : le *renforcement* et la *punition*. Ceux-ci peuvent être positifs ou négatifs. Les *renforceurs* peuvent être *primaires* : répondant à un besoin essentiel de l'individu (nourriture, douleur...) ou *secondaires* : nécessitant un apprentissage préalable (jouet, argent...). Cette branche de la psychologie demeure sujette à controverses dès lors qu'elle considère l'agent comme une "boîte noire" et n'envisage pas la construction des connaissances ou les phénomènes de l'inconscient. Elle est cependant utilisée dans les thérapies comportementales et cognitives afin de traiter les troubles anxieux et la dépression. L'apprentissage par renforcement est également largement inspiré des travaux réalisés dans le domaine du contrôle optimal et plus précisément de la *programmation dynamique* inventée par le professeur Richard Bellman [Bellman 57]. Les algorithmes associés permettent le calcul d'une politique optimale étant donné un modèle de l'environnement parfaitement modélisé dans le cadre d'un *Processus Décisionnel Markovien* (MDP). Bien que ces techniques soient de peu d'utilité dans le cadre de la robotique mobile étant donnée l'hypothèse du modèle parfait et l'ampleur des calculs à réaliser rendant difficile une utilisation temps-réel, l'équation de Bellman demeure un des fondements de l'apprentissage par renforcement dont les principes seront exposés plus loin. Plus largement, l'intelligence artificielle a contribué au développement de ces techniques par la planification dans l'espace d'état notamment. Ces domaines partagent une représentation commune des interactions entre l'agent et l'environnement sous la forme d'états et d'actions et l'utilisation d'un signal de gain et/ou d'utilité. Cette partie de l'exposé est consacrée à la présentation des principes et méthodes de l'apprentissage par renforcement permettant de mettre en évidence les mécanismes similaires employés dans notre système et d'en souligner les différences.

II. Principes généraux

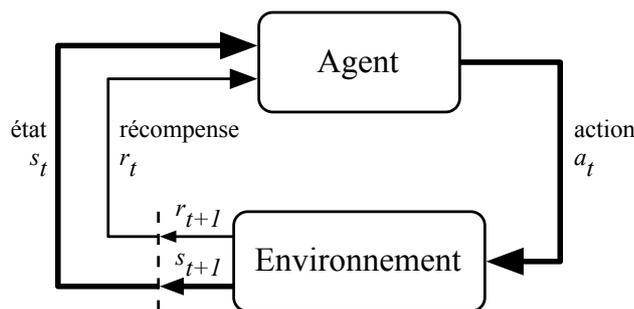


FIG. 2.11 – Interaction entre agent et environnement dans l'apprentissage par renforcement. Traduit de [Sutton 98].

Dans le cadre de l'apprentissage par renforcement, l'agent est en interaction constante

¹L'ontogénèse (ou ontogénie) décrit le développement progressif d'un organisme depuis sa conception jusqu'à sa forme mature, voire jusqu'à sa mort.

avec l'environnement. Le temps est discrétisé, $t = 1, 2, 3, \dots$. A chaque pas de temps t , l'agent reçoit de l'environnement une information concernant son *état* $s_t \in \mathcal{S}$, où \mathcal{S} représente l'ensemble des états possibles et sélectionne une *action* $a_t \in \mathcal{A}(s_t)$ avec $\mathcal{A}(s_t)$, l'ensemble des actions possibles à partir de l'état s_t . A l'issue de son action a_t , l'agent reçoit une valeur de *récompense* r_{t+1} ainsi que le nouvel état courant s_{t+1} . Ce mécanisme est illustré par la Figure 2.11. A chaque instant t l'agent met en œuvre une *politique* $\pi_t : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ qui associe à chaque état s et à chaque action a possible dans s la probabilité $\pi_t(a, s)$ de choisir a dans s . Si l'agent est déterministe, la politique désigne une fonction de \mathcal{S} dans \mathcal{A} qui à tout état associe l'action à entreprendre. On a alors $\pi(s) \in \mathcal{A}(s)$. La dynamique de l'environnement peut se résumer en la réalisation de deux fonctions dans $\mathcal{S} : (s_t, a_t) \rightarrow s_{t+1}$ et dans $\mathbb{R} : (s_t, a_t) \rightarrow r_{t+1}$ définissant d'une part l'interaction physique (composante objective) de l'agent avec l'environnement et d'autre part le gain associé à cette interaction (composante subjective). La notion de gain est donc ici incluse dans la dynamique de l'environnement. L'agent doit donc à chaque instant t choisir une séquence d'actions maximisant un gain cumulé R_t défini à horizon fini par :

$$R_t = r_{t+1} + r_{t+2} + r_{t+3} + \dots + r_T = \sum_{i=t+1}^T r_i \quad (2.1)$$

Lorsque l'horizon est infini, le calcul du gain cumulé avec intérêt devient :

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots + r_T = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (2.2)$$

Avec γ un taux d'intérêt fixé tel que $0 \leq \gamma \leq 1$. Ce taux permet de donner la priorité aux gains les plus proches dans le temps.

La notion d'*utilité* permet d'envisager les potentialités de gain d'une politique étant donné un *état* ou un couple (*état, action*) par l'intermédiaire de deux fonctions $V^\pi(s)$ et $Q^\pi(s, a)$ et de l'espérance de gain E_π en suivant cette politique :

$$V^\pi(s) = E_\pi \{R_t | s_t = s\} \quad (2.3)$$

$$Q^\pi(s, a) = E_\pi \{R_t | s_t = s, a_t = a\} \quad (2.4)$$

Si l'horizon est infini, ces fonctions d'*utilité* deviennent :

$$V^\pi(s) = E_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s \right\} \quad (2.5)$$

$$Q^\pi(s, a) = E_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a \right\} \quad (2.6)$$

Grâce à ces fonctions il est alors possible de définir une relation d'ordre sur les politiques :

$$\pi \geq \pi' \Leftrightarrow V^\pi(s) \geq V^{\pi'}(s), \forall s \in \mathcal{S} \quad (2.7)$$

Une politique optimale π^* est alors supérieure ou égale à toute autre politique. Les fonctions d'utilité associées à une politique optimale sont les suivantes :

$$V^*(s) = \max_{\pi} V^{\pi}(s), \forall s \in \mathcal{S} \quad (2.8)$$

$$Q^*(s, a) = \max_{\pi} Q^{\pi}(s, a), \forall s \in \mathcal{S} \text{ et } \forall a \in \mathcal{A} \quad (2.9)$$

Nous venons de présenter brièvement sous quelle forme sont représentées, dans le cadre de l'apprentissage par renforcement, les interactions entre l'agent et l'environnement. L'objectif d'un tel apprentissage est de permettre à cet agent d'accumuler l'information nécessaire à une décision immédiate maximisant l'espérance de gain dans un horizon fini ou non. Cette information locale est représentée par le biais des fonctions d'utilité $V^{\pi}(s)$ et $Q^{\pi}(s, a)$ que le système doit construire tout en améliorant sa politique π .

III. Méthodes d'évaluation des politiques

A. APPRENTISSAGE EN ENVIRONNEMENT CONNU : LA PROGRAMMATION DYNAMIQUE

Dans cette section nous allons donc considérer un agent possédant une connaissance locale de la dynamique de l'environnement sous la forme de probabilités de transition entre états $\mathcal{P}_{ss'}^a$, ainsi que les renforcements associés $\mathcal{R}_{ss'}^a$. Etant donnée une politique π , une première phase vise le calcul des fonctions d'utilité. Nous considérerons l'évaluation de la fonction $V^{\pi}(s)$ sachant que la fonction $Q^{\pi}(s, a)$ peut être évaluée de façon similaire. La deuxième phase devra permettre l'amélioration de la politique courante grâce au calcul des valeurs locales d'utilité.

$$\begin{aligned} V^{\pi}(s) &= E_{\pi} \left\{ R_t \middle| s_t = s \right\} \\ &= E_{\pi} \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \middle| s_t = s \right\} \\ &= E_{\pi} \left\{ r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} \middle| s_t = s \right\} \\ &= \sum_a \pi(s, a) \sum_{s'} \mathcal{P}_{ss'}^a \left[\mathcal{R}_{ss'}^a + \gamma E_{\pi} \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} \middle| s_{t+1} = s' \right\} \right] \\ &= \sum_a \pi(s, a) \sum_{s'} \mathcal{P}_{ss'}^a \left[\mathcal{R}_{ss'}^a + \gamma V^{\pi}(s') \right] \end{aligned} \quad (2.10)$$

L'équation 2.10 montre le développement de la fonction d'utilité sous la forme d'une accumulation de gains suivant les trajectoires fixées par la politique π . La dernière ligne traduit la dépendance existant entre l'espérance de gain à l'état s et l'espérance de gain des états successeurs de s : c'est l'équation de Bellman pour V^{π} [Bellman 57]. A partir

de cette formulation, il est relativement aisé de mettre en place une procédure itérative d'approximation de la fonction d'utilité telle que :

$$\begin{aligned} V_{k+1}(s) &= E_{\pi} \left\{ r_{t+1} + \gamma V_k(s_{t+1}) \middle| s_t = s \right\} \\ &= \sum_a \pi(s, a) \sum_{s'} \mathcal{P}_{ss'}^a \left[\mathcal{R}_{ss'}^a + \gamma V_k(s') \right] \end{aligned} \quad (2.11)$$

On montre facilement que $V_k = V_{\pi}$ est un point fixe de cette règle itérative et que la séquence $\{V_k\}$ converge vers V_{π} si $\gamma < 1$ ou si l'horizon est fini.

L'évaluation d'une politique est donc un processus réalisable en ligne en utilisant les informations locales de gain. Nous allons à présent nous intéresser à l'amélioration d'une politique existante. L'amélioration de politique se base sur un théorème essentiel reliant une amélioration locale de l'espérance de gain à l'amélioration globale d'une politique. Ce théorème est ici énoncé dans le cas de politiques déterministes mais peut être étendu au cas des politiques non-déterministes.

Théorème 2.1. *Soient π et π' deux politiques déterministes, telles que, pour tout état $s \in \mathcal{S}$:*

$$Q^{\pi}(s, \pi'(s)) \geq V^{\pi}(s) \quad (2.12)$$

Alors la politique π' doit être au moins aussi bonne que la politique π , ce qui signifie que, pour tout état $s \in \mathcal{S}$:

$$V^{\pi'}(s) \geq V^{\pi}(s) \quad (2.13)$$

Ce théorème ouvre la voie aux processus itératifs d'amélioration de politique dans la mesure où une amélioration locale à quelques états permet une amélioration globale de la politique. Les processus itératifs d'amélioration de politique que nous allons envisager s'appuient sur une alternance entre phase d'évaluation et phase d'amélioration. De plus on montre qu'il n'est pas nécessaire d'attendre la convergence de la phase d'évaluation. Si l'on alterne une seule itération de la phase d'évaluation de politique entre chaque phase d'amélioration on obtient l'*algorithme itérateur de valeur* :

$$\begin{aligned} V_{k+1}(s) &= \max_{a \in \mathcal{A}} E \left\{ r_{t+1} + \gamma V_k(s_{t+1}) \middle| s_t = s, a_t = a \right\} \\ &= \max_{a \in \mathcal{A}} \sum_{s'} \mathcal{P}_{ss'}^a \left[\mathcal{R}_{ss'}^a + \gamma V_k(s') \right] \end{aligned} \quad (2.14)$$

Les méthodes de programmation dynamique sont, comme nous l'avons vu, basées sur une connaissance parfaite de l'environnement. Ces méthodes présentent donc un intérêt limité dans le cadre de l'apprentissage ouvert en robotique. Toutefois, elles introduisent l'idée de mise à jour incrémentale de la politique courante qui est également exploitée dans les méthodes prenant en compte une connaissance partielle de l'environnement qui vont à présent nous intéresser.

B. MÉTHODE DE MONTE-CARLO

Les probabilités de transition entre état $\mathcal{P}_{ss'}^a$ et la valeur des renforcements $\mathcal{R}_{ss'}^a$ n'étant pas disponibles, cette méthode va consister en une mesure empirique des espérances de gain de chaque état ou couple (état, action) (équations 2.3 et 2.4). Cette mesure sera obtenue en moyennant les gains amassés à partir de ces états au cours d'*épisodes* ou séquences d'états et d'actions se terminant dans des états terminaux. Les méthodes de Monte-Carlo s'appliquent donc à des tâches épisodiques pour lesquelles sont caractérisés des états terminaux. Cette caractéristique ne nous permettra donc pas de les appliquer au problème de l'apprentissage ouvert. L'apprentissage par renforcement est cependant à l'origine de méthodes hybrides prenant en compte des tâches continues. Ces méthodes constitueront la suite de notre exposé.

C. MÉTHODE DES DIFFÉRENCES TEMPORELLES

Les méthodes des différences temporelles sont un croisement entre la méthode de la programmation dynamique et la méthode de Monte-Carlo. Tout comme la programmation dynamique, cette méthode utilise la corrélation entre les états pour mettre à jour l'évaluation de l'espérance de gain sans recourir à un modèle de l'environnement. Lors de la phase d'évaluation, la méthode de Monte-Carlo effectue une mesure empirique de l'accumulation de gain durant un épisode donné afin d'approcher l'espérance de gain (équation 2.15) par le biais du calcul de la moyenne des gains sur de multiples épisodes.

$$V(s_t) \leftarrow V(s_t) + \alpha \left[R_t - V(s_t) \right] \quad (2.15)$$

La programmation dynamique, pour sa part, évalue l'espérance de gain en s'appuyant sur la relation liant l'utilité de deux états successeurs :

$$V_{k+1}(s) = \sum_a \pi(s, a) \sum_{s'} \mathcal{P}_{ss'}^a \left[\mathcal{R}_{ss'}^a + \gamma V_k(s') \right] \quad (2.16)$$

La méthode des différences temporelles utilise une mise à jour locale de cette estimation :

$$V_t^\pi(s) \leftarrow V_t^\pi(s) + \alpha \left[r_{t+1} + \gamma V_t^\pi(s_{t+1}) - V_t^\pi(s) \right] \quad (2.17)$$

Cette méthode permet donc une mise à jour en ligne de l'estimation des espérances de gains. Elle semble donc être un bon candidat pour un apprentissage ouvert lorsque le modèle de l'environnement est incomplet, voir inconnu.

IV. Positionnement de notre approche

Etant donné que notre système d'apprentissage doit construire ses propres représentations de l'environnement celui-ci ne dispose, à un instant donné que d'un modèle incomplet des transitions entre états et des valeurs de récompense associés aux états. Cette limitation mise à part, il serait tout du moins possible de mettre à jour à chaque instant l'ensemble des valeurs d'utilité, parmi le modèle courant, par les méthodes de

la programmation dynamique. Nous avons cependant choisi d'écartier cette première solution car celle-ci produirait un grand nombre d'itérations à chaque mise à jour du modèle, ce qui rendrait impossible un calcul en temps réel. Nous nous sommes donc focalisé sur les méthodes de mise à jour incrémentales des fonction d'utilité tout en tirant avantage de l'existence d'un modèle en construction. Nous souhaitons en effet exploiter les temps d'attente disponibles durant la réalisation d'une action, pour mettre en œuvre une mise à jour partielle des fonctions d'utilité lors d'une phase de prédiction à horizon fini. Dans le cas d'un temps d'attente nul, une mise à jour semblable à celle des différences temporelles aura lieu, sinon cette mise à jour sera précédée d'une réévaluation de l'utilité des états successeurs dans un horizon de prédiction proportionnel au temps d'attente. L'implémentation de ce mécanisme au sein de la mémoire associative sera présentée dans le Chapitre 5.

Références

- [Arleo 99] A. Arleo, J. del R. Millán & D. Floreano. *Efficient learning of variable-resolution cognitive maps for autonomous indoor navigation*. IEEE Trans. Robot. Automat., vol. 15, pages 990–1000, 1999.
- [Arleo 00a] A. Arleo & W. Gerstner. *Modeling Rodent Head-direction Cells and Place Cells for Spatial Learning in Bio-mimetic Robotics*. Dans Floreano D. Roitblat H.L. Meyer J.A. Berthoz A. & Wilson S.W., editeurs, *From Animals to Animats VI*, pages 236–245. MIT Press, Cambridge MA, 2000.
- [Arleo 00b] A. Arleo & W. Gerstner. *Spatial cognition and neuro-mimetic navigation : A model of hippocampal place cell activity*. Dans Biological Cybernetics, 2000.
- [Arleo 01] A. Arleo & W. Gerstner. *Spatial Orientation in Navigating Agents : Modeling Head-direction Cells*. Neurocomputing, vol. 38-40, pages 1059–1065, 2001.
- [Arleo 04] A. Arleo, F. Smeraldi & W. Gerstner. *Cognitive Navigation Based on Nonuniform Gabor Space Sampling, Unsupervised Growing Networks, and Reinforcement Learning*. Dans IEEE Trans Neural Networks, volume 15, pages 639–52, 2004.
- [Bellman 57] R. Bellman. *Dynamic programming*. Princeton University Press, 1957.
- [Biederman 87] I. Biederman. *Recognition-by-components : a theory of human image understanding*. Psychological Review, vol. 94, pages 115–147, 1987.
- [Biederman 93] I Biederman & P. C. Gerhardstein. *Recognizing depth-rotated objects : evidence and conditions for three-dimensionnal viewpoint invariance*. Journal of Experimental Psychology : Human Perception and Performance, vol. 19, no. 6, pages 1162–1182, 1993.

- [Biederman 00] I. Biederman. *Recognizing depth-rotated objects : A review of recent research and theory*. Spatial Vision, vol. 13, no. 2-3, pages 241–253(13), 2000.
- [Bülthoff 94] H. Bülthoff, S. Edelman & M. Tarr. *How Are Three-Dimensional Objects Represented in the Brain ?* Rapport technique, Massachusetts Institute of Technology Artificial Intelligence Laboratory and Center for Biological and Computational Learning Department of Brain and Cognitive Sciences, 1994.
- [Chatila 85] R. Chatila & J. P. Laumond. *Position referencing and consistent world modeling for mobile robots*. Dans IEEE Int. Conf. Robotics Automation, 1985.
- [Cuperlier 06] N. Cuperlier, M. Quoy, C. Giovannangeli, Ph. Gaussier & Ph. Laroque. *Transition Cells for navigation and planning in an unknown environment*. Dans The Society For Adaptive Behavior SAB'2006, pages 286–297, <http://publi-etis.ensea.fr/2006/CQGGLO6>.
- [Elfes 87] A. Elfes. *Sonar-based real-world mapping and navigation*. Dans IEEE Journal of Robotics and Automation, volume 3, pages 249–265, 1987.
- [Elfes 89] A. Elfes. *Occupancy Grids : A Probabilistic Framework for Robot Perception and Navigation*. Thèse de doctorat, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1989.
- [Gaussier 02] P. Gaussier, A. Revel, J.P. Banquet & V. Badaeu. *From view cells and place cells to cognitive map learning : processing stages of the hippocampal system*. Biological cybernetics, vol. 86, no. 1, pages 15–28, 2002.
- [Georges-François 99] P. Georges-François, E.T. Rolls & R.G. Robertson. *Spatial view cells in the primate hippocampus : allocentric view not head direction or eye position or place*. Cerebral Cortex, vol. 9, pages 197–212, 1999.
- [Gibson 79] James J. Gibson. *The ecological approach to visual perception*. Lawrence Erlbaum Associates, 1979.
- [Kortenkamp 94] D. Kortenkamp & T. Weymouth. *Topological mapping for mobile robots using a combination of sonar and vision sensing*. Dans the Twelfth National Conference on Artificial Intelligence, pages 979–984, 1994.
- [Kuipers 78] B. Kuipers. *Modeling Spatial Knowledge*. Cognitive Science, vol. 2, pages 129–153, 1978.
- [Kuipers 91] B. Kuipers & Y.T. Byun. *A Robot Exploration and Mapping Strategy Based on a Semantic Hierarchy of Spatial Representations*. Dans Robot. Autonomous Syst, volume 8, pages 47–63, 1991.

- [Leprêtre 00] S. Leprêtre, P. Gaussier & J.P. Cocquerez. *From Navigation to Active Object Recognition*. Dans *Simulation of Adaptive Behavior*, SAB'2000, 2000.
- [Moravec 85] H. P. Moravec & A. Elfes. *High resolution maps from wide angle sonar*. Dans *IEEE Int. Conf. Robotics Automation*, pages 116–121, 1985.
- [Moravec 88] H. P. Moravec. *Sensor fusion in certainty grids for mobile robots*. *A.I. Magazine*, 1988.
- [O'Keefe 71] J. O'Keefe & J. Dostrovsky. *The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat*. *Brain Research*, vol. 34, no. 1, pages 171–175, 1971.
- [O'Keefe 78] J. O'Keefe & L. Nadel. *The hippocampus as a cognitive map*. Oxford University Press, 1978.
- [O'Keefe 96] J. O'Keefe & N. Burgess. *Geometric determinants of the place fields of hippocampal neurons*. *Nature*, vol. 381, pages 425–428, 1996.
- [Poggio 90] T. Poggio & S. Edelman. *A Network That Learns to Recognize Three Dimensional Objects*. *Letters to Nature*, vol. 343, pages 263–266, 1990.
- [Redish 97a] A.D. Redish & D.S. Touretzky. *Cognitive maps beyond the hippocampus*. *Hippocampus*, vol. 7, pages 15–35, 1997.
- [Redish 97b] A.D. Redish & D.S. Touretzky. *Navigating with landmarks : Computing goal locations from place codes*. Dans Ikeuchi K. & Veloso M., éditeurs, *Symbolic Visual Learning*, chapitre 12, pages 325–351. Oxford University Press, 1997.
- [Riesenhuber 99] M. Riesenhuber & T. Poggio. *Hierarchical models of object recognition in cortex*. *Nature Neuroscience*, vol. 28, pages 1019–1025, 1999.
- [Riesenhuber 00] M. Riesenhuber & T. Poggio. *Models of Object Recognition*, 2000.
- [Rolls 95] E.T. Rolls & S.M. O'Mara. *View-responsive neurons in the primate hippocampal complex*. *Hippocampus*, vol. 5, no. 409-424, 1995.
- [Rolls 99] E. T. Rolls. *Spatial View Cells and the Representation of Place in the Primate Hippocampus*. *Hippocampus*, vol. 9, pages 467–480, 1999.
- [Rolls 04] E. T. Rolls, J. Xiang & L. Franco. *Object, Space, and Object-Space Representations in the Primate Hippocampus*. *Journal of Neurophysiology*, vol. 94, pages 833–844, 2004.
- [Rolls 05] E. T. Rolls & J. Xiang. *Reward-Spatial View Representations and Learning in the Primate Hippocampus*. *The Journal of Neuroscience*, vol. 25, no. 26, pages 6167–6174, 2005.

- [Rolls 06] E.T. Rolls & J. Xiang. *Spatial view cells in the primate hippocampus and memory recall*. Reviews in neurosciences, vol. 17, no. 1-2, pages 175–200, 2006.
- [Sharp 91] P.E. Sharp. *Computer simulation of hippocampal place cells*. Psychobiology, vol. 19, no. 2, pages 103–115, 1991.
- [Sutton 98] Richard S. Sutton & Andrew G. Barto. Reinforcement learning : An introduction (adaptive computation and machine learning). The MIT Press, March 1998.
- [Tarr 95] M. Tarr & H. Bülthoff. *Is human object recognition better described by geon structural description or by multiple views ? Comment on Biedermann and Gerhardstein (1993)*. Journal of Experimental Psychology : Human Perception and Performance, vol. 21, pages 1494–1505, 1995.
- [Tarr 98] M. Tarr & H. Bülthoff. *Image-based object recognition in man, monkey and machine*. Cognition, no. 67, pages 1–20, 1998.
- [Thrun 98] S. Thrun. *Learning maps for indoor mobile robot navigation*. Artificial Intelligence, vol. 89, pages 21–71, 1998.
- [Touretzky 96] D.S. Touretzky & A.D. Redish. *A theory of rodent navigation based on interacting representations of space*. Hippocampus, vol. 6, no. 3, pages 247–270, 1996.
- [Wan 94] H.S. Wan, D.S. Touretzky & A.D. Redish. *Towards a Computational Theory of Rat Navigation*. Dans Proceedings of the 1993 Connectionist Models Summer School, pages 11–19, 1994.
- [y Cajal 11] Santiago Ramón y Cajal. *Histologie du système nerveux de l’homme et des vertébrés ; vol 1 et 2*. A. Maloine Paris, 1911.
- [Yamauchi 96] B. Yamauchi & R. Beer. *Spatial learning for navigation in dynamic environments*. IEEE Transactions on Systems, Man, and Cybernetics - Part B : Cybernetics. Special Issue on Learning Autonomous Robots, 1996.

APPRENDRE ET RECONNAÎTRE

les mécanismes fondamentaux

3.1	Problématique	48
	I. Apprentissage en ligne	48
	II. Apprentissage non-supervisé	49
3.2	Neurone et modèles	50
	I. Le neurone biologique	51
	II. Les modèles du neurone, décrire c'est choisir	53
	III. Les modèles biophysique	54
	IV. L'approche fréquentielle	55
	V. L'approche impulsionnelle	57
	VI. L'approche stochastique	57
	VII. Positionnement de notre approche	58
3.3	Neurones, cartes et connectivité	59
	I. Cartes neuronales, invariance à la translation	59
	II. Hiérarchie et compétition	60
3.4	Le modèle "intègre et tire"	60
	I. Intégration	60
	II. Tir	62
3.5	Apprentissage et compétition	65
	I. La règle de Hebb	65
	II. La compétition par inhibitions latérales	65
	III. Compétition à l'activation	67
	IV. Loi d'apprentissage	68
	V. Matrices de poids adaptatives	69
3.6	Les 3 voies de la perception	71
3.7	La représentation des objets	72
	I. Représentation par vues, structure du réseau	72
	II. Association temporelle et spatiale de vues, invariance à l'objet	74
3.8	La représentation des positions et des actions	75
	I. Cartes temporisées	75

II.	La représentation égocentrique de l'espace	79
III.	La représentation proprioceptive des actions	80
IV.	L'apprentissage incrémental de motifs de position et d'action	80
3.9	Récompense perçue, récompense induite	81
3.10	Conclusion	83
	Références	83

3.1 Problématique

L'état du monde n'est pas directement accessible. Seule la mesure, l'observation par l'intermédiaire de capteurs permet de mettre en lumière la structure de l'environnement. Les capteurs selon leur nature, ne fournissent qu'un reflet incomplet et plus ou moins dégradé de la réalité : l'image enregistrée par une caméra ne rendra pas compte de l'information de profondeur dans la scène observée, un capteur laser sera incapable de mesurer la couleur d'un objet... Il semble donc à première vue difficile de concevoir un système capable de construire de manière autonome une représentation fidèle de l'environnement dans lequel il évolue. Le robot dispose en plus de ses capteurs d'une mémoire capable de conserver la trace des observations passées. Cette mémoire est un outil essentiel pour détecter des structures stables au cours du temps et lors du changement de point de vue et ainsi identifier des classes d'objets perceptifs. Mesure et mémoire sont les deux fonctions essentielles de nos unités de traitement, les neurones. Nous présenterons dans ce chapitre le modèle de réseau de neurones artificiel implanté dans le simulateur appelé NeuSter, qui a servi à la fois d'outil et de plate-forme d'expérimentation durant toute la durée de nos travaux. Dans un premier temps, nous présenterons les principes de l'apprentissage en ligne non supervisé ainsi que les solutions proposées dans la littérature. Dans un deuxième temps, nous décrirons le modèle de neurone et les schémas de connexion qui sont exploités pour la tâche de décomposition du contexte perceptif dans la boucle sensori-motrice présentée au Chapitre 1. Cette présentation sera précédée d'un bref panorama de trois différentes familles de modèles de neurones proposées dans la littérature. Enfin, la dernière partie de ce chapitre sera consacrée à l'exploitation de notre modèle de neurone pour la tâche de décomposition du contexte perceptif selon trois axes : l'apprentissage des représentations d'objet, l'apprentissage de motifs de position et l'apprentissage de motifs d'actions.

I. Apprentissage en ligne

Notre système doit permettre un apprentissage en ligne (on-line learning) de données perçues en temps réel. Ce type d'apprentissage est à opposer aux méthodes d'apprentissage hors-ligne (batch learning) pour lesquelles toutes les données d'apprentissage sont fournies au système durant une phase d'apprentissage indépendante de la phase d'exploitation sur des données réelles. Il est donc primordial de limiter au maximum le temps de calcul global qui, dans le cadre des réseaux de neurones formels, inclut à la fois le temps de traitement unitaire du neurone et le temps de propagation de l'infor-

mation au sein du réseau. Notre objectif étant d'autre part un apprentissage ouvert, donc une acquisition incrémentale de connaissances, nous devons limiter l'incidence de l'augmentation des connaissances sur la performance du réseau.

Le simulateur NeuSter a été conçu afin de permettre la distribution des calculs sur de multiples machines connectées au sein d'un réseau local. Il est ainsi bien adapté au calcul de réseaux de grandes tailles. Cependant, pour que le parallélisme du calcul soit vraiment efficace, il est nécessaire d'y adjoindre une forme compacte de transmission de l'information. Une solution à ce problème est d'utiliser un modèle de propagation "event driven" où seules les informations de décharge des neurones sont propagées. Le nombre de neurones actifs à un instant donné étant limité, la quantité d'information véhiculée est donc de taille relativement faible et celle-ci varie selon l'activité du réseau.

II. Apprentissage non-supervisé

Notre système doit être capable d'apprendre sur la base d'informations perçues au cours de son expérience. Nous considérons qu'il existe une structure, une cohérence liant cette information et générant des données perceptives stables et reproductibles. Aucun oracle (ou expert) ne doit renseigner le système sur les structures, les catégories ou plus généralement les données cachées à l'observation. Il existe dans la littérature divers modèles répondant à ces attentes. Tout d'abord, dans le domaine de la fouille de données (data mining) divers algorithmes ont été proposés dans le but d'extraire d'un ensemble de données une classification (ou partitionnement) pouvant prendre divers aspects. D'une part les classifications hiérarchiques très utilisées dans le domaine de la biostatistique visent à construire une hiérarchie de classes d'objets sous forme d'arbres, où les classes doivent être les plus homogènes possible et les plus différentes les unes des autres. Cependant, ces approches sont itératives et ne sont donc pas adaptées à un apprentissage en ligne de catégories. Des algorithmes de classification non hiérarchique ont d'autre part été proposés afin de partitionner des ensembles de données numériques issues d'observations. Nous renvoyons le lecteur à la méthode des k-moyennes et à l'algorithme espérance-maximisation (en anglais "Expectation-maximisation algorithm", souvent abrégé "EM") qui est par exemple utilisé dans le domaine des modèles de Markov cachés (HMM) pour l'évaluation des paramètres du modèle. Le principal désavantage de ce type d'algorithmes est qu'ils nécessitent une connaissance a priori du nombre de classes pour le partitionnement. Ces méthodes ne sont donc pas adaptées à un apprentissage ouvert de représentations. Dans le domaine des réseaux de neurone artificiels, les cartes auto-organisatrice (en anglais "Self Organizing Maps", SOM) proposées par Teuvo Kohonen [Kohonen 89] constituent une approche intéressante de catégorisation non-supervisée permettant par ailleurs de représenter des relations topologiques entre les catégories. La règle d'apprentissage que nous allons présenter utilise, à l'image des SOMs, une compétition entre les neurones pour sélectionner le meilleur candidat lors d'une phase d'apprentissage (principe "Winner-Take-All"). En revanche, le mécanisme de compétition ne sera pas confié à un algorithme global d'évaluation mais sera intégré localement au fonctionnement de chaque neurone.

3.2 Neurone et modèles

Depuis la découverte de l'effet transistor en 1947 par les américains John Bardeen, William Shockley et Walter Brattain et l'invention du premier microprocesseur, l'Intel 4004 en 1971, la puissance de calcul des ordinateurs n'a cessée de croître à un rythme exponentiel (Figure 3.1). Capables d'effectuer des opérations en des temps inférieurs à la nanoseconde et bénéficiant d'une quantité de mémoire considérable, les ordinateurs modernes sont devenus les outils indispensables à la réalisation de calculs complexes.

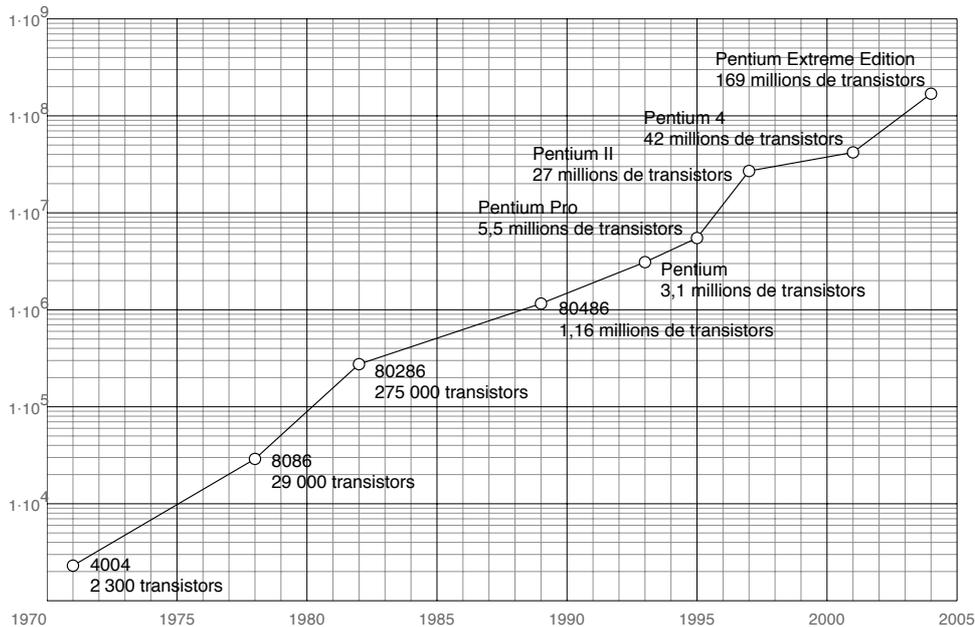


FIG. 3.1 – Evolution du nombre de transistors dans les processeurs Intel. Cette augmentation exponentielle fut rapidement nommée Loi de Moore du nom d'un ingénieur de Fairchild Semiconductor, un des trois fondateurs d'Intel, qui fut le premier à décrire ce phénomène.

Comparé à un ordinateur, le cerveau humain est incapable d'atteindre de telles performances en termes de précision, de vitesse de calcul et dans la capacité de restituer de façon fidèle une grande quantité d'informations mémorisées. Notre cerveau est cependant capable de traiter des informations imprécises capturées par nos sens à une vitesse considérable. Nos capacités d'apprendre et de reconnaître la parole, les gestes, les visages ainsi que de généraliser durant l'expérience n'ont pu être atteintes à l'heure actuelle par les ordinateurs modernes.

Pour comprendre les différences importantes de performances caractérisant ces deux systèmes, il est utile de comparer leurs architectures (Tableau 3.1). A la différence de l'ordinateur, le cerveau est assez lent (la durée de traitement de l'information par le neurone est de l'ordre de la milliseconde) mais ce défaut est largement compensé par

	Ordinateur Von Neumann	Système biologique
Processeur	complexe rapide (ns) un/quelques-un	simple lent (ms) un très grand nombre
Mémoire	séparée du processeur localisée non adressable par le contenu	intégrée au processeur distribuée adressable par le contenu
Calculs	centralisés séquentiels programmes mémorisés	distribués parallèles apprentissage
Fiabilité	vulnérable	robuste

TAB. 3.1 – Comparaison entre ordinateur et système biologique.

le parallélisme de son architecture. Le cerveau bénéficie en effet de capacités de calculs et d'une mémoire totalement distribués. De ce fait, la perte d'une unité de traitement n'est pas fatale à l'intégrité du système, contrairement aux ordinateurs pour lesquels la destruction d'un seul transistor peut être critique. Cette grande robustesse, associée aux capacités de généralisation du cerveau, ont donc été naturellement à l'origine de travaux désireux de s'inspirer du fonctionnement massivement parallèle des réseaux de neurones.

1. Le neurone biologique

C'est il y a plus de cent ans que l'histologue espagnol Santiago Ramón y Cajal décrivit pour la première fois le neurone comme une cellule polarisée véhiculant un signal électrique de son *arbre dendritique*, le long de son *axone* vers ses *synapses* (Figure 3.2). Le cerveau humain est constitué de près de 10^{11} neurones, chacun connecté à un ensemble composé de 10^3 à 10^4 autres neurones. Bien que la forme des neurones soit très variable, ils partagent, pour leur grande majorité, le même principe de fonctionnement. Des impulsions électriques appelées *potentiels d'action* sont véhiculées par chaque cellule grâce à une série d'échanges ioniques ayant lieu au niveau de la membrane (Figure 3.4). A la terminaison de l'axone, la synapse, l'arrivée d'un potentiel d'action entraîne une libération de neurotransmetteurs chimiques dans la fente synaptique séparant les éléments *présynaptiques* et *postsynaptiques* (Figure 3.3). Les *neurotransmetteurs* ainsi libérés sont captés par des récepteurs situés au niveau de la membrane du neurone postsynaptique. Des récepteurs particuliers dits *ionotropiques* participent à la création de courant ioniques à l'origine d'un *potentiel postsynaptique* qui peut être soit *excitateur* soit *inhibiteur*. Si la membrane dépasse un seuil de dépolarisation, un potentiel d'action est alors produit. Les milliers de synapses connectées à chaque neurone sont à l'origine d'une sommation à la fois spatiale et temporelle grâce à la capacité électrique de la membrane.

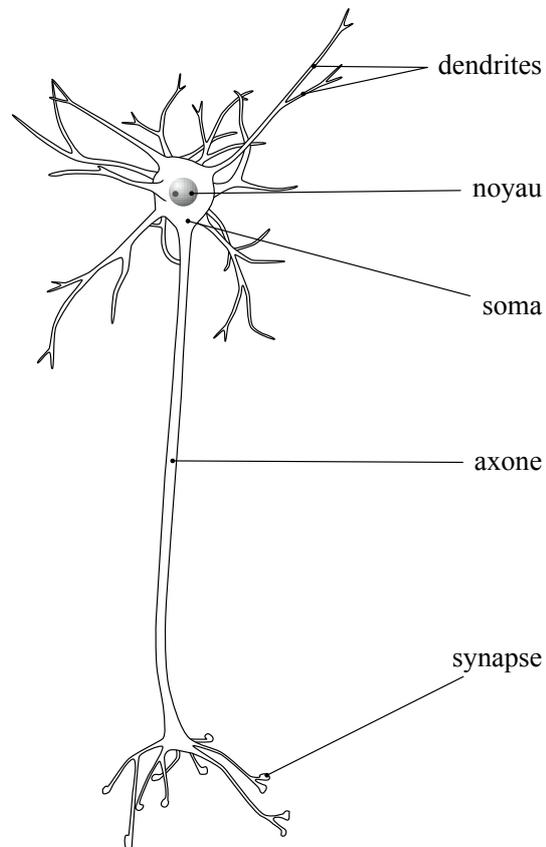


FIG. 3.2 – Schéma d'un neurone

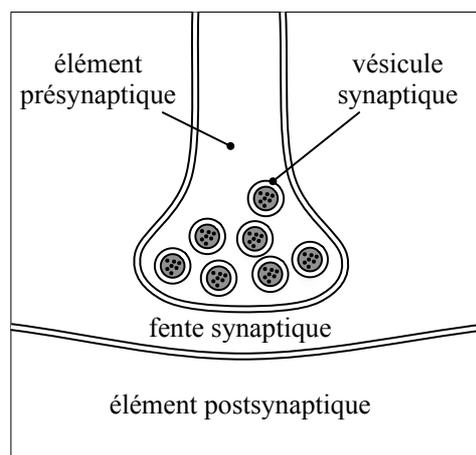


FIG. 3.3 – Schéma d'une synapse

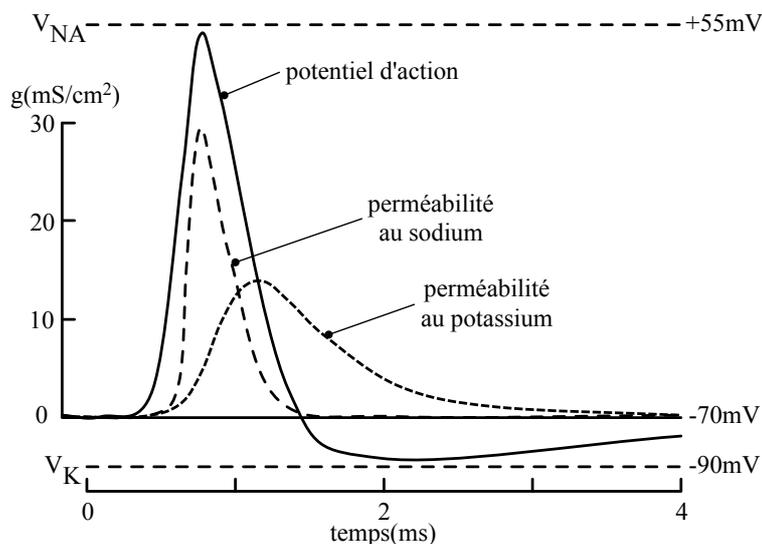


FIG. 3.4 – Evolution au cours du temps du potentiel d'action et des modifications de la perméabilité au sodium et au potassium

II. Les modèles du neurone, décrire c'est choisir

Il existe dans la littérature de nombreux modèles ayant pour objet de décrire le fonctionnement des neurones. Selon que l'on envisage le neurone dans son fonctionnement électro-chimique ou que l'on étudie les phénomènes d'apprentissage à l'échelle du réseau, les variables du modèle peuvent être très différentes. Les différents niveaux d'abstraction sont également liés à la nature de l'information qui doit être transmise entre les neurones au sein du modèle. Il est dès lors possible de distinguer les modèles biophysiques des modèles computationnels. Les premiers ont pour but l'étude des mécanismes internes et des échanges transmembranaires qui sont à l'origine du potentiel d'action et de sa propagation. Ces modèles sont complexes par nature, car ils mettent en jeu de nombreuses équations différentielles et ne sont pas, de ce fait, utilisés dans la simulation de réseaux de grande taille. Un deuxième type de modèles que nous qualifions de modèles computationnels étudient le neurone en tant qu'unité de traitement de l'information. Dès lors, se pose la question de la nature de cette information. Nous allons décrire brièvement trois types de modèles computationnels. Tout d'abord, les modèles impulsifs, dont la célèbre famille des modèles "intègre-et-tire", étudie la formation de trains d'impulsions à l'aide d'une représentation du temps explicite. Les modèles fréquentiels, qui se situent à un niveau d'abstraction supérieur, ont une représentation de l'activité du neurone sous la forme des fréquences d'activation et négligent donc l'aspect impulsif de cette activité. Nous présenterons enfin une troisième classe de modèle, les modèles stochastiques, qui introduisent la notion de probabilité de décharge, qui peut être reliée à la notion de fréquence instantanée de décharge, tout en conservant le caractère impulsif de l'information échangée entre les neurones.

III. Les modèles biophysique

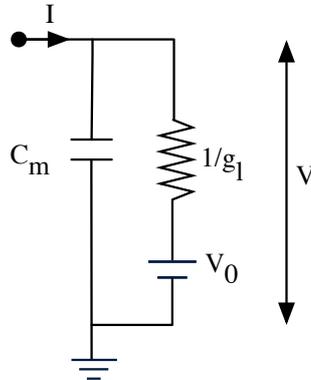


FIG. 3.5 – Schéma du modèle de Lapicque

Un premier type de modélisation consiste à considérer le neurone dans son fonctionnement électro-chimique. Les échanges ioniques ainsi que l'évolution du potentiel de membrane sont simulés afin de créer artificiellement les potentiels d'action. La membrane du neurone peut-être modélisée par un circuit électrique dans lequel une résistance R_m est associée en parallèle à une capacité C_m (Figure 3.5). Ce modèle introduit par Lapicque [Lapicque 07] dès 1907 permet de décrire l'évolution du potentiel de membrane selon l'équation suivante :

$$C_m \frac{dV}{dt} = -g_l(V - V_0) + I \quad (3.1)$$

Où V est le potentiel de membrane, I le courant entrant et g_l la conductance de fuite. Bien qu'il ne fût pas en mesure de formuler les lois régissant la création du potentiel d'action, ce modèle d'intégration avec fuite fut à l'origine de nombreux modèles dont le célèbre "intègre-et-tire" dont nos travaux sont inspirés.

Il faut attendre 1952 et les travaux de Hodgkin et Huxley pour mettre en lumière les mécanismes générateurs du potentiel d'action par le biais des canaux ioniques cellulaires. L'étude de l'axone géant de calamar par l'utilisation de techniques du *potentiel stabilisé* a en effet mis en évidence l'influence de la polarisation de l'axone sur la perméabilité ionique de la membrane vis à vis des ions sodium et potassium. La Figure 3.6 montre le schéma électrique prenant en compte ces deux types de canaux. Pour chaque ion x , un courant I_x est induit d'une part par la conductance membranaire relative à cet ion g_x et d'autre part par une force électromotrice $(E_x - V)$ où E_x est le potentiel d'équilibre de l'ion calculé grâce à l'équation de Nernst. Si l'on note P_x la proportion des canaux ouverts, on obtient l'équation suivante :

$$I_x = g_x P_x (E_x - V) \quad (3.2)$$

Ce modèle permet donc de simuler la génération d'un potentiel d'action pour un neurone soumis à un courant électrique (Figure 3.6). Les modèles biophysiques ont

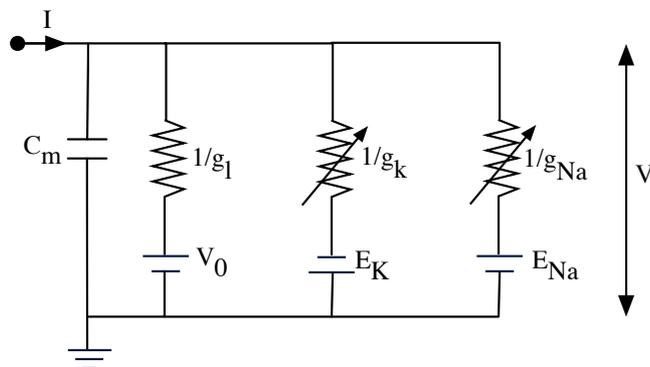


FIG. 3.6 – Schéma du modèle de Hodgkin et Huxley

offre une meilleure compréhension des phénomènes d'intégration et de transmission de l'information par les neurones, ouvrant ainsi la voie aux modèles computationnels.

IV. L'approche fréquentielle

L'histoire des modèles computationnels du neurone a débuté grâce aux travaux de quatre chercheurs : Warren McCulloch, Walter Pitts, Donald O. Hebb et Frank Rosenblatt. Les deux premiers sont à l'origine de travaux fondateurs poursuivis pendant les années 40. D.O. Hebb, psychologue canadien proposa une règle d'apprentissage synaptique qui subsiste aujourd'hui dans de nombreux modèles, dont celui décrit dans ce manuscrit [Hebb 49]. F. Rosenblatt a étendu le modèle de McCulloch et Pitts en développant le célèbre perceptron.

Le modèle de McCulloch et Pitts [McCulloch 43] décrit un neurone formel binaire à temps discret. La sortie du neurone $y(t)$ vaut 1 si un potentiel d'action est émis, 0 sinon. Chaque connexion est associée à un poids qui peut prendre 2 valeurs : 1 s'il s'agit d'une entrée excitatrice, -1 si c'est une entrée inhibitrice. Un potentiel d'action est émis si l'effet des connexions excitatrices et inhibitrices dépasse un seuil noté θ :

$$y(t+1) = \begin{cases} 1 & \text{si } \sum_{i=1}^n w_i x_i(t) > \theta, \\ 0 & \text{si } \sum_{i=1}^n w_i x_i(t) < \theta. \end{cases} \quad (3.3)$$

où $x_i(t)$ est la i ème entrée et w_i le i ème poids de connexion également appelé *poids synaptique*. Ce type de neurone est capable de reproduire des opérations logiques simples comme le montre la Figure 3.7.

Le perceptron de F. Rosenblatt [Rosenblatt 62] est une extension de ce modèle par l'utilisation de valeurs réelles de poids au lieu des simples connexions excitatrices et inhibitrices et la généralisation de la notion de seuil avec l'apparition d'une fonction d'activation notée ici f :

$$y(x) = f\left(\sum_{i=1}^n w_i x_i\right) \quad (3.4)$$

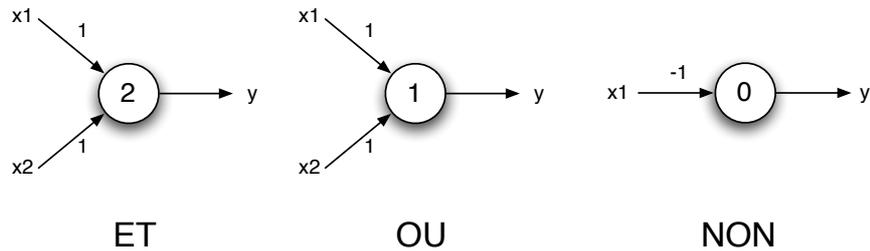


FIG. 3.7 – Fonctions logiques réalisées par le perceptron. Les valeurs des poids synaptiques sont indiquées près de chaque connexion, les valeurs de seuil sont inscrites à l'intérieur des cercles.

Les fonctions d'activation les plus communément utilisées sont linéaires, sigmoïdes et gaussiennes (Figure 3.8). Ce type de modèle est largement utilisé en neurosciences computationnelles [Rolls 98]. Différentes règles d'apprentissage peuvent leur être associées comme la règle du perceptron [Rosenblatt 62], la rétropropagation [Rumelhart 86], l'algorithme de Kohonen [Kohonen 77] ou l'apprentissage par renforcement [Sutton 98].

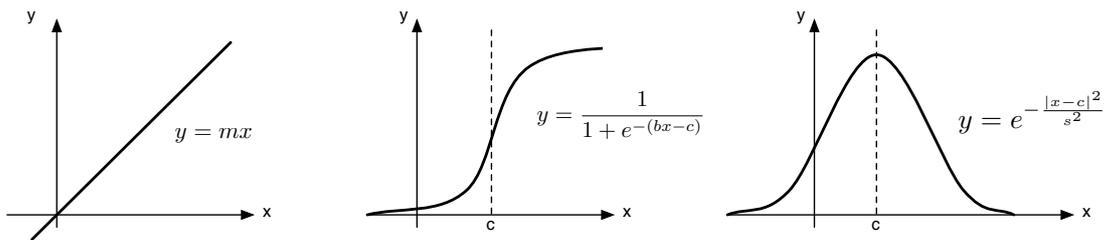


FIG. 3.8 – Fonctions d'activation couramment utilisées : linéaires, sigmoïdes et gaussiennes

Le perceptron, bien que prometteur ne pouvait pas être entraîné à reconnaître toutes les formes de motifs. Ce constat a entraîné un désintérêt pour ce champ de recherche pendant de nombreuses années avant l'avènement du perceptron multicouche. Le perceptron était en effet incapable d'apprendre des motifs non-linéairement séparables comme l'ont montrés Marvin Minsky et Seymour Papert dans un ouvrage nommé "Perceptrons" [Minsky 69]. Ils affirmèrent même faussement que cette incapacité ne serait pas levée par l'apport de couches supplémentaires. C'est Stephen Grossberg, trois ans plus tard qui contredit cette affirmation en présentant un réseau capable, entre autres, d'apprendre un OU-exclusif (XOR). Cette démonstration ne suffit pas à lever le doute sur les capacités du neurone formel, si bien que l'engouement pour la discipline ne réapparut que bien plus tard, durant les années 80¹.

¹Minsky et Papert publièrent en 1987 une version corrigée de leur ouvrage appelé "Perceptrons - Expanded Edition".

V. L'approche impulsionnelle

Pour présenter cette approche nous allons nous placer dans le formalisme SRM (Spike Response Model) [Gerstner 95]. L'état d'un neurone i est décrit par une variable d'état u_i qui vaut 0 lorsque le neurone est au repos. On dit que le neurone "décharge" lorsque la valeur u_i dépasse un seuil noté ϑ . L'instant de cette décharge est noté $t_i^{(f)}$. L'ensemble de tous les instants de décharge du neurone i est défini par :

$$\mathcal{F}_i = \{t_i^{(f)}; 1 \leq f \leq n\} = \{t | u_i(t) = \vartheta\} \quad (3.5)$$

La valeur de la variable d'état u_i est affectée par deux phénomènes. D'une part, après chaque décharge ayant lieu à un instant $t_i^{(f)}$, la valeur de u_i est réinitialisée en lui soustrayant une composante de réfraction noté $\eta_i(t - t_i^{(f)})$. Cette réinitialisation a pour but de simuler la période réfractaire, mesurée chez les neurones réels, qui succède à chaque potentiel d'action et durant laquelle le neurone devient moins sensible aux signaux d'excitation entrants. $\eta_i(s)$ est une fonction croissante négative qui vaut $-\vartheta$ pour $s \leq 0$ et tend vers 0 pour $s \rightarrow \infty$. La valeur de u_i est d'autre part affectée par les décharges provenant des neurones présynaptiques $j \in \Gamma_i$, où Γ_i est l'ensemble des neurones présynaptiques du neurone i . Selon la valeur du poids w_{ij} associé à la connexion ij , une décharge présynaptique $t_j^{(f)}$ peut augmenter ou diminuer la valeur de u_i par une quantité $w_{ij}\epsilon_{ij}(t - t_j^{(f)})$. La fonction ϵ_{ij} permet de définir un temps de transmission affectant la connexion ainsi que la durée d'influence d'une décharge présynaptique. L'évolution de l'état du neurone i est donc donnée par la formule :

$$u_i(t) = \sum_{t_i^{(f)} \in \mathcal{F}_i} \eta_i(t - t_i^{(f)}) + \sum_{j \in \Gamma_i} \sum_{t_j^{(f)} \in \mathcal{F}_j} w_{ij}\epsilon_{ij}(t - t_j^{(f)}) \quad (3.6)$$

Nous retrouvons, à droite de l'expression, la combinaison linéaire des deux types de contribution mentionnées plus haut. La prise en compte explicite du temps dans l'évolution de l'état du neurone et la génération des décharges a ouvert la voie à l'étude du codage temporel de l'information et notamment le codage par phase [Maass 96] et l'étude des synchronies et corrélations dans des groupes de neurones [Ritz 97].

VI. L'approche stochastique

Le calcul stochastique, qui a vu le jour dans les années 1960, est basé sur la manipulation de flux de données binaires représentant des valeurs réelles sous une forme fréquentielle. A la différence du codage par impulsion, le temps n'est pas explicitement exploité dans les calculs. Ce type de modèle sera, de ce fait, beaucoup plus simple à mettre en œuvre. D'un point de vue biologique, Srinivasan et Bernard furent les premiers à proposer l'existence d'un calcul de cette nature dans le neurone [Srinivasan 76], cette idée fut par la suite reprise par Koch et Poggio [Koch 92]. Du point de vue des neurosciences computationnelles, le modèle que nous allons présenter est celui qui constitue la première réalisation concrète de cette approche [Shawe-Taylor 91].

Une séquence de Bernoulli de paramètre p (appelée aussi séquence de Bernoulli p) est un flux de données binaires obtenu par des épreuves de Bernoulli indépendantes de paramètre p . Chaque bit de la séquence a donc une probabilité p d'être égal à 1. Notons $p \circ q$ la séquence de Bernoulli obtenue en appliquant bit à bit l'opération binaire \circ aux séquences p et q , par exemple :

$$p \text{ AND } q = pq \quad (3.7)$$

$$p \text{ XNOR } q = pq + (1 - p)(1 - q) \quad (3.8)$$

La représentation stochastique d'un nombre réel $x \in [a, b]$ est une application $S : [a, b] \rightarrow [0, 1]$ qui à ce nombre réel associe une séquence de Bernoulli de paramètre $S(x)$. Deux représentations stochastiques sont principalement utilisées :

$$S_{\text{And}} : [0, 1] \rightarrow [0, 1], S_{\text{And}}(x) = x \quad (3.9)$$

$$S_{\text{Xnor}} : [-1, 1] \rightarrow [0, 1], S_{\text{Xnor}}(x) = (x + 1)/2 \quad (3.10)$$

Chacune de ces représentations permet la multiplication de nombre réels par l'utilisation d'opérations binaires bit-à-bit :

$$S_{\text{And}}(x) \text{ AND } S_{\text{And}}(y) = S_{\text{And}}(xy) \quad (3.11)$$

$$S_{\text{Xnor}}(x) \text{ XNOR } S_{\text{Xnor}}(y) = S_{\text{Xnor}}(xy) \quad (3.12)$$

Ainsi, la multiplication des séquences d'entrée du neurone par la valeur de poids associée aux connexions peut être obtenue en utilisant chacune de ces représentations. Pour produire la séquence de sortie du neurone, les bits de même indice des séquences postsynaptiques sont additionnés et le résultat est comparé à une valeur de seuil. Si le seuil est dépassé, le bit de sortie situé à cet indice sera affecté à la valeur 1. La valeur du seuil n'est pas fixe, celle-ci est associée à une distribution de probabilité. Dans le cas d'une distribution uniforme de la valeur de seuil, le neurone réagit comme s'il était soumis à une fonction d'activation linéaire. En revanche, si la distribution est concentrée en un point, la fonction d'activation du neurone sera comparable à une fonction sigmoïde. L'approche stochastique, par l'utilisation d'opérations binaires simples, est la source de réalisations matérielles basées sur des circuits logiques programmables tels que les FPGA (field-programmable gate array). Un apprentissage stochastique supervisé dérivé de l'apprentissage par rétropropagation a également été proposé. Cependant, la simplicité de cette approche est obtenue au prix d'une plus faible précision dans les calculs, comparée aux approches fréquentielles classiques.

VII. Positionnement de notre approche

Le modèle que nous allons présenter, qui est dérivé de celui proposé dans [Paquier 04], est principalement inspiré des approches fréquentielles et stochastiques. En

effet, il n'existe pas dans ce modèle de représentation explicite du temps. Nous exploitons cependant un intégrateur de potentiel à temps discret associé à un coefficient de fuite. La fonction d'activation exploitée pendant la phase de tir est pour sa part dérivée des techniques stochastiques par l'utilisation d'un seuil probabiliste et d'un flux binaire de sortie. L'apprentissage non-supervisé est rendu possible par un processus de compétition et une mise à jour des poids synaptiques dérivée de la loi de Hebb. Ces différentes caractéristiques sont le sujet des prochaines sections.

3.3 Neurones, cartes et connectivité

Nous allons dans cette section présenter notre modèle, en plaçant dans un premier temps le neurone au sein d'une structure qui jouera un rôle majeur dans notre système : la carte neuronale. Nous présenterons par la suite le modèle de neurone en tant que tel, les phases d'intégration, de tir et d'apprentissage ainsi que les mécanismes de compétition qui sont indispensables à l'apprentissage non-supervisé.

I. Cartes neuronales, invariance à la translation

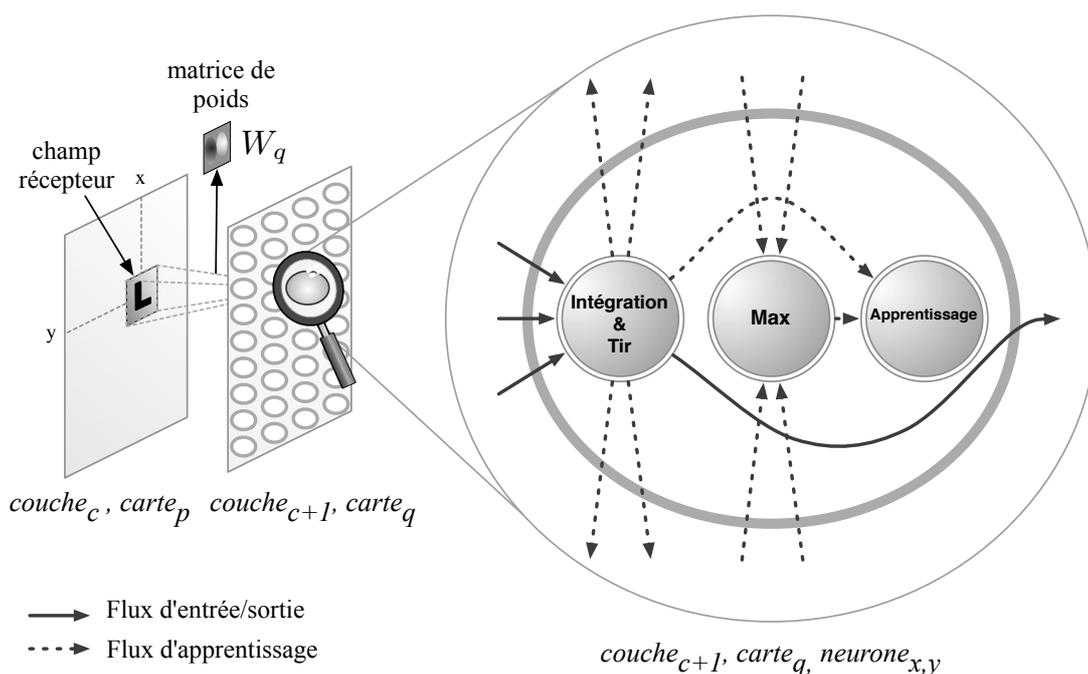


FIG. 3.9 – Schéma d'une carte neuronale. Une carte est un ensemble de neurones partageant les mêmes poids synaptiques. L'organisation des neurones dans les cartes est rétinotopique.

Nous introduisons dans cette section un concept fondamental de notre architecture : la carte neuronale. Une carte est un ensemble de neurones partageant le même ensemble

de poids synaptiques. La position d'un neurone dans sa carte correspond à la position de son champ récepteur dans le/les carte(s) afférente(s). Cette organisation qualifiée de rétinotopique est illustrée par la Figure 3.9. L'activité d'un neurone dans une carte traduit la détection au sein de/des carte(s) afférente(s) d'un motif mémorisé par la mise à jour des poids synaptiques. Ainsi, une carte neuronale permet la détection d'un motif quelle que soit sa position : c'est une reconnaissance invariante à la translation. Bien que biologiquement peu réaliste, cette technique a permis l'utilisation de réseaux de neurones multi-couches de grande taille de façon efficace pour notamment la reconnaissance de caractères manuscrits [Fukushima 03]. Comme le montre la Figure 3.9, chaque neurone est constitué de trois modules fonctionnels. Le premier module a pour but l'intégration des signaux d'entrée et la génération du signal de sortie. Le second module reçoit les signaux inhibiteurs de compétition et transmet la valeur maximale d'inhibition au troisième module d'apprentissage. Ce dernier module est à l'origine de la mise à jour des poids synaptiques dans le cas où le neurone a remporté la compétition. Ces trois fonctions seront présentées dans les prochaines sections (3.4 et 3.5).

II. Hiérarchie et compétition

Les Figures 3.10 et 3.11 illustrent les connexions reliant des cartes neuronales dans le cadre d'un réseau simple à trois couches. Chaque carte est connectée à un ensemble de cartes afférentes. Ainsi chaque neurone est associé à un ou plusieurs champs récepteurs qui définissent ses neurones afférents dans les cartes afférentes. Les poids synaptiques de chaque neurone peuvent eux-mêmes être regroupés en matrices de poids indépendantes pour chaque champ récepteur. Le but d'une telle hiérarchie de cartes est de permettre la réutilisation des informations extraites à chaque niveau (ou couche). Dans l'exemple présenté ici, le premier niveau correspond à l'image brute tandis que le deuxième niveau permet l'extraction d'orientations locales exploitée dans un troisième niveau afin de détecter la présence des lettres "L" et "T". La Figure 3.11 montre en outre la présence de signaux latéraux d'inhibitions nécessaires à l'apprentissage non-supervisé comme nous le présenterons dans la Section 3.5.

3.4 Le modèle "intègre et tire"

I. Intégration

Etant donné un neurone N_i , soit $\beta_i(t) \in \{0, 1\}$ la valeur binaire représentant la présence ou l'absence de décharge mesurée à sa sortie à l'instant t . Nous notons $\Omega_i(t)$ l'ensemble des neurones afférents au neurone N_i . Soit $w_{ij}(t) \in W_i(t), N_j \in \Omega_i(t)$ le poids synaptique associé à la connexion reliant le neurone N_j au neurone N_i ainsi que le sous-ensemble $\Omega_i^+(t)$ de $\Omega_i(t)$ définis par :

$$W_i(t) = \{w_{ij}(t), N_j \in \Omega_i\} \quad (3.13)$$

$$\Omega_i^+(t) = \{N_j \in \Omega_i, \beta_j(t) = 0\} \quad (3.14)$$

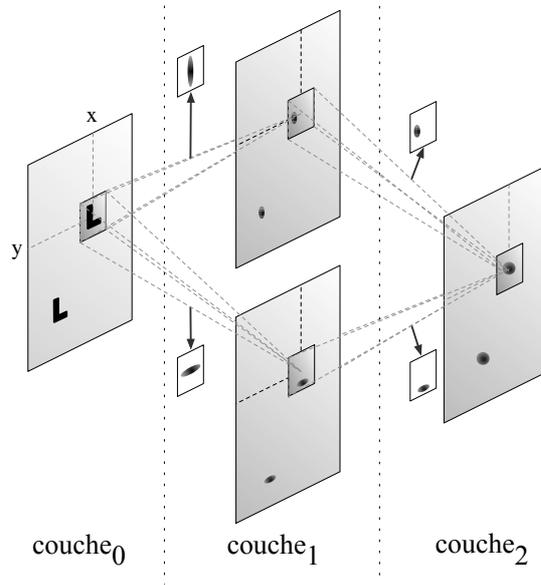


FIG. 3.10 – Hierarchie de cartes neuronales. Le signal d'entrée (couche 0), est décomposé par deux cartes d'orientation (couche 1). La carte située sur la dernière couche (couche 2) détecte la cooccurrence des deux orientations extraites dans la couche de niveau inférieur.

Les sous-ensembles $W_i^+(t)$ et $W_i^-(t)$ des poids positifs et négatifs à l'instant t sont alors :

$$W_i^+(t) = \{w_{ij}(t) > 0, N_j \in \Omega_i\} \quad (3.15)$$

$$W_i^-(t) = \{w_{ij}(t) < 0, N_j \in \Omega_i\} \quad (3.16)$$

Soient $S_i^{\beta+}(t)$, $S_i^+(t)$ et $S_i^-(t)$ respectivement la somme des poids synaptiques activés, la somme des poids synaptiques positifs et la somme des poids synaptiques négatifs à l'instant t telles que :

$$S_i^{\beta+}(t) = \sum_{N_j \in \Omega_i^+(t)} w_{ij}(t) \quad (3.17)$$

$$S_i^+(t) = \sum_{w_{ij}(t) \in W_i^+(t)} w_{ij}(t) \quad (3.18)$$

$$S_i^-(t) = \sum_{w_{ij}(t) \in W_i^-(t)} w_{ij}(t) \quad (3.19)$$

Nous appelons $P_i(t)$ le potentiel postsynaptique. Cette valeur, résultat de l'intégration, représente en un sens le niveau de reconnaissance du motif mémorisé au sein des poids synaptiques. La valeur du potentiel est donnée par la formule suivante :

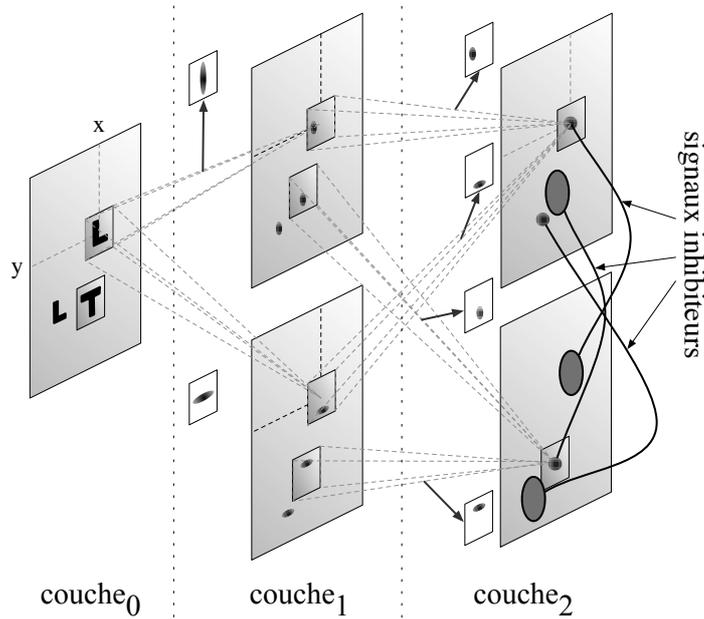


FIG. 3.11 – Compétition entre cartes neuronales. La détection d'un motif par une carte neuronale est à l'origine de signaux inhibiteurs pour bloquer l'apprentissage du même motif par les cartes de la même couche.

$$\mathcal{P}_i(t+1) = \alpha_{\mathcal{P}} \mathcal{P}_i(t) + (1 - \alpha_{\mathcal{P}}) \frac{S_i^{\beta^+}(t)}{S_i^+(t)} \quad (3.20)$$

Où $\alpha_{\mathcal{P}} \in [0, 1]$ est une valeur de fuite prédéfinie. Ainsi, le potentiel attendra sa valeur maximale si seuls les poids synaptiques positifs sont activés : cette activation correspondant au motif idéal. En présence d'une entrée stable et si les poids synaptiques demeurent constants, la valeur du potentiel atteindra à la limite une valeur comprise dans l'intervalle $[\frac{S_i^-(t)}{S_i^+(t)}, 1]$.

II. Tir

Le rôle de la phase de tir est de décider de l'état d'activation du neurone étant donné le potentiel calculé pendant la phase d'intégration. Durant cette phase, seuls les potentiels positifs sont considérés, les potentiels négatifs interdisant la décharge. La fonction d'activation de la phase de tir prend donc en entrée une valeur comprise entre 0 et 1. La détermination du niveau de potentiel "suffisant" pour générer une décharge est dans une grande mesure liée au type d'information à catégoriser. En effet, suivant que l'on souhaite autoriser une forte variation du motif d'activation autour du motif idéal, ou au contraire obtenir une catégorisation plus stricte, la fonction d'activation devra avoir des formes très différentes. Or, du fait de l'apprentissage, ce motif idéal défini par les poids synaptiques est en évolution constante et l'apprentissage, comme nous le verrons

par la suite, n'a lieu que lorsque le neurone a déchargé (apprentissage hebbien). De plus, il n'est pas trivial de connaître par avance le niveau de spécialisation nécessaire pour un neurone donné lorsque l'on ne dispose que des seuls valeurs caractérisant ce neurone. C'est pour ces raisons que nous proposons une fonction de transfert qui devra s'adapter au cours de l'apprentissage tout en analysant l'activité du neurone pour en extraire un niveau de spécialisation convenable. La sortie de chaque neurone étant binaire, leur fonction d'activation aura pour but de définir une probabilité de décharge. Ce modèle stochastique d'activation impulsionnelle partage donc avec les modèles fréquentiels le fait que la fonction d'activation traduise aussi la fréquence de décharge du neurone : une probabilité égale à 1 correspond à la fréquence maximale, celle du programme de traitement simulant l'activité des neurones.

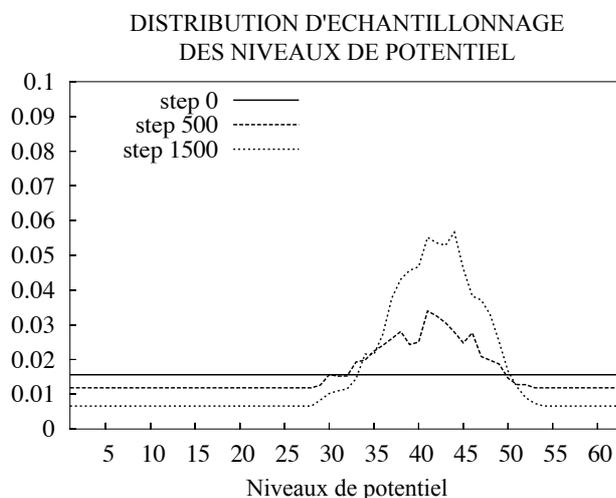


FIG. 3.12 – Distributions d'échantillonnage obtenues lors de l'apprentissage d'un neurone. 3 distributions sont représentées pour $t = 0$, $t = 500$ et $t = 1500$.

En observant les valeurs du potentiel d'un neurone particulier soumis à des décharges entrantes et dont les poids synaptiques ont amorcé leur apprentissage, il est possible de construire une distribution des niveaux de potentiel ψ_i par échantillonnage lorsque le neurone a déchargé (Figure 3.12). La valeur du potentiel est discrétisée en 64 niveaux :

$$\psi_i(k, t) \simeq P\left(\frac{k}{64} < \mathcal{P}_i(t) \leq \frac{k+1}{64} \mid \beta_i(t) = 1\right) \text{ avec } k \in [0, 63] \cap \mathbb{N} \quad (3.21)$$

Cette distribution ψ_i est mise à jour de la façon suivante :

$$\begin{aligned} & \text{si } \beta_i(t) = 1 \text{ alors } \forall k \in [0, 63] \cap \mathbb{N} : \\ \psi_i(k, t+1) & \leftarrow \begin{cases} (1 - \alpha_\psi)\psi_i(k, t) + \alpha_\psi & \text{si } \frac{k}{64} < \mathcal{P}_i(t) \leq \frac{k+1}{64}, \\ (1 - \alpha_\psi)\psi_i(k, t) & \text{sinon.} \end{cases} \end{aligned} \quad (3.22)$$

Avec α_ψ un coefficient d'apprentissage choisi dans l'intervalle $[0, 1]$. On obtient ainsi une distribution dont la forme gaussienne traduit la présence de bruit en entrée. Ce bruit

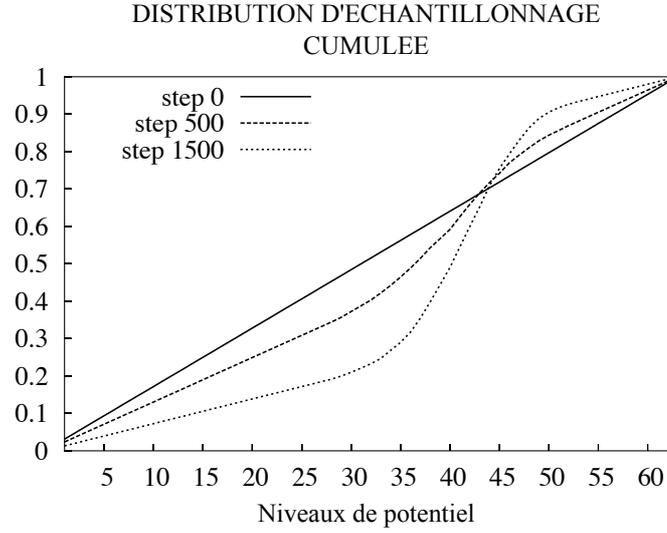


FIG. 3.13 – Distributions cumulées calculées à partir des distributions d'échantillonnage illustrées par la figure 3.12.

est dû soit directement aux capteurs soit à l'activation stochastique des neurones afférents. Dans un cas comme dans l'autre, cette distribution nous informe sur les niveaux de potentiel pour lesquels l'activation, donc l'apprentissage est nécessaire. En tant que première approche, il serait possible de placer un seuil "au pied de la gaussienne" en construisant une fonction de transfert de type fonction-seuil permettant l'activation à partir de cette valeur. Cette approche simpliste, bien qu'efficace durant les premiers pas d'apprentissage entraîne rapidement un sur-apprentissage. En effet, au cours de l'apprentissage, la somme des poids positifs S_i^+ augmente or, selon l'équation 3.20 une plus grande ressemblance au motif idéal $S_i^{\beta^+}$ est nécessaire afin d'obtenir une même valeur de potentiel. Le neurone aura donc tendance à décharger de moins en moins, tout en apprenant de plus en plus : c'est le sur-apprentissage. Cette première approche montre la nécessité d'une catégorisation plus souple pour contrôler la spécialisation.

Nous proposons d'utiliser une valeur de seuil \mathcal{T}_i régie par une distribution de probabilité $\phi(t)$ évoluant au cours du temps. Comment construire cette fonction ϕ en prenant en compte l'histoire des activations du neurone ? Nous souhaitons en effet que celui-ci conserve une trace du niveau de spécialisation courant de sorte que :

$$P(\beta_i(t) = 1 | \mathcal{P}_i(t) = x) = P(\mathcal{P}_i(t) \leq x | \beta_i(t) = 1) \quad (3.23)$$

or, étant donnée un niveau de potentiel x , le tir n'a lieu que si $\mathcal{T}_i \leq x$:

$$P(\beta_i(t) = 1 | \mathcal{P}_i(t) = x) = P(\mathcal{T}_i(t) \leq x) \quad (3.24)$$

on obtient alors l'équation suivante :

$$P(\mathcal{T}_i(t) \leq x) = P(\mathcal{P}_i(t) \leq x | \beta_i(t) = 1) \quad (3.25)$$

on en déduit donc que :

$$P(\mathcal{T}_i(t) = x) = P(\mathcal{P}_i(t) = x | \beta_i(t) = 1) \quad (3.26)$$

nous utiliserons en pratique la distribution $\psi_i(k, t)$ construite par échantillonnage de sorte que :

$$P\left(\frac{k}{64} < \mathcal{T}_i(t) \leq \frac{k+1}{64}\right) \simeq \psi_i(k, t), k \in [0, 63] \cap \mathbb{N} \quad (3.27)$$

La probabilité de décharge à l'instant t est donc obtenue de la façon suivante :

$$P\left(\beta_i(t) = 1 \mid \frac{k}{64} < \mathcal{P}_i(t) \leq \frac{k+1}{64}\right) \simeq \sum_{l=0}^k \psi_i(l, t) \quad (3.28)$$

De façon pratique, étant donné un niveau de potentiel, nous effectuons un tirage aléatoire de répartition uniforme dans l'intervalle $[0, 1]$. La valeur binaire de sortie du neurone N_i à l'instant t , notée $o_i(t)$, vaut donc 1 si le tirage est inférieur à la probabilité donnée par l'équation 3.28, 0 sinon.

3.5 Apprentissage et compétition

I. La règle de Hebb

D'après Donald O. Hebb [Hebb 49], les poids synaptiques sont renforcés lorsque les activités présynaptiques et postsynaptiques coïncident durant une courte période de temps. Cette règle d'apprentissage est d'un grand intérêt pour l'apprentissage non-supervisé par le fait que seules les coïncidences sont à l'origine de l'apprentissage. Il est à présent admis que ce mode d'apprentissage joue un rôle important dans de nombreuses fonctions cognitives et sensori-motrices telles que la mémoire [Bliss 93], la vision [Fregnac 88] [Soodak 91] et les fonctions motrices [Ito 93] mais aussi dans certaines formes de développement du système nerveux [Yuille 89]. Cette règle montre néanmoins ses limites car elle provoque notamment l'instabilité des poids lorsque les activités présynaptiques et postsynaptiques sont totalement non-corrélées. Pour palier ce type de limitation, on remplace l'apprentissage de la cooccurrence par celui de la covariance, cet aspect sera abordé dans les prochaines lignes.

II. La compétition par inhibitions latérales

La compétition entre neurones intervient dès lors que des motifs définis dans les mêmes domaines doivent être appris. Les neurones en compétition ont donc en commun les mêmes neurones afférents. Ils sont en outre connectés les uns aux autres par le biais de connexions inhibitrices latérales. A la différence des connexions excitatrices, ces connexions véhiculent directement le niveau de potentiel et non l'information binaire de décharge. Ces connexions sont dites "latérales" car leurs directions sont orthogonales à celles des connexions excitatrices. La compétition consiste à comparer le potentiel de

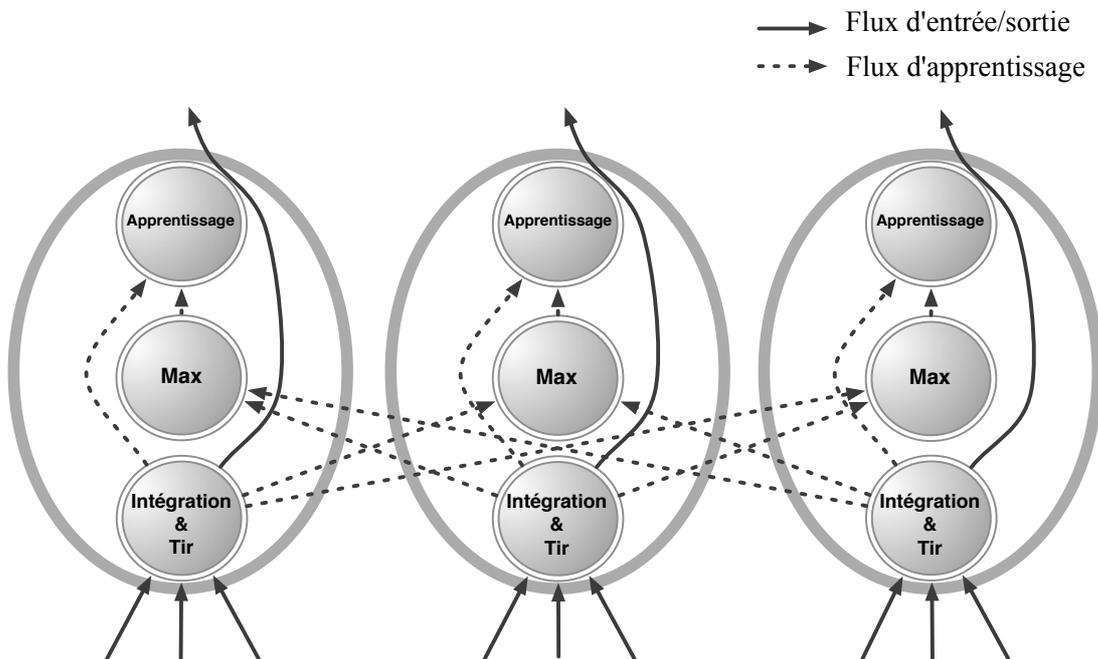


FIG. 3.14 – Ce schéma met en évidence les connexions inhibitrices existant entre trois neurones appartenant à trois cartes neuronales distinctes en compétition. Ces neurones sont situés en des coordonnées identiques dans leurs cartes respectives. Le module "Max" de chaque neurone fournit au module d'apprentissage la valeur maximale de potentiel mesurée parmi l'ensemble des neurones inhibiteurs. L'apprentissage a lieu lorsque la valeur du potentiel local est supérieure à celles des potentiels distants.

chaque neurone à ceux des neurones inhibiteurs. Ainsi, seul le neurone dont le niveau de reconnaissance est le plus élevé pourra apprendre. Ce processus est illustré par la Figure 3.14. Dans cette figure sont représentés trois neurones interconnectés par des connexions inhibitrices. La cellule "Max" correspond au comparateur de chaque neurone. Celui-ci reçoit de chaque neurone inhibiteur sa valeur de potentiel et transmet à la cellule d'"Apprentissage" la valeur maximale reçue. Le processus d'apprentissage est alors en mesure de comparer le potentiel local à la valeur maximale d'inhibition et déclencher l'apprentissage si le neurone a gagné la compétition. Ce processus est qualifié de "Winner Take All" car seul le neurone ayant gagné la compétition est autorisé à apprendre le motif à l'origine de l'intégration. Ce mécanisme exclut donc toute notion d'apprentissage induit par voisinage qui est utilisé dans les cadre des carte auto-organisatrices (SOM) de Kohonen [Kohonen 89]. Du point de vue des cartes neuronales (Figure 3.11), la compétition intervient entre des neurones appartenant à des cartes différentes mais dont les positions au sein de leur carte sont les mêmes. La localisation de la compétition permet ainsi d'apprendre des motifs présents, au même instant mais à des positions différentes, dans la carte d'entrée. Nous notons Ω_i^\ominus , l'ensemble des neurones en compétition avec le neurone N_i .

III. Compétition à l'activation

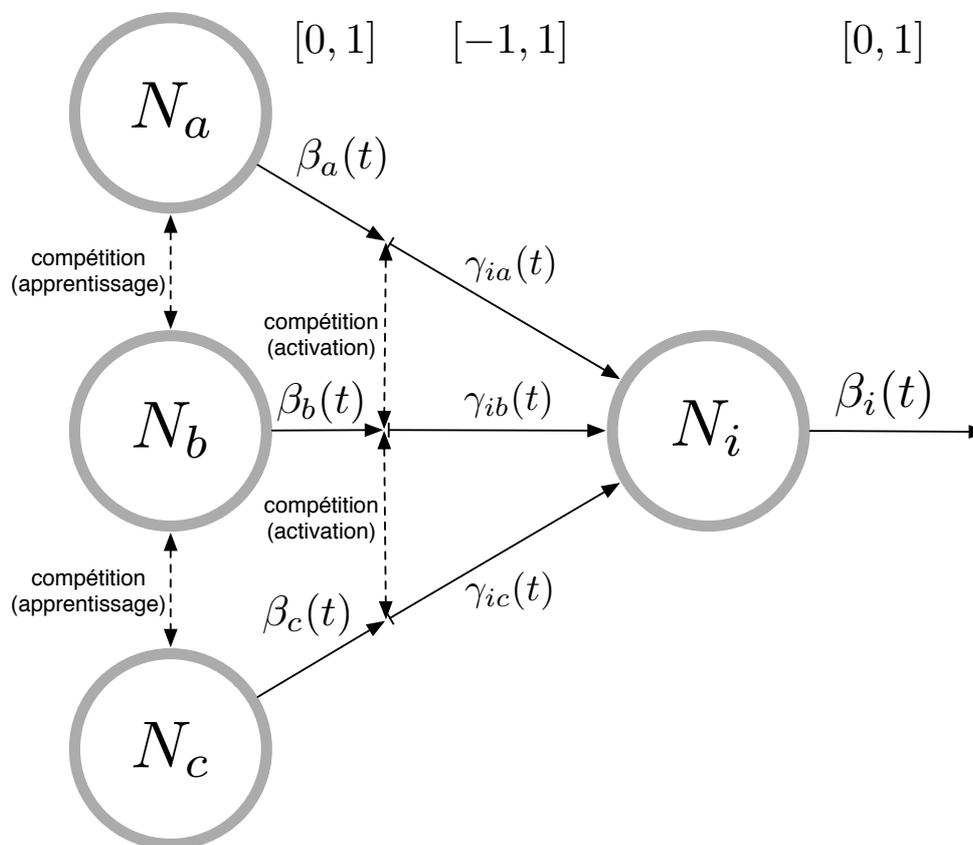


FIG. 3.15 – Le processus de compétition à l'activation permet de convertir les valeurs $\beta_j(t) \in [0, 1]$ en valeurs $\gamma_j(t) \in [-1, 1]$. Ces dernières seront les cibles de l'apprentissage hebbien.

	$\beta_c(t) = 0$		$\beta_c(t) = 1$	
	$\beta_b(t) = 0$	$\beta_b(t) = 1$	$\beta_b(t) = 0$	$\beta_b(t) = 1$
$\beta_a(t) = 0$	$\gamma_{ia}(t) = 0$	$\gamma_{ia}(t) = -1$	$\gamma_{ia}(t) = -1$	$\gamma_{ia}(t) = -1$
	$\gamma_{ib}(t) = 0$	$\gamma_{ib}(t) = 1$	$\gamma_{ib}(t) = -1$	$\gamma_{ib}(t) = 1$
	$\gamma_{ic}(t) = 0$	$\gamma_{ic}(t) = -1$	$\gamma_{ic}(t) = 1$	$\gamma_{ic}(t) = 1$
$\beta_a(t) = 1$	$\gamma_{ia}(t) = 1$	$\gamma_{ia}(t) = 1$	$\gamma_{ia}(t) = 1$	$\gamma_{ia}(t) = 1$
	$\gamma_{ib}(t) = -1$	$\gamma_{ib}(t) = 1$	$\gamma_{ib}(t) = -1$	$\gamma_{ib}(t) = 1$
	$\gamma_{ic}(t) = -1$	$\gamma_{ic}(t) = -1$	$\gamma_{ic}(t) = 1$	$\gamma_{ic}(t) = 1$

TAB. 3.2 – Tableau des valeurs γ_{ij} pour 3 neurones afférents en compétition (Figure 3.15).

L'information d'activation transmise par les neurones est binaire et la probabilité

de tir traduit la détection plus ou moins certaine d'un motif. Lorsque une valeur binaire de décharge est transmise à un neurone, celle-ci a la capacité d'exciter ou au contraire d'inhiber le neurone cible en fonction de la valeur de poids associée à la connexion (valeur positive ou négative respectivement). Le rôle de la compétition à l'activation est étroitement lié à l'apprentissage de poids synaptiques négatifs. La présence d'un poids négatif signifie que la composante d'entrée, associée à la connexion, est absente du motif idéal. Mais quelle est alors la différence entre un poids nul, qui peut être comparé à l'absence de connexion et un poids négatif? Dans le premier cas, la présence de la composante d'entrée n'a aucune influence sur la détection du motif appris. Or, dans le cas d'un poids négatif, la composante associée doit être absente pour atteindre une probabilité maximale d'activation. Si les composantes d'entrée sont issues d'une compétition dans leur couche, la présence d'une composante signifie également l'absence d'autres composantes en compétition avec celle-ci (étant donnée la règle "Winner-take-all" ayant lieu lors de l'apprentissage). A l'inverse, si une composante est absente en entrée mais qu'à cette même position une autre composante s'active, alors l'absence de la première devra être à l'origine de l'apprentissage d'un poids synaptique négatif (cette composante doit être absente du motif). Enfin, si toutes les composantes d'entrée sont inactives en une même coordonnée, alors les poids associés à leurs connexions respectives devront converger vers 0 (ces connexions n'auront à terme pas d'influence sur la détection du motif). Si un neurone N_i est connecté à des neurones en compétition (N_a, N_b, N_c dans la Figure 3.14), ces neurones ayant donc les mêmes positions sur des cartes différentes, alors les valeurs d'activation perçues par le neurone N_i seront données par :

$$\forall N_j \in \Omega_i, \gamma_{ij}(t) = \begin{cases} 1 & \text{si } \beta_j(t) = 1 \\ -1 & \text{si } \beta_j(t) = 0 \text{ et } \exists N_k \in \Omega_i \cap \Omega_j^\ominus, \beta_k(t) = 1 \\ 0 & \text{sinon.} \end{cases} \quad (3.29)$$

La Figure 3.15 et le Tableau 3.2 illustrent le principe de compétition à l'activation générant le signe des activations $\gamma_{ij}(t)$ mesurées à l'entrée du neurone N_i .

En revanche si les cartes afférentes d'un neurone N_i ne sont pas en compétition, par exemple dans le cas de cartes issues de modalités perceptives différentes, le calcul de $\gamma_{ij}(t)$ devient :

$$\forall N_j \in \Omega_i, \gamma_{ij}(t) = \begin{cases} 1 & \text{si } \beta_j(t) = 1 \\ -1 & \text{sinon.} \end{cases} \quad (3.30)$$

IV. Loi d'apprentissage

L'apprentissage a lieu lorsque un neurone s'est activé et qu'il a remporté la compétition : si $\beta_i(t) = 1$ et $\mathcal{P}_i(t) > \mathcal{P}_j(t), \forall N_j \in \Omega_i^\ominus$. Lorsque ces deux conditions sont respectées, le processus d'apprentissage met à jour trois valeurs notées μ_{ij} , σ_{ij} et w_{ij} , respectivement la moyenne, l'écart-type des valeurs d'activation et le poids synaptique

de la connexion entre le neurone N_j au neurone N_i tels que :

$$\mu_{ij}(t+1) = (1 - \alpha_W)\mu_{ij}(t) + \alpha_W |\gamma_{ij}(t)| \quad (3.31)$$

$$\sigma_{ij}(t+1)^2 = (1 - \alpha_W)\sigma_{ij}(t) + \alpha_W [\mu_{ij}(t+1) - \gamma_{ij}(t+1)]^2 \quad (3.32)$$

$$w_{ij}(t+1) = (1 - \alpha_W)w_{ij}(t) + \alpha_W [1 - 2\sigma_{ij}(t+1)]\gamma_{ij}(t+1) \quad (3.33)$$

avec $\alpha_W \in [0, 1]$ un taux d'apprentissage fixé. Cette mise à jour des poids synaptiques permet de filtrer les entrées dont l'activité présente un écart-type élevé et donc d'éviter l'instabilité de la valeur des poids observée lors d'un apprentissage hebbien classique (équation 3.31). Afin de stabiliser l'apprentissage nous ajoutons au taux d'apprentissage un coefficient de fuite, fixé expérimentalement à une valeur proche de 1 (0.99 pour la plupart des applications).

V. Matrices de poids adaptatives

Les matrices de poids, à savoir les sous-ensembles de poids associés aux connexions entrantes provenant d'une même carte afférente ont été jusqu'à présent supposées de taille fixe. Cette limitation constitue un obstacle important à l'apprentissage ouvert de motifs perceptifs comme la reconnaissance de motifs visuels. Nous proposons une méthode simple permettant d'adapter la taille des champs récepteurs du neurone tout en conservant leur forme rectangulaire d'origine. Cette première approche adaptative sera bien sûr réalisée en ligne en analysant des données locales : la valeur des poids synaptiques. Nous avons défini jusqu'à présent l'apprentissage neuronal comme la détection d'invariants locaux. Les motifs visuels que nous souhaitons apprendre sont en effet localisés et forment des zones d'activité connexes. Partant de ce constat, nous allons étudier les conditions d'expansion ou de rétraction des matrices de poids lors de l'apprentissage. La taille des champs récepteurs doit être identique pour chaque carte afférente et cela pour deux raisons. Tout d'abord, il est impératif de maintenir la compétition lors de l'activation pour chaque neurone afférent. D'autre part, l'expansion d'un champ récepteur doit pouvoir profiter à toute carte afférente afin d'enrichir le motif courant.

Pour pouvoir décider de l'expansion ou de la rétraction d'un champ récepteur, nous fixons un seuil pour la valeur absolue des poids synaptiques au delà duquel un poids sera considéré comme intéressant pour la détection du motif mémorisé. De plus, nous attribuons à chaque poids une valeur entière comptabilisant le nombre de voisins ayant dépassés ce seuil. Ainsi, en observant les poids situés au bord de leur matrice nous décidons de l'expansion ou de la rétraction de la façon suivante : l'expansion a lieu sur un bord des matrices si un poids dépasse le seuil d'intérêt sur ce bord sur une des matrices (Figure 3.16) ; la rétraction a lieu sur un bord des matrices si tous les compteurs de voisinage sont égaux à 0 pour tous les poids situés sur ce bord sur toutes les matrices (Figure 3.17). Cette méthode exploite donc la stabilité du motif lors de l'expansion et l'instabilité des zones situées au bord de l'objet (sur l'image) lors de la rétraction. Dans le cadre de l'approche écologique de la perception visuelle de Gibson et lors du déplacement d'un objet, ces zones sont nommées des zones de délétion (lorsque l'image de fond passe derrière l'objet) ou d'accrétion (lorsque l'image de fond est découverte

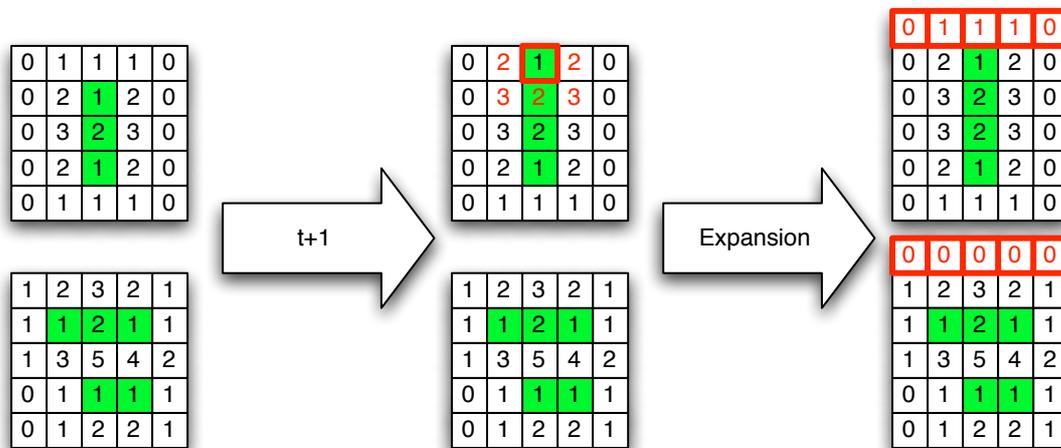


FIG. 3.16 – Expansion des matrices de poids. Chaque case correspond à un poids où est inscrit la valeur du compteur de voisinage. Les cases colorées en vert correspondent aux poids ayant dépassé le seuil d'intérêt. A gauche : matrices de poids initiales associées à deux cartes afférentes. Au centre : lorsqu'une valeur de poids dépasse le seuil d'intérêt, les valeurs de voisinage sont mises à jour. A droite : expansion des matrices de poids due à la présence sur le bord d'une matrice d'un poids ayant dépassé le seuil d'intérêt.

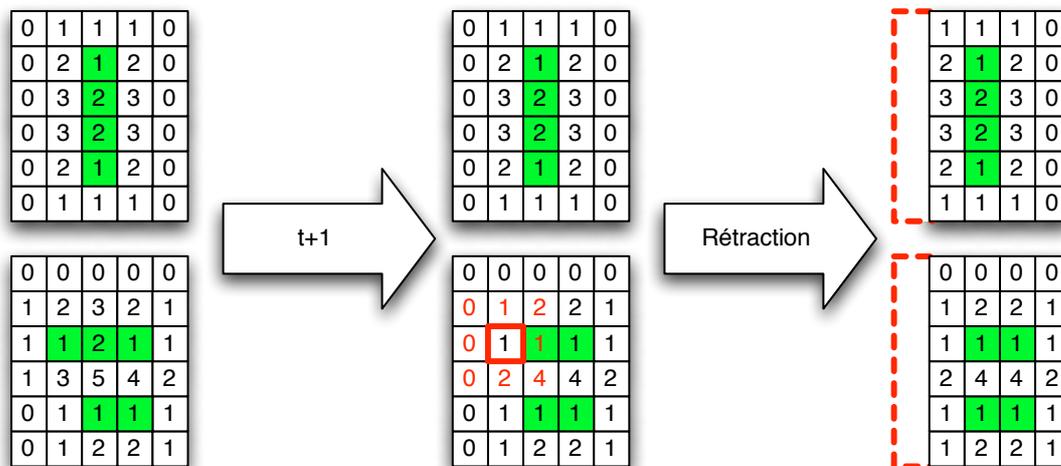


FIG. 3.17 – Rétraction des matrices de poids. A gauche : matrices de poids initiales associées à deux cartes afférentes. Au centre : lorsque la valeur d'un poids devient inférieure au seuil d'intérêt, les valeurs de voisinage sont modifiées. A droite : la rétraction des matrices de poids a lieu lorsqu'un bord contient dans toutes les matrices des valeurs de voisinage égales à 0 (ici le bord de gauche).

lors du déplacement de l'objet). Cette méthode de redimensionnement des matrices de poids trouve son efficacité maximale lorsqu'elle est appliquée à des images acquises par

stéréovision où les flux d'images gauches et droites sont entrelacés. Par cette méthode, qui simule le déplacement rapide de l'observateur, les zones d'accrétion et de déléition apparaissent distinctement sans recourir au déplacement de l'objet. L'utilisation de cette méthode est illustrée dans les prochaines sections.

3.6 Les 3 voies de la perception

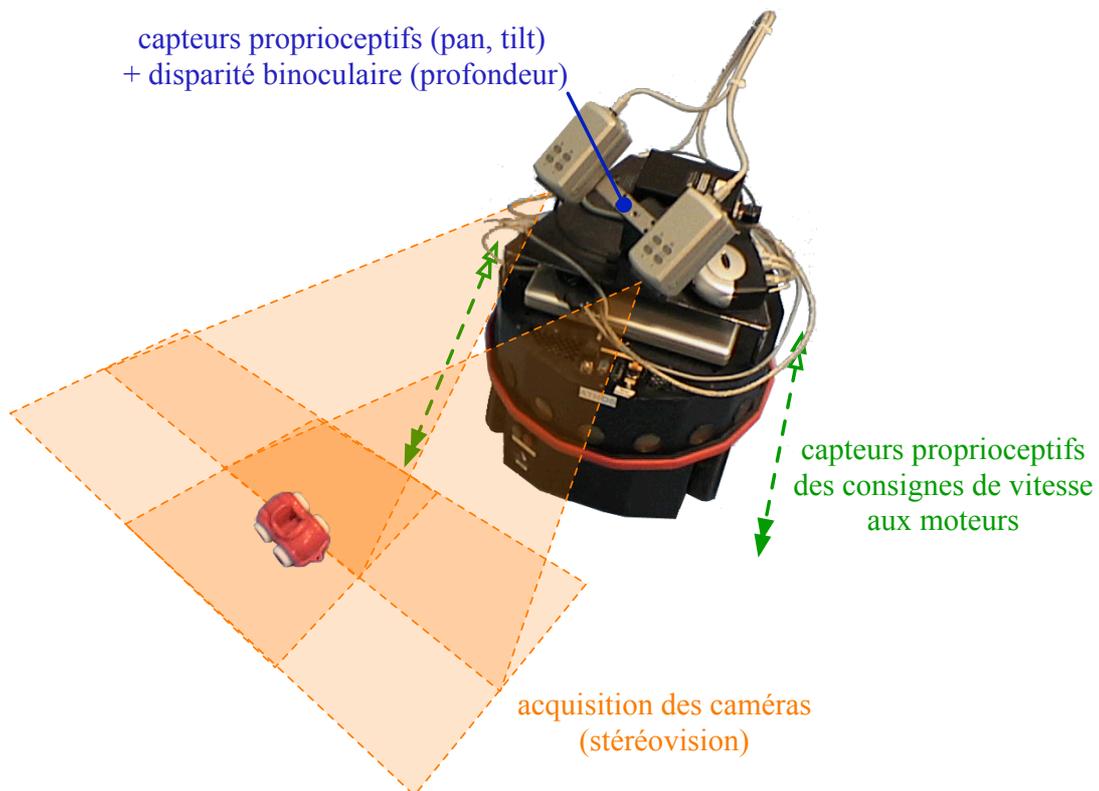


FIG. 3.18 – Plateforme expérimentale. Notre robot dispose de trois types de capteurs : deux caméras, des capteurs proprioceptifs mesurant la position des caméras et la distance à l'objet, des capteurs proprioceptifs mesurant les consignes de vitesse données aux moteurs.

Notre plate-forme expérimentale est un robot à deux roues motrices différentielles de type SuperScout 2 équipé de deux caméras montées sur une platine motorisée (Figure 3.18). Nous souhaitons étudier l'apprentissage en ligne de représentations sensori-motrices lors de l'interaction entre le robot et un objet présent dans son environnement. L'objet considéré sera supposé de taille réduite afin qu'il puisse être entièrement perçu dans une seule image acquise par les caméras. Le comportement du robot devra être autonome. L'exécution du programme de traitement ne sera pas uniquement située dans l'unité de calcul embarquée mais distribuée parmi différentes machines appartenant au réseau local et ce, afin de permettre une simulation en temps réel du réseau de

neurones.

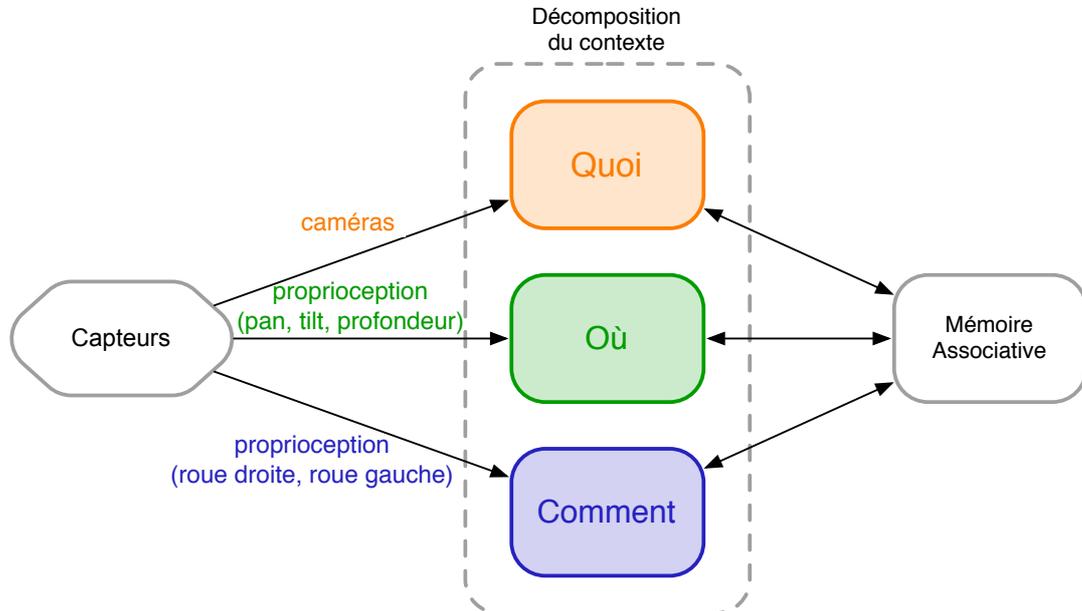


FIG. 3.19 – La phase de décomposition du contexte perceptif produit trois groupes de représentations. 'Quoi' : représentations visuelles des objets. "Où" : représentations de la position des objets. "Comment" : représentations des actions du robot.

Lors de la décomposition du contexte perceptif, nous distinguons trois classes de représentations (Figure 3.19) :

- 'Quoi' : classe des représentations visuelles des objets
- 'Où' : classe des représentations de la position des objets
- 'Comment' : classe des représentations des actions du robot

Ces trois classes de représentation sont caractérisées par l'utilisation de données issues de capteurs spécifiques :

- 'Quoi' : caméras (stéréovision)
- 'Où' : capteurs proprioceptifs mesurant la position de la platine et la disparité binoculaire
- 'Comment' : capteurs proprioceptifs mesurant les consignes de vitesse des moteurs

Cette décomposition nous permettra d'identifier et de construire de façon systématique des associations sensori-motrices dans la mémoire associative.

3.7 La représentation des objets

I. Représentation par vues, structure du réseau

Nous avons fait le choix d'une représentation des objets par vues, car celle-ci est la mieux adaptée aux capacités d'apprentissage des neurones. Nous associons à chaque vue une carte neuronale dont la matrice d'apprentissage sera redimensionnée suivant la

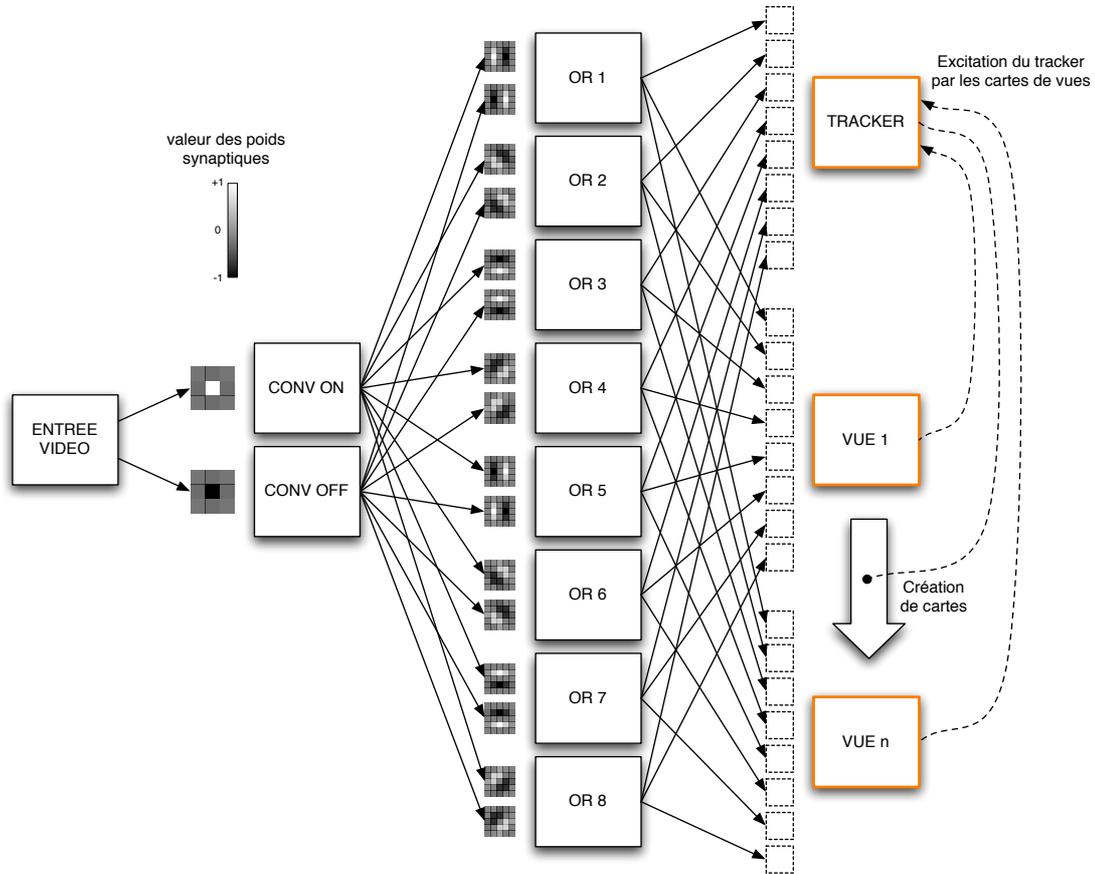


FIG. 3.20 – Schéma d'un réseau utilisé pour l'apprentissage de vues d'objet. Les connexions entre cartes sont représentées par les flèches en trait plein et les interactions entre le tracker et les cartes de vues par des flèches en pointillés.

taille de l'objet. Les cartes neuronales ne sont pas directement connectées aux données brutes provenant des caméras. Il est nécessaire de convertir l'information de luminance de l'image composée d'un ensemble de valeurs réelles comprises entre 0 et 1, en une information compatible avec le processus d'intégration efférent. Cette fonction est remplie par des cartes particulières appelées "convertisseurs". La fonction de ces cartes peut être comparée à celle des cellules de la rétine : l'information de luminance est traitée pour en extraire des contrastes locaux. Nous utilisons deux cartes de conversion afin d'extraire des contrastes locaux "positifs" (pixel lumineux sur zone sombre) et "négatifs" (pixel sombre sur zone lumineuse). La phase de tir est réalisée en utilisant une fonction de transfert sigmoïde et une activation stochastique semblable à celle des neurones classiques. La sortie des cartes de conversion est donc binaire et exploitable pour le traitement des neurones efférents. Une deuxième couche de cartes est alors utilisée pour extraire des orientations locales de l'image. Bien que le nombre de cartes d'orientation ne soit pas limité par cette architecture, huit cartes donnent expérimentalement

des résultats satisfaisants. Chaque carte d'orientation est connectée aux deux cartes de conversion par le biais de deux matrices de poids semblables à des filtres de Gabor. L'activité d'un neurone dans chaque carte représente donc la position d'une orientation particulière à cette position. Le troisième niveau est destiné à l'apprentissage. Dans ce niveau, les cartes sont connectées à chaque carte d'orientation. L'apprentissage d'un motif dans ces cartes correspond donc à la détection de huit motifs d'orientation cooccurrents. Les matrices de poids de ces cartes de vue sont initialement de taille faible : 15x15 pour une image de taille 80x60 pixels. Les valeurs des poids synaptiques sont tirées aléatoirement suivant une distribution gaussienne centrée : les poids situés au centre de la matrice auront des valeurs plus élevées que les poids situés sur son contour. Ainsi nous assurons que chaque carte sera en mesure d'amorcer son apprentissage sur la zone des cartes afférentes produisant le maximum de décharges (zone saillante).

II. Association temporelle et spatiale de vues, invariance à l'objet

La structure en réseau hiérarchique offre la possibilité d'apprendre une vue particulière d'un objet. L'association temporelle et spatiale est un moyen d'apprendre de façon incrémentale ses différentes vues [Wallis 96a, Wallis 96b]. Pour suivre la position d'un objet mais aussi construire une représentation vue-indépendante nous utilisons une carte d'apprentissage particulière : le tracker. Un tracker est une carte dont le coefficient d'apprentissage est élevé et constant ($\alpha_W = 0.2$) lui procurant ainsi une mémoire à court terme. Le motif appris sera en effet mis à jour en grande partie, à chaque pas de temps du simulateur. Un tracker est utilisé pour chaque objet et connecté à chaque carte de vue par une connexion latérale excitatrice. Lorsqu'une carte de vue est activée le tracker est excité à la position de cette activation, entraînant l'apprentissage du motif situé à cette position, à savoir une vue particulière de l'objet. La fonction des connexions excitatrices latérales peut être comparée à celle de l'opérateur MAX proposé par Riesenhuber et Poggio dans le modèle hiérarchique de reconnaissance d'objet HMAX [Riesenhuber 99]. Les mémoires à long terme des cartes de vue sont donc "prioritaires" sur la mémoire à court-terme lors de la reconnaissance. Il est cependant possible que le tracker ne reçoive aucune impulsion excitatrice latérale si une vue inconnue de l'objet est présentée en entrée, lors de la rotation de l'objet par exemple. C'est alors que la plasticité de la mémoire à court terme trouve tout son intérêt. Lors de la rotation de l'objet, à partir d'une vue apprise et à la différence d'une mémoire à long terme, le tracker peut suivre la "déformation" du motif : la mémoire à court terme exploite la cohérence temporelle et spatiale de cette déformation. Si le tracker ne reçoit aucun signal d'excitation latéral (le tracker est la seule carte active), durant une période fixée à 30 pas de temps (entre 1 et 2 secondes), alors une nouvelle carte de vue est créée, connectée et excitée à la position d'intérêt (la position de la nouvelle vue dans l'image) pour amorcer son apprentissage.

Cette méthode permet de construire de façon incrémentale un ensemble de cartes neuronales spécialisées pour la détection de vues spécifiques d'un objet. Le tracker a la double fonction de fournir une détection vue-invariante et de recruter de nouvelles cartes lorsqu'une nouvelle vue est présentée. Le nombre de cartes de vues nécessaires

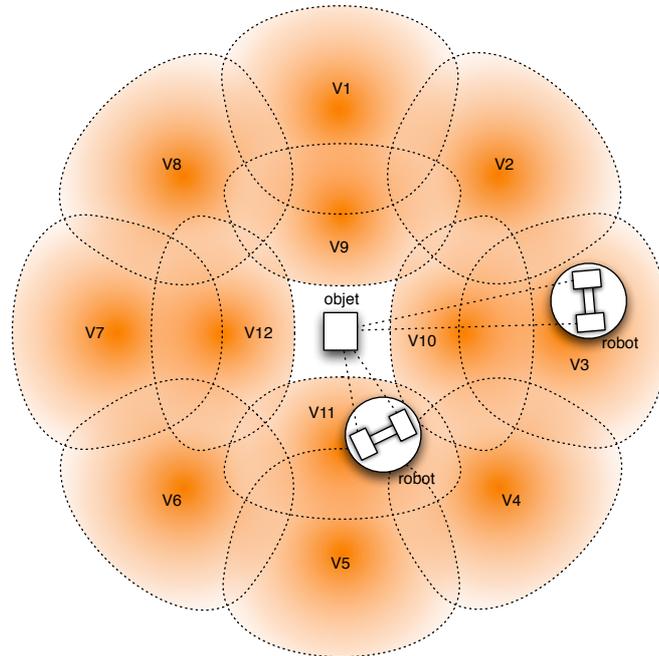


FIG. 3.21 – Schéma des champs de vue d'un objet.

à l'apprentissage d'un objet est très dépendante du champ de vue de chacune de ces cartes. Le champ de vue est défini par la zone de l'espace dans laquelle le robot est capable de reconnaître une vue particulière (Figure 3.21). La taille de cette zone dépend de plusieurs facteurs comme les symétries de l'objet ou le nombre d'orientations locales utilisées pour l'apprentissage. En effet, plus le nombre de cartes d'orientations est grand plus le motif appris par la carte de vue sera contraint ce qui réduira la taille du champ de vue correspondant. Il est important de noter que selon les symétries de l'objet, certains champs de vue peuvent être représentés comme l'union de zones non connexes. C'est le cas par exemple lors de l'apprentissage d'objets ayant deux faces visuellement identiques et opposées. Le chevauchement des champs de vue voisins semble en contradiction avec le mécanisme "Winner-Take-All" de l'apprentissage. Ce chevauchement s'explique par le fait que la compétition n'intervient que pour sélectionner la carte qui sera autorisée à apprendre, alors que lors de la phase de tir toute carte ayant accumulé un potentiel suffisant pourra s'activer.

3.8 La représentation des positions et des actions

I. Cartes temporisées

Afin de présenter l'apprentissage des motifs de position et d'action nous introduisons la notion de cartes neuronales temporisées (Figures 3.24, 3.25 et 3.26), ce type de représentation étant largement exploité par la suite. Ces cartes sont constituées de neurones organisés en tableaux à deux dimensions. L'axe vertical est utilisé pour la mesure d'une

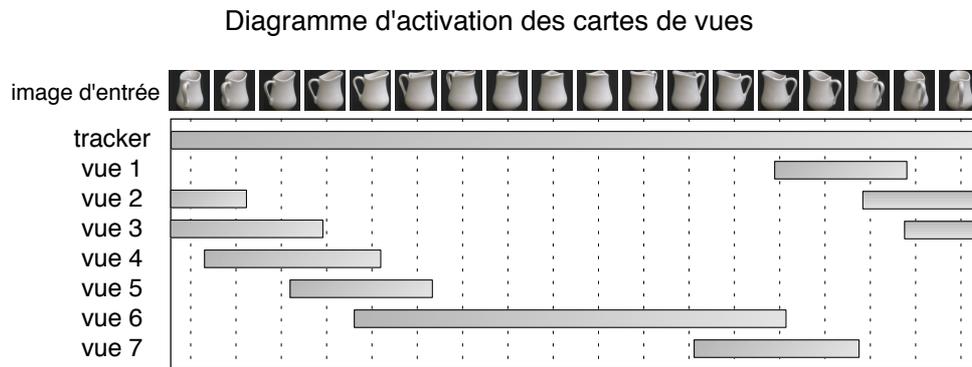


FIG. 3.22 – Diagramme des activations des cartes de vues mesurées après un apprentissage non supervisé lors de la rotation d'un objet. Pour chaque carte, le domaine d'activation est représenté par un rectangle gris.

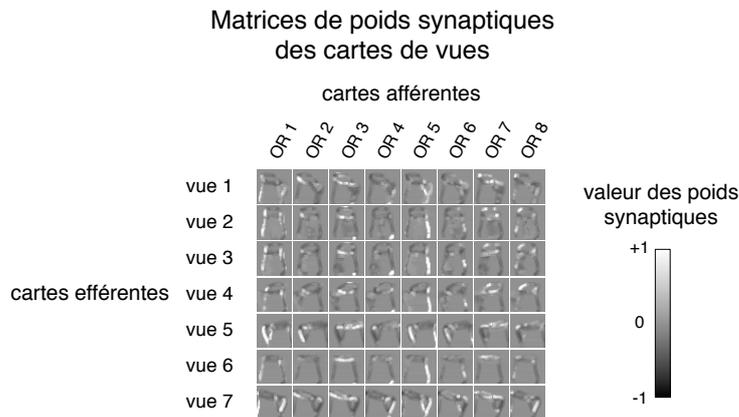


FIG. 3.23 – Matrices de poids des cartes de vues obtenues après apprentissage. Se référer à la figure 3.20 pour le schéma des connexions.

valeur perceptive particulière comme la position d'un degré de liberté, ou la vitesse de rotation d'un moteur. L'axe horizontal permet quant à lui de représenter l'évolution de cette valeur au cours du temps. La propagation des activations s'effectue de gauche à droite et la valeur mesurée au temps présent est représentée par la colonne de neurones située à l'extrême gauche de chaque carte. Les cartes temporisées ont été utilisées pour apprendre des séquences d'activation comme pour la reconnaissance de phonème à base de spectrogrammes présentée par W. Paquier [Paquier 04]. Notre but est ici d'apprendre des motifs de transitions impliquant les trois classes de représentations présentées plus haut : "Quoi", "Où" et "Comment". Un tel apprentissage ne nécessite donc pas d'enregistrer l'historique complet des activations et seul l'état précédent semblerait suffire. L'algorithme d'apprentissage présenté plus haut nous contraint cependant à maintenir les motifs perceptifs durant un temps suffisant afin de permettre aux poids et surtout à la fonction de transfert de se stabiliser. La propagation temporelle des activations n'est

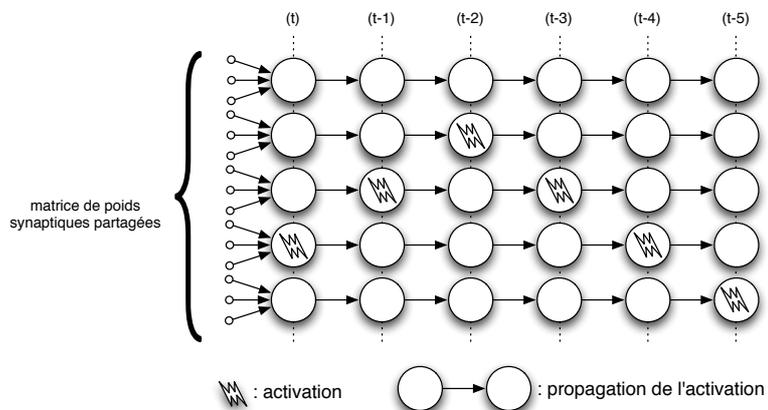


FIG. 3.24 – Propagation des activation dans une carte de conversion temporisée.

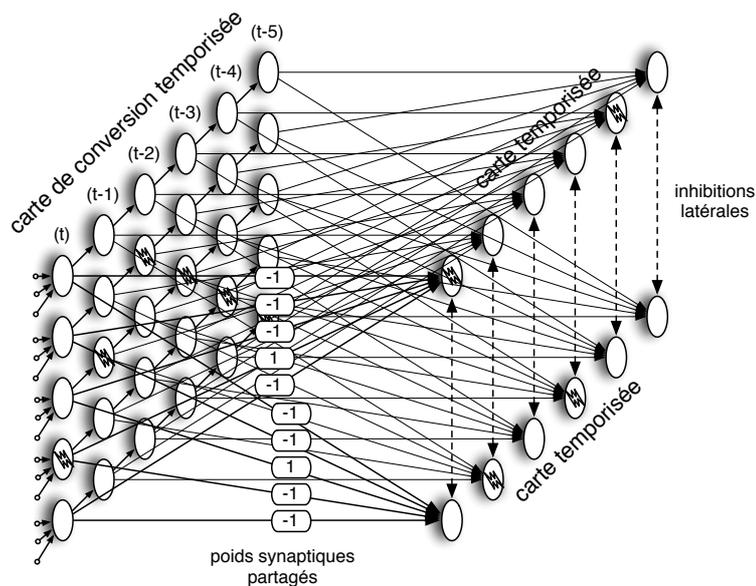


FIG. 3.25 – Schéma de connexions d'une carte de conversion temporisée vers deux cartes en compétition.

nécessaire que pour les cartes de conversion car l'effet de propagation est diffusé au travers du réseau par le schéma de connexion. Nous introduisons dans la figure 3.26 la représentation simplifiée d'un réseau de cartes temporisées, c'est cette représentation qui sera utilisée par la suite.

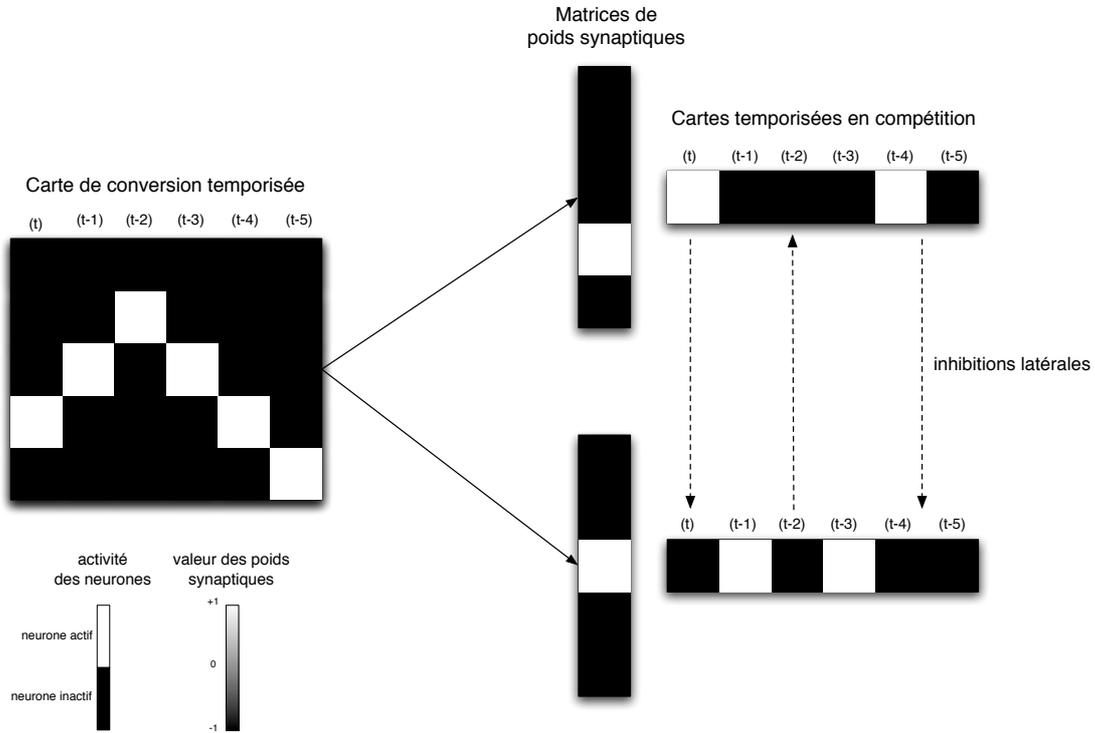


FIG. 3.26 – Schéma de connexions simplifié d’une carte de conversion temporisée vers deux cartes en compétition. Ce schéma est équivalent à celui de la Figure 3.24.

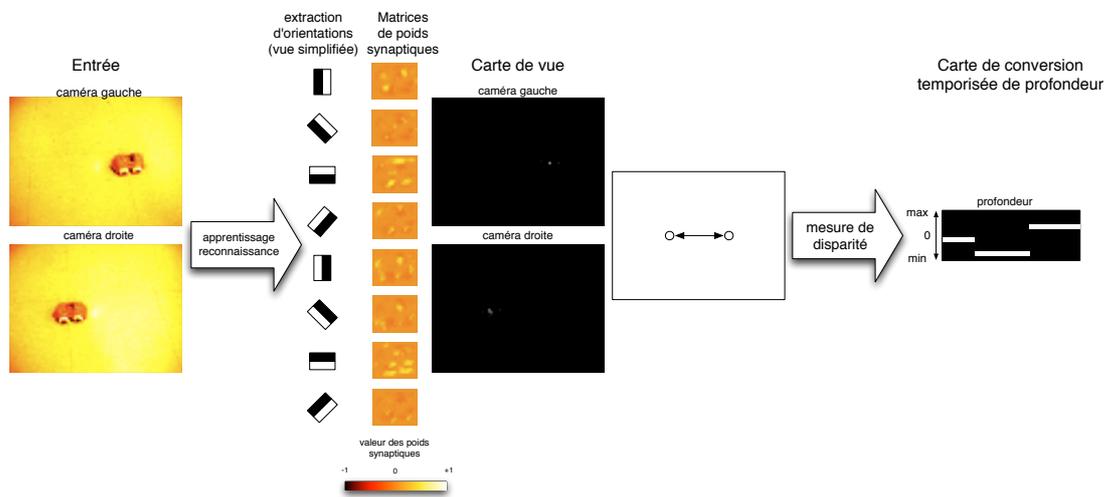


FIG. 3.27 – Mesure de disparité binoculaire et carte de conversion de profondeur. La carte de conversion temporisée de profondeur fournit une mesure de disparité binoculaire lorsqu’un objet est détecté.

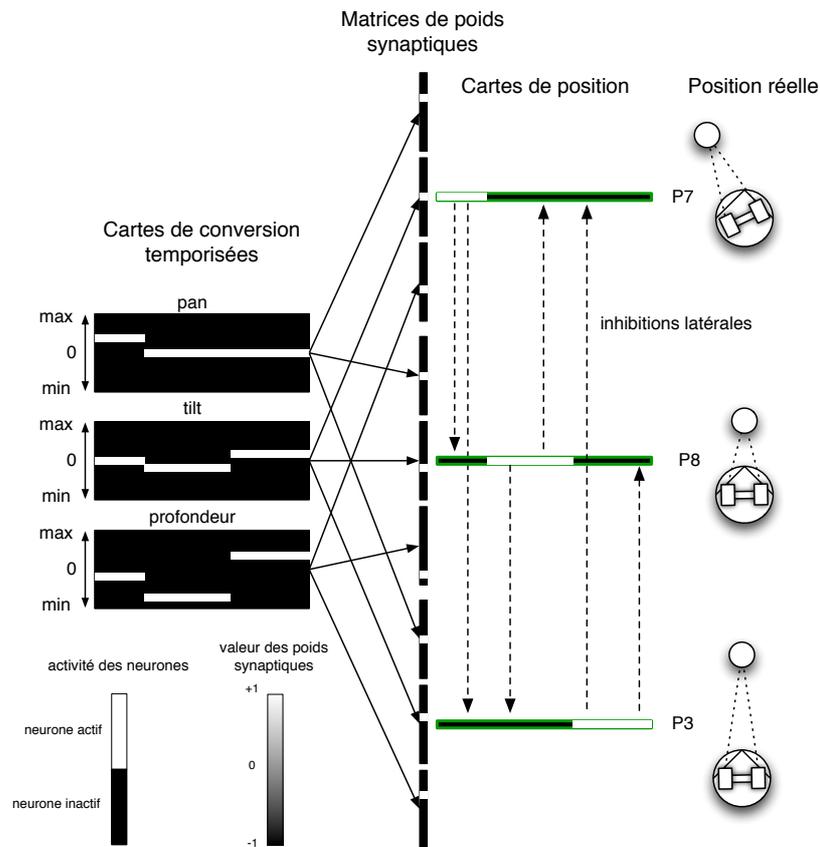


FIG. 3.28 – Schéma d'un réseau utilisé pour l'apprentissage des motifs de positions.

II. La représentation égocentrique de l'espace

À la différence de la navigation, les interactions d'un robot avec un objet ne nécessitent pas une représentation allothétique de l'espace. C'est la représentation de la position relative entre le robot et l'objet qui sera prise en compte à partir de données proprioceptives de l'orientation de la caméra et de la profondeur du motif perçu. La position d'un objet est encodée séparément de sa représentation visuelle, bien que celle-ci soit nécessaire pour le calcul de disparité binoculaire. Une représentation de position pourra donc être successivement activée par des objets de nature différente. La séparation du "où" et du "quoi" permet également d'apprendre des actions généralisables à tout type d'objet comme par exemple s'approcher, contourner, etc... Nous utilisons trois types de cartes de conversion temporisées pour l'apprentissage des motifs de position :

- une carte mesurant l'azimut de la platine (pan)
- une carte mesurant l'élévation de la platine (tilt)
- une carte mesurant la profondeur par disparité binoculaire (profondeur)

Le dernier type de carte de conversion mesure la distance entre deux activations d'une carte tracker à partir des images gauches et droites issues des caméras. La Figure

3.27 montre les images obtenues par stéréovision, l'activation d'une carte de vue et la carte de conversion de profondeur produite après avoir mesuré la disparité binoculaire de la paire de détecteurs. Le réseau de cartes utilisé pour l'apprentissage de motifs de positions ne contient qu'une seule couche de cartes en compétition (Figure 3.28). Chacune de ces cartes possède trois matrices de poids connectées aux trois cartes de conversion. L'initialisation des poids est obtenue par un tirage aléatoire de distribution uniforme. Il n'existe pas ici de compétition à l'activation entre les cartes de conversion car elles ne sont pas issues de la même modalité, les poids négatifs sont donc engendrés par l'absence d'activation à l'entrée du neurone (voir III.). L'apprentissage ainsi que la compétition interviennent durant toute la durée de la propagation de l'activation (rangs $(t - 1)$, $(t - 2)$, etc...).

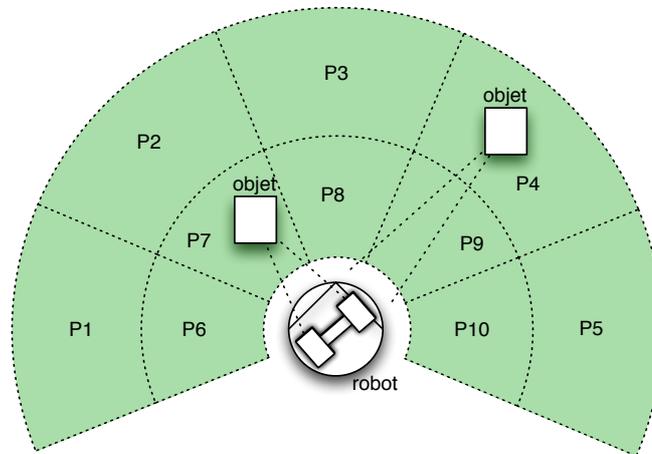


FIG. 3.29 – Schéma des champs de position.

Contrairement aux champs de vue, les champs de position dépendent de l'orientation de la platine du robot (pan) et il n'existe pas de chevauchement entre champs voisins. Ce dernier point est dû à la discrétisation des signaux de proprioception là où la transition entre vue est continue.

III. La représentation proprioceptive des actions

L'apprentissage de motifs d'action est très semblable à celui des motifs de position. L'entrée mesure ici l'information proprioceptive des commandes en vitesse sur chaque roue. Ce réseau est donc composé de deux cartes de conversions temporisées et chaque carte d'action est associée à deux matrices de poids synaptiques. Les champs d'action constituent un pavage uniforme dans l'espace des commandes de vitesse sur les roues.

IV. L'apprentissage incrémental de motifs de position et d'action

Contrairement à l'apprentissage des vues d'objet, les motifs d'actions et de positions pourraient être intégrés dès la conception. Mais un nombre trop important de ces cartes serait alors un handicap au traitement en temps réel du simulateur. C'est pourquoi

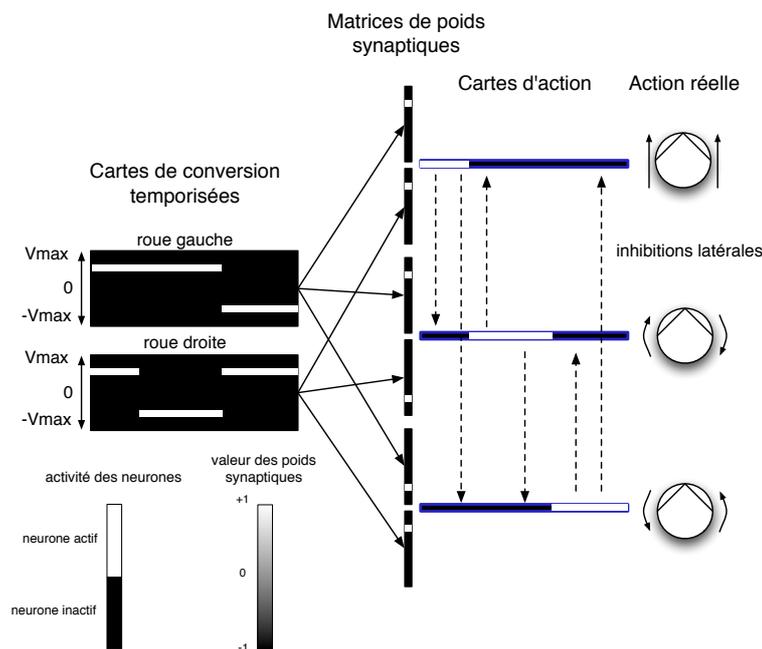


FIG. 3.30 – Schéma d'un réseau utilisé pour l'apprentissage des motifs d'actions.

nous avons opté pour un apprentissage incrémental de ce type de représentations. Le système doit donc décider quelles représentations sont utiles à l'apprentissage sensori-moteur. C'est la composante subjective, le niveau global de récompense perçu lors des interactions, qui doit guider l'apprentissage. Une variation d'effet enregistrée à un instant donné doit informer le système de la présence d'un événement favorable ou non, et celui-ci doit extraire l'information pertinente dans le but de favoriser ou d'éviter cette variation dans le futur. A chaque pas de temps, nous mesurons la variation de récompense globale et nous la rendons disponible à toutes les cartes du réseaux. Tous les neurones d'une carte temporisée, situés au rang (t) enregistrent cette variation et n'autorisent l'apprentissage que si celle-ci dépasse un seuil de récompense. Il est également nécessaire de favoriser l'apprentissage des représentations qui ont précédées la variation d'effet, le mécanisme décrit précédemment s'applique donc aussi aux neurones situés au rang ($t - 1$) des cartes temporisées.

3.9 Récompense perçue, récompense induite

Nous avons montré l'utilisation de techniques d'apprentissage non-supervisé dans le but de permettre au système l'acquisition de représentations du monde sous la forme de cartes neuronales. Chaque représentation objective ainsi formée est stockée au sein d'un ensemble de poids synaptiques et d'une fonction de transfert adaptative. Ce premier type d'apprentissage permet de catégoriser les informations enregistrées par les capteurs du robot et de définir des représentations stables de haut niveau. Cependant,

les caméras ou les capteurs proprioceptifs, encodant la position des différents degrés de liberté, ne sont pas les uniques sources d'informations disponibles. Comme nous l'avons mentionné en introduction, notre système est également sensible à une valeur de récompense globale $\mathcal{R}(t)$ ($0 \leq \mathcal{R}(t) \leq 1$) qui peut être à la fois perçue (représentations initiales) ou induite par le système lui-même (représentations apprises). La valeur de récompense globale est obtenue en calculant la moyenne des contributions locales de récompense, à la fois initiales et apprises. Cette valeur est donc un nombre réel compris dans l'intervalle $[-1, 1]$ pour lequel les valeurs négatives représentent les situations que le système devra éviter (punitions) et les valeurs positives, les situations que le système aura pour but de reproduire (récompenses). Il est à noter que seules les représentations actives à un instant donné participent au calcul de la valeur de récompense globale. Les représentations apprises par le système ont donc un double rôle de descripteur du monde (aspect objectif) et de capteur de récompense (aspect subjectif). Les représentations initiales peuvent être comparées à des capteurs de récompense primitifs donnés au système au moment de sa conception. Par exemple, un niveau négatif de récompense (punition) peut être associé aux consignes de vitesse sur les moteurs suivant une valeur proportionnelle à l'énergie consommée, de sorte que les trajectoires de déplacement directes soient privilégiées, car maximisant le niveau de récompense total perçu (en minimisant la punition). Ainsi, le choix de l'ensemble des représentations initiales aura une forte influence sur le comportement du robot, il incombe cependant au système d'apprendre à générer les séquences d'actions menant dans des états associés à des niveaux de récompense élevées. La valeur de récompense perçue au travers des représentations subjectives initiales ne nécessite aucun apprentissage. De plus, les représentations initiales ne sont pour le robot rattachées à aucune représentation objective de l'environnement (vue, position ou action). En d'autres termes le système est initialement incapable de discerner les situations qui pourraient lui fournir une récompense ou une punition. Cette information n'a aucune "incarnation physique" perceptible par le robot. C'est pourquoi il lui est nécessaire d'apprendre à associer à chaque instant le niveau de récompense perçue aux représentations susceptibles d'être en rapport avec cette récompense. C'est une nouvelle fois la présence simultanée de ces deux informations (niveau global de récompense et activation d'une représentation) qui est la clé du mécanisme associatif : à chaque instant tout neurone actif est en mesure d'apprendre le niveau de récompense $r_i(t)$ coïncidant avec son activation selon l'équation :

$$r_i(t+1) = \alpha_r \mathcal{R}(t) + (1 - \alpha_r) r_i(t) \quad (3.34)$$

où $\mathcal{R}(t)$ est le niveau de récompense global à l'instant t et α_r ($0 \leq \alpha_r \leq 1$) le coefficient d'apprentissage de la récompense.

La valeur de récompense est donc au centre d'un mécanisme d'apprentissage dédié (Figure 3.31). Les représentations initiales permettent d'amorcer cet apprentissage en fournissant une mesure primitive du niveau de récompense. Puis, parallèlement à l'apprentissage des représentations objectives de haut-niveau (vues, positions, actions), la valeur instantanée de récompense globale est associée à chacune de ces représentations lors de leurs activations (diffusion de la récompense dans le réseau). Finalement, chaque représentation active contribue au calcul du niveau de récompense globale (récompense

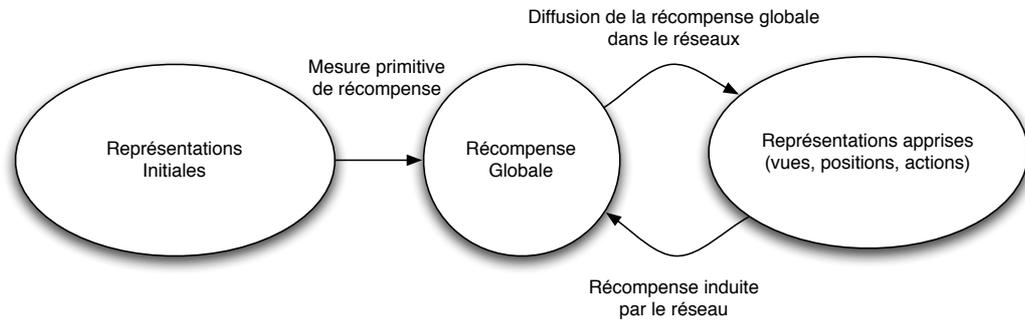


FIG. 3.31 – Schéma du cycle de la récompense.

induite).

3.10 Conclusion

Nous venons de présenter dans ce chapitre le modèle de neurone qui a été développé pour permettre l'apprentissage non-supervisé de motifs perceptifs. Ce modèle est caractérisé par l'utilisation d'une fonction de transfert capable d'adapter le niveau de spécialisation du neurone en mémorisant les niveaux d'activations passés. La taille de l'information transmise entre les neurones a également été fortement réduite par l'introduction d'une phase de tir probabiliste inspirée des modèles de neurones stochastiques. Ce dernier point a également permis de limiter les phénomènes de sur-apprentissage car un neurone est à présent capable de décharger, malgré un niveau de potentiel bas, et ainsi remettre en cause son motif idéal en explorant les motifs voisins. Nous avons présenté dans un second temps les trois classes de motifs perceptifs qui vont servir de base à la création d'association puis à la prise de décision. Ces deux points vont être développés au sein des prochains chapitres.

Références

- [Bliss 93] T.V. Bliss & G.L. Collingridge. *A synaptic model of memory : long-term potentiation in the hippocampus*. Nature, vol. 361, no. 6407, pages 31–39, 1993.
- [Fregnac 88] Y. Fregnac, D. Shulz, S. Thorpe & E. Bienenstock. *A cellular analogue of visual cortical plasticity*. Nature, vol. 333, no. 6171, pages 367–370, 1988.
- [Fukushima 03] K. Fukushima. *Neocognitron for handwritten digit recognition*. Neurocomputing, vol. 51, pages 161–180, April 2003.
- [Gerstner 95] W. Gerstner. *Time structure of the activity in neural network models*. Phys. Rev. E, vol. 51, pages 738–758, 1995.
- [Hebb 49] Donald O. Hebb. *The organization of behavior : A neuropsychological theory*. Wiley, New York, June 1949.

- [Ito 93] M. Ito. *Synaptic plasticity in the cerebellar cortex and its role in motor learning*. The Canadian journal of neurological sciences, vol. 20, no. 3, pages S70–S74, 1993.
- [Koch 92] c. Koch & T. Poggio. *Multiplying with synapses and neurons*. Dans Academic Press, editeur, Single neuron computation, pages 315–345. Academic Press Professionnal, Inc., 1992.
- [Kohonen 77] T. Kohonen. *Associative memory : A system-theoretical approach*. Springer-Verlag, Berlin, 1977.
- [Kohonen 89] T. Kohonen. *Self-organization and associative memory*. Springer-Verlag New York, Inc., 1989.
- [Lapicque 07] L. Lapicque. *Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation*. J. Physiol. Pathol. Gen, no. 9, pages 620–635, 1907.
- [Maass 96] W. Maass. *Lower Bounds for the Computational Power of Networks of Spiking Neurons*. Neural Computation, vol. 8, no. 1, pages 1–40, 1996.
- [McCulloch 43] W.S. McCulloch & W. Pitts. *A Logical Calculus of the Ideas Immanent in Nervous Activity*. Bulletin of Mathematical Biophysics, vol. 5, pages 115–133, 1943.
- [Minsky 69] M. Minsky & S. Papert. *Perceptrons : An introduction to computational geometry*. MIT Press, 1969.
- [Paquier 04] W. Paquier. *Apprentissage ouvert de représentations et de fonctionnalités en robotique : analyse, modèles et implémentation*. Thèse de doctorat, Université Paul Sabatier de Toulouse, 2004.
- [Riesenhuber 99] M. Riesenhuber & T. Poggio. *Hierarchical models of object recognition in cortex*. Nature Neuroscience, vol. 28, pages 1019–1025, 1999.
- [Ritz 97] R. Ritz & T.J. Sejnowski. *Synchronous oscillatory activity in sensory systems : new vistas on mechanisms*. Current Opinion in Neurobiology, vol. 7, no. 4, pages 536–546, 1997.
- [Rolls 98] E.T. Rolls & A. Treves. *Neural networks and brain function*. Oxford University Press, 1998.
- [Rosenblatt 62] F. Rosenblatt. *Principles of neurodynamics ; perceptrons and the theory of brain mechanisms*. Washington, Spartan Books, 1962.
- [Rumelhart 86] D. E. Rumelhart, G. E. Hinton & R. J. Williams. *Learning Internal Representations by Error Propagation*. Dans D. E. Rumelhart & J. L. McClelland, éditeurs, Parallel Distributed Processing, volume 1, pages 318–362. MIT Press, 1986.
- [Shawe-Taylor 91] J. Shawe-Taylor, P. Jeavons & M. Daalen. *Probabilistic Bit Stream Neural Chip : Theory*. Connection Science, vol. 3, no. 3, pages 317–328, 1991.

- [Soodak 91] R.E. Soodak. *Reverse-Hebb plasticity leads to optimization and association in a simulated visual cortex*. Vis Neurosci., vol. 6, no. 5, pages 507–518, 1991.
- [Srinivasan 76] M.V. Srinivasan & G.D. Bernard. *A proposed mechanism for multiplication of neural signals*. Biological Cybernetics, vol. 21, no. 4, pages 227–36, 1976.
- [Sutton 98] Richard S. Sutton & Andrew G. Barto. Reinforcement learning : An introduction (adaptive computation and machine learning). The MIT Press, March 1998.
- [Wallis 96a] G. Wallis. *How neurons learn to associate 2D-views in invariant object recognition*. Rapport technique 37, Max-Planck-Institute für Biologische Kybernetik, Tübingen, Germany, 1996.
- [Wallis 96b] G. Wallis. *Using spatio-temporal correlations to learn invariant object recognition*. Neural Networks, vol. 9, no. 9, pages 1513–1519, 1996.
- [Yuille 89] A.L. Yuille, D.M. Kammer & D.S. Cohen. *Quadrature and the development of orientation selective cortical cells by Hebb rules*. Biological Cybernetics, vol. 61, no. 3, pages 183–194, 1989.

ASSOCIER

4.1	De la statique à la dynamique	87
4.2	Structure de la mémoire associative	88
	I. Vue générale	88
	II. Cartes référentes et groupes associatifs	89
4.3	Association sensori-motrice de motifs de position : <i>déplacements</i>	89
4.4	Association de vues d'objets	90
4.5	Actions, préconditions et <i>réflexes</i>	95
4.6	Associations et récompense : évaluation des gains	95
4.7	Complexité du réseau, élagage	97
4.8	Conclusion	97

4.1 De la statique à la dynamique

Les méthodes d'apprentissage présentées dans le précédent chapitre ont pour but d'extraire de façon non supervisée des motifs visuels ou proprioceptifs. Ces motifs sont statiques, ils forment la représentation d'un état perceptif à un instant donné et ne rendent compte ni de l'évolution de la perception ni de l'évolution des actions du robot. L'aspect temporel n'est cependant pas écarté lors de l'apprentissage des vues d'un objet, mais il n'est qu'un indice utile à la classification et n'a aucun pouvoir structurant. Nous allons présenter dans ce chapitre les méthodes employées afin de relier, par la création incrémentale d'associations, les trois classes de représentations nées de l'interaction entre le robot et l'objet : "Quoi", "Où", "Comment". Ces associations seront elles-mêmes divisées en trois classes. Une première classe aura pour objectif l'apprentissage de transitions entre deux positions successives d'un objet engendrées par une action du robot (classe *déplacement*). Ces transitions ne seront pas liées à la nature de l'objet et seront donc généralisables aux interactions entre le robot et des objets différents. La seconde classe d'associations devra capturer les transitions

entre les vues d'un objet pendant le déplacement du robot (classe *transformation*). Ces transitions permettront de structurer la représentation de l'objet en associant aux cartes de vues des cartes de positions et d'actions. La dernière classe d'associations permettra au robot d'associer des contextes perceptifs (vue, position) aux actions sous la forme de préconditions (classe *réflexe*). Plus qu'une représentation statique de l'environnement, ces associations tendent à identifier des invariances dans la dynamique des interactions entre le robot et l'objet.

4.2 Structure de la mémoire associative

I. Vue générale

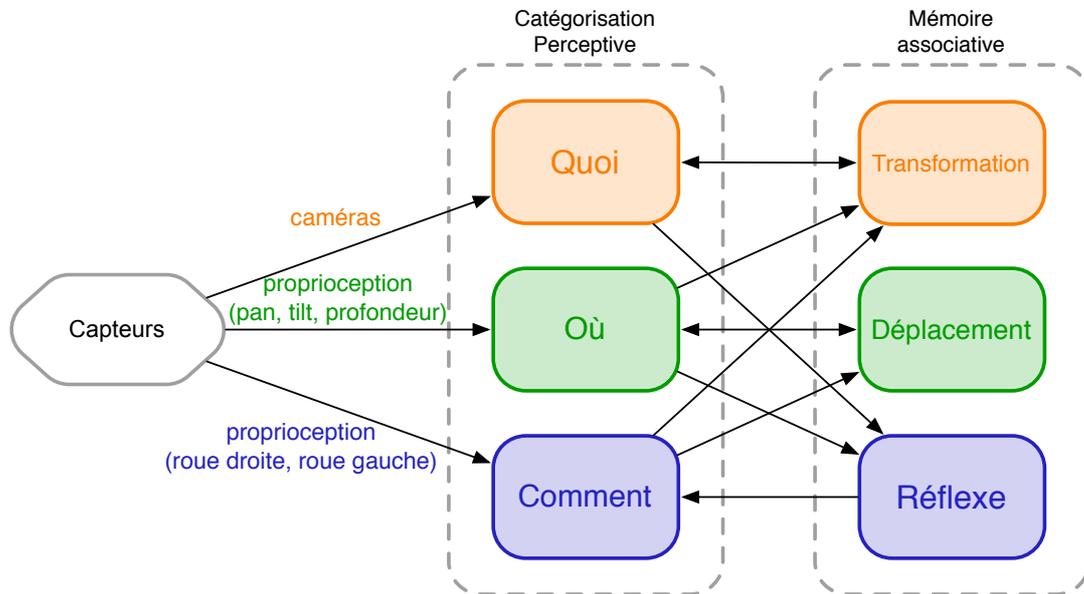


FIG. 4.1 – La mémoire associative est formée de trois classes d'associations : les *transformations*, les *déplacements* et les *réflexes*. A chacune de ces classes correspond un schéma de connexion particulier.

La mémoire associative reçoit en entrée les activations issues des cartes neuronales construites lors de la phase de catégorisation perceptive. Les connexions vers les cartes associatives sont donc elles-mêmes créées de façon incrémentale. Conformément à la règle de Hebb, les neurones associatifs (neurones cibles) ont le pouvoir de renforcer leurs connexions entrantes lorsqu'ils sont activés. Ces neurones doivent donc bénéficier lors de l'apprentissage d'un ensemble de connexions entrantes de taille maximale afin de détecter dans cet ensemble la cooccurrence d'activations. Le choix d'un schéma de connexion complet (de l'ensemble des représentations vers les associations) n'est cependant pas envisageable car il serait trop gourmand en temps de calcul. C'est pourquoi nous avons décidé d'interdire certaines connexions non pertinentes et de procéder à un

élagage des connexions inutiles (poids synaptique nul) après l'apprentissage. La Figure 4.1 montre le schéma simplifié de connexions qui a été utilisé.

II. Cartes référentes et groupes associatifs

Le rôle d'une association est d'anticiper l'activation d'une carte neuronale de vue ("Quoi") de position ("Où") ou d'action ("Action"). Ainsi, chacune de ces cartes se voit associer un ensemble de cartes associatives qui anticiperont son activation dans différents contextes. Cet ensemble est appelé un groupe associatif. Tout comme leurs cartes cibles, les cartes associatives sont temporisées : elles gardent en mémoire une séquence d'activation passée. Chaque carte cible d'un groupe associatif est qualifiée de carte référente. Un groupe associatif peut donc être composé d'un ensemble d'associations de *déplacement*, de *transformation* ou de *réflexe* selon que la carte référente est respectivement une position, une vue ou une action. Pour assurer l'apprentissage non supervisé des associations, toutes les cartes d'un groupe associatif sont reliées par des connexions latérales inhibitrices. De plus, la carte référente est en mesure de signaler son activation au groupe associatif lorsqu'une transition vient d'avoir lieu pour activer l'apprentissage. En retour, chaque association est connectée à sa carte référente pour reporter dans celle-ci le résultat d'une prévision (cette fonction sera développée au prochain chapitre). La construction des groupes associatifs est incrémentale. Lorsqu'une carte de vue, de position ou d'action a atteint un niveau de spécialisation suffisant, un nouveau groupe ne contenant qu'une carte associative est créé pour permettre l'apprentissage d'association pour cette nouvelle carte. Par la suite, de nouvelles cartes associatives seront ajoutées et connectées au groupe dès qu'une carte associative aura terminé son apprentissage.

4.3 Association sensori-motrice de motifs de position : *déplacements*

Dans le cas des *déplacements*, les cartes associatives reçoivent en entrée les flux d'activation de l'ensemble des cartes d'actions et de positions, à l'exception de ceux provenant de la carte de position référente. Chacune de ces connexions se voit attribuer un poids synaptique unique (matrice 1x1) qui, en fin d'apprentissage vaudra 1 si l'activité de la carte afférente a précédé celle de la carte référente, -1 sinon. À terme, seules deux connexions auront conservé des poids égaux à 1 définissant ainsi un couple (position,action) à l'origine du déplacement (Figure 4.2). Les autres connexions sont effacées du réseau au terme de l'apprentissage (voir 4.7). Chaque carte référente doit piloter l'apprentissage de toutes les cartes de *déplacement* qui lui sont associées en générant des signaux d'excitation lorsqu'une transition est détectée. C'est l'identification de fronts montants d'activation qui est à l'origine des signaux excitateurs (Figure 4.3). Lorsque le robot pénètre un champ de lieu, la carte de position associée devient active, un front montant est alors détecté et un signal d'excitation transmis par connexion latérale vers toutes les cartes de *déplacement* du groupe associatif. Cette excitation engendre un signal autorisant l'apprentissage pour tous les neurones du groupe associatif situés au rang inférieur à celui du front montant détecté (pas de temps précédent).

Le rôle de la carte référente est donc de permettre l'apprentissage d'associations au rang précis où ont lieu les transitions vers l'état référent. C'est alors qu'intervient la compétition entre les cartes de *déplacement* et la carte ayant le plus fort potentiel peut finalement apprendre.

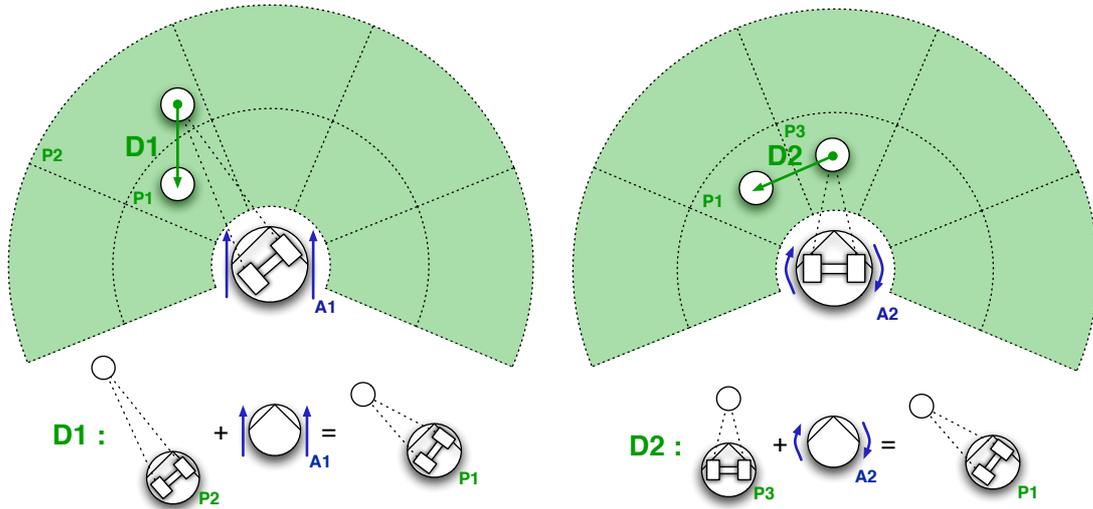


FIG. 4.2 – *Déplacements* et champs de lieu : les associations de *déplacement* d'un même groupe associatif permettent d'identifier des couples (position,action) menant à la même position référente.

4.4 Association de vues d'objets

À la différence des représentations de positions, les cartes de vues d'objet ne sont pas temporisées, or cette caractéristique est nécessaire à l'apprentissage de motifs de transition. Il faut donc convertir l'activation rétinotopique des cartes de vue (voir 3.7) en une activation temporisée. Cette conversion est obtenue en connectant chacun des neurones d'une carte de vue rétinotopique au premier neurone d'une carte temporisée. À l'instar des cartes de conversion temporisées, ces cartes propagent à chaque pas de temps l'activation des représentations de vue fournies en entrée (Figure 4.4). L'apprentissage de transitions entre vues nécessite en outre de définir des limites de détection strictes. Lorsque le robot se déplace à la frontière d'un champ de vue, le potentiel de la carte de vue active réduit progressivement, entraînant une réduction simultanée de la probabilité de décharge et donc l'instabilité du signal de sortie. Cette instabilité est un obstacle à l'utilisation des détections de fronts montant s'il n'existe pas de filtrage préalable du signal de détection. En effet, si le robot quitte un champ de vue lors d'un déplacement, l'instabilité de la détection donnerait lieu à de nombreux front montant que le système pourrait interpréter comme le signe d'une transition menant vers cette vue, alors qu'il s'agit en fait de la manifestation d'une reconnaissance approximative. Pour atteindre une dissociation plus nette des zones de détections, nous utilisons un filtrage à hystérésis

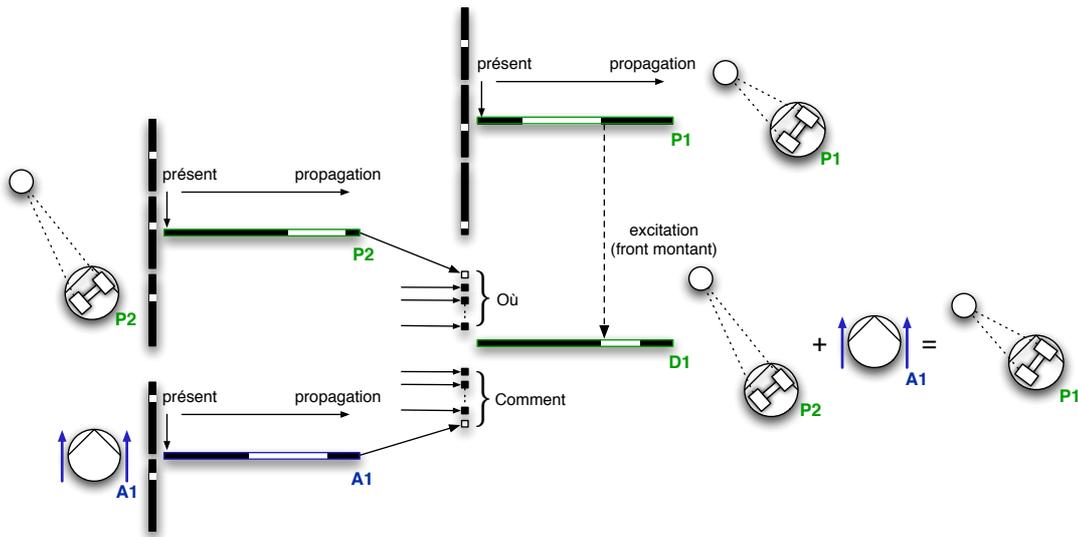


FIG. 4.3 – Schéma de connexion d'une association de *déplacement* : chaque carte de *déplacement* reçoit en entrée les activations de l'ensembles de cartes d'action et de position. Lorsqu'un front montant est détecté dans sa carte référente, l'apprentissage est activé dans le groupe associatif (ici seule la carte de *déplacement* ayant remporté la compétition est représentée).

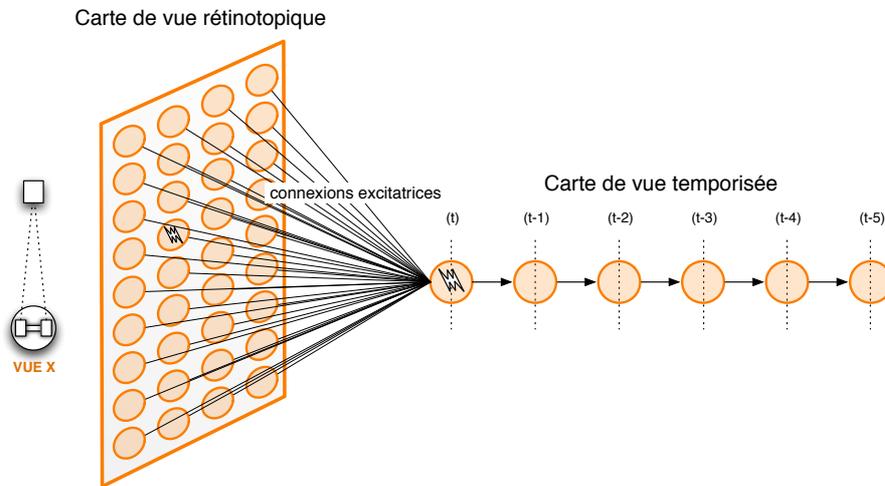


FIG. 4.4 – Conversion d'une carte de vue rétinotopique vers une carte de vue temporisée : le premier neurone de la carte de vue temporisée mesure la valeur maximale de potentiel provenant des neurones de la carte rétinotopique. Les activations de ce premier neurone sont ensuite propagées d'un rang (vers la droite) à chaque pas de temps.

sur la valeur maximale de détection obtenue en sortie des cartes rétinotopiques : le niveau de détection nécessaire au passage de l'état inactif à l'état actif est supérieur à celui nécessaire au passage de l'état actif à l'état inactif. Ce comparateur est réalisé en utilisant une connexion récurrente et un seuil fixe au niveau du premier neurone de la carte associative (Figure 4.5).

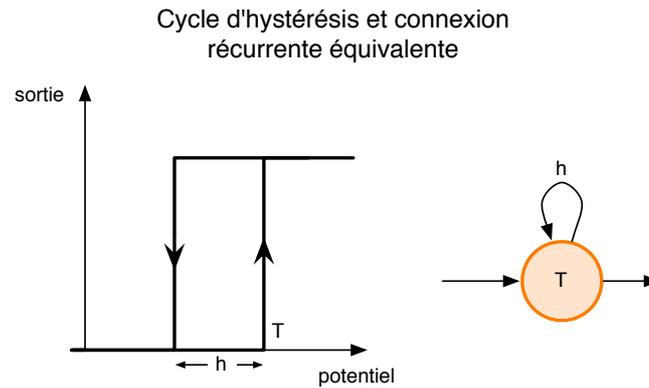


FIG. 4.5 – Un neurone (à droite) pourvu d'une connexion récurrente de poids h et de seuil T équivaut à un filtre à hystérésis dont le cycle est donné à gauche.

Les associations de *transformation* menant à une représentation de vue référente (carte 'Quoi') sont obtenues par l'apprentissage de contextes perceptifs comprenant une vue, une position ainsi que l'action menant à la vue référente. Comme le montre la Figure 4.6, les seuls paramètres (vue initiale, action) ne permettent pas de faire la distinction entre deux transitions différentes car une partie de l'information relative à la direction du mouvement par rapport à l'objet est contenue dans les représentations de position. Le schéma des connexions des cartes associatives de *transformation* est similaire à celui des cartes *déplacement* : la détection de front montant, obtenue après filtrage, dans la carte de vue référente permet d'activer l'apprentissage d'un ensemble de cartes associatives en compétition (Figure 4.7).

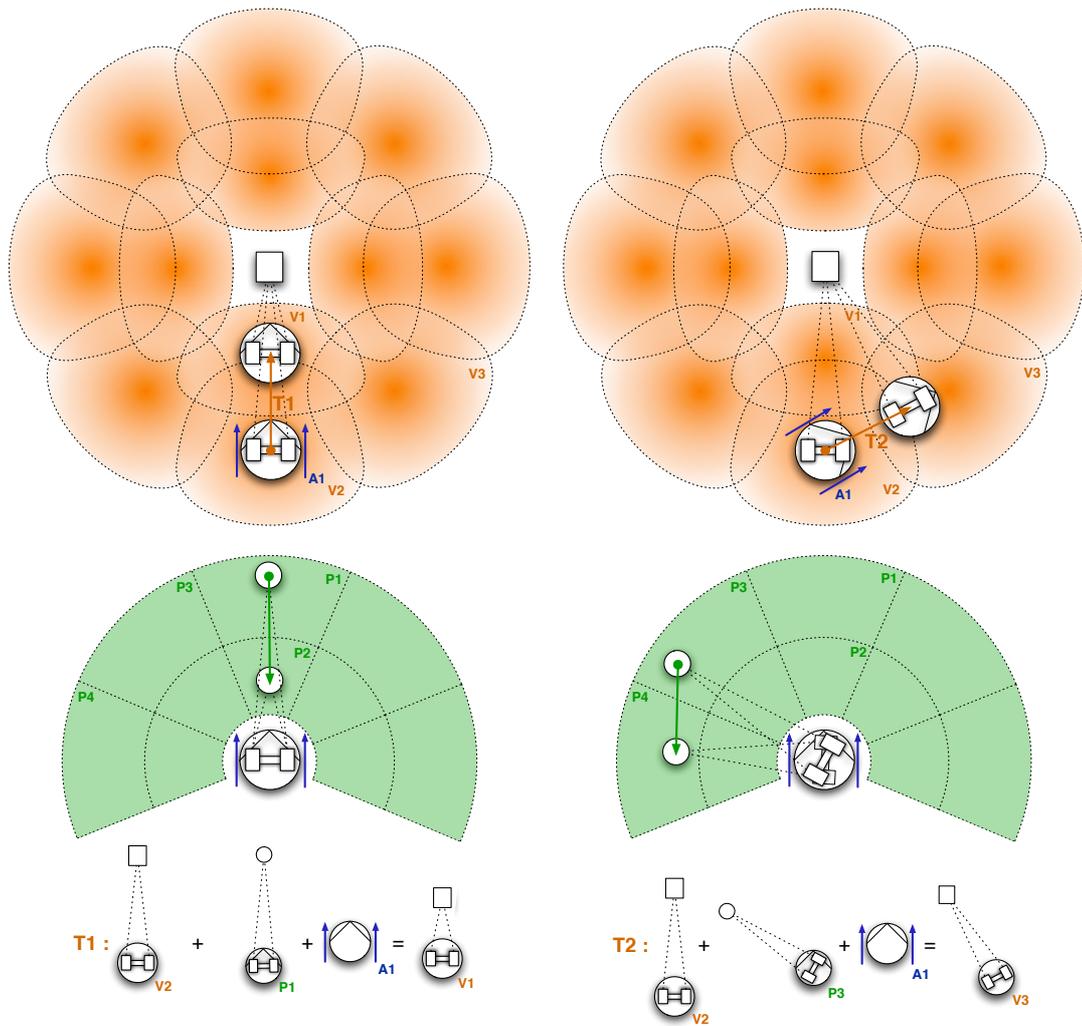


FIG. 4.6 – Ce schéma illustre deux *transformations* : T1 à gauche et T2 à droite dont les positions initiales sont différentes (respectivement P1 et P3) menant à deux vues distinctes.

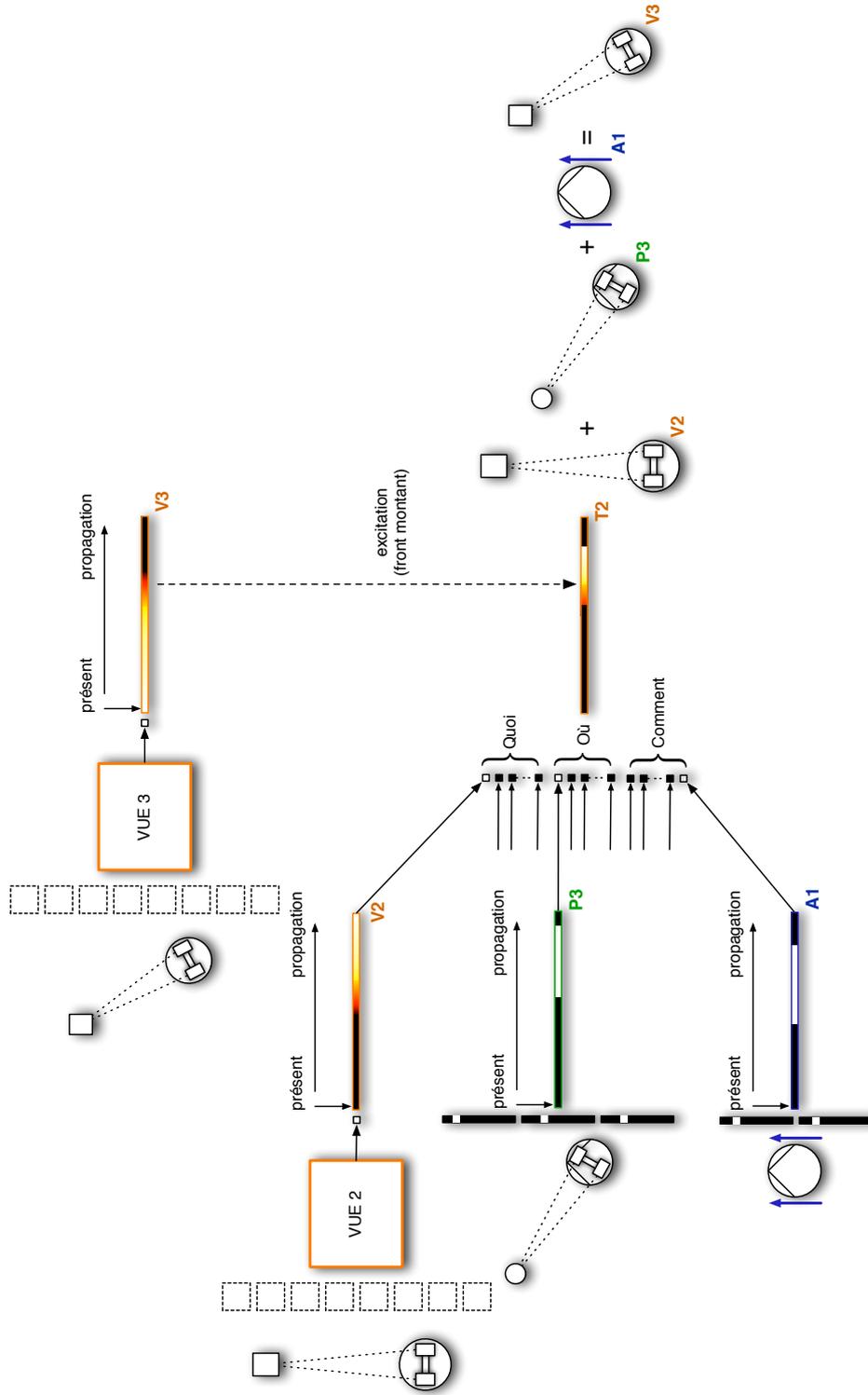


FIG. 4.7 – Schéma de connexion d'une association de *transformation* : chaque carte de *transformation* reçoit en entrée les activations de l'ensemble de cartes de vue, d'action et de position. Lorsqu'un front montant est détecté dans sa carte référente, l'apprentissage est activé dans le groupe associatif (ici encore, seule la carte de *déplacement* ayant remporté la compétition est représentée).

4.5 Actions, préconditions et *réflexes*

Le troisième type d'associations que nous allons à présent envisager se distingue des précédents par l'introduction de la valeur de récompense dans le pilotage de l'apprentissage. Le rôle de ces associations, qualifiées d'associations *réflexe*, est d'associer à un couple (position,vue), les actions engendrant une variation significative de récompense pour, par la suite, faciliter ou d'inhiber leur activation. Or en règle générale, cette variation de récompense n'a pas lieu au moment où débute l'action mais plus tard, lorsque les modifications sur l'environnement apportées par l'action seront perçues et induiront de fait une variation de récompense. Le signal d'apprentissage ne peut donc être obtenu en exploitant les fronts montant d'activation de la carte référente, mais en observant une variation de récompense ayant lieu pendant le déroulement l'action. Si une carte d'action est active lors d'une variation significative de la valeur globale de récompense, un signal excitateur est propagé en direction des cartes d'association *réflexe* du groupe associatif de l'action courante. Ainsi, chaque carte associative pourra mémoriser et par la suite reconnaître les préconditions de l'action référente à laquelle elle est associée. Le schéma des connexions utilisé pour l'apprentissage des cartes d'association *réflexe* est représenté par la Figure 4.8.

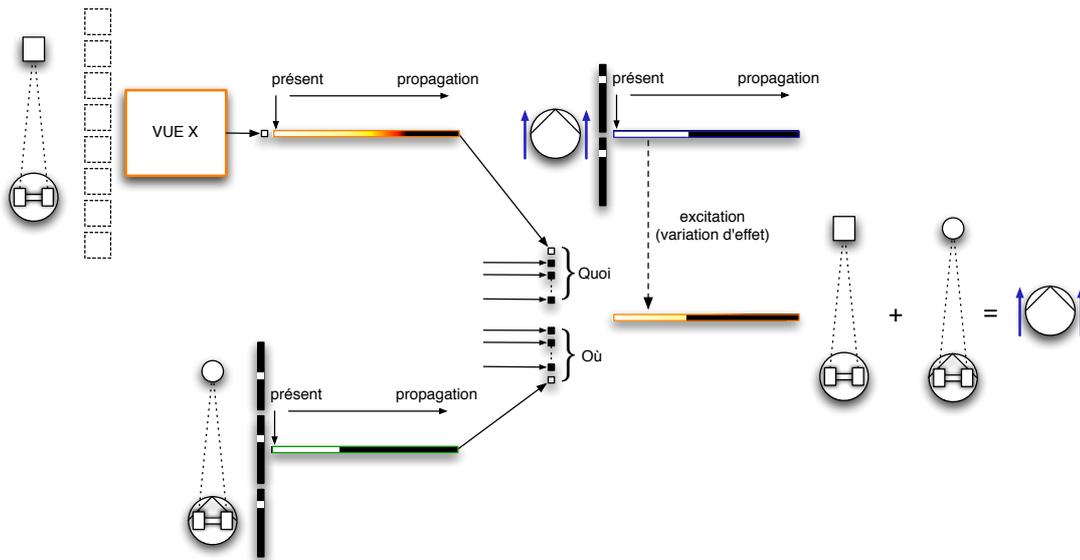


FIG. 4.8 – Schéma de connexion d'une association *réflexe* : chaque carte *réflexe* reçoit en entrée les activations de l'ensemble des cartes de vue et de position. Lorsqu'une variation de récompense est détectée, l'apprentissage est activé dans le groupe associatif.

4.6 Associations et récompense : évaluation des gains

Nous avons présenté, lors du chapitre précédent, une première forme d'apprentissage des niveaux de récompense associés aux représentations de vue, de position et d'action.

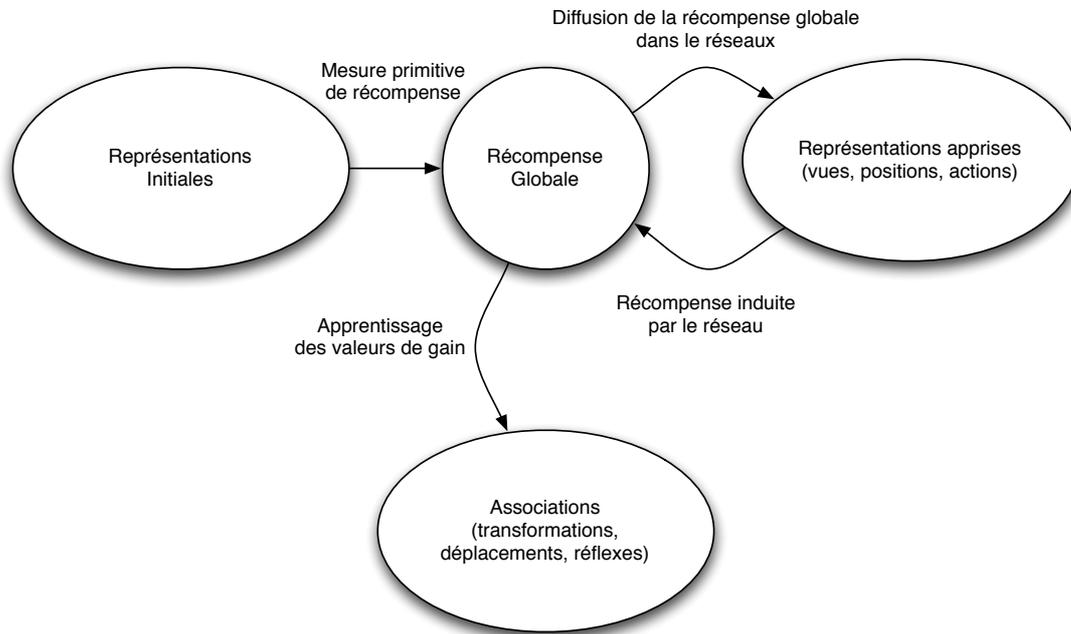


FIG. 4.9 – L'apprentissage des valeurs de gain par les associations

Cet apprentissage doit permettre au système d'évaluer le contexte courant en lui associant une valeur de récompense globale (voir 3.9). Dans cette phase, il s'agit donc de construire un modèle statique de l'environnement dans ses composantes objectives et subjectives. Nous souhaitons à présent apprendre la dynamique des interactions entre le robot et l'environnement et donc, du point de vue de la récompense, apprendre les gains associés aux transitions entre états.

La valeur de récompense perçue par le système à un instant donné est le résultat d'une récompense mesurée par un ensemble de capteurs primitifs de récompense : les représentations initiales et l'expression des récompenses locales induites par l'activation de représentations apprises : les représentations de vues, de positions et d'actions. Chaque carte associative doit pour sa part apprendre la variation de récompense mesurée lorsque se produit la transition qu'elle représente (*déplacement*, *transformation* ou *réflexe*). L'apprentissage d'une variation de récompense (le gain) entre deux instants consécutifs permettra par la suite de calculer une valeur de gain espéré (l'utilité) pour chaque action envisagée par le système. La variation de récompense engendrée par une association ne peut-être mesurée qu'à l'instant où ces effets sont réellement perçus. Cela correspond au moment où la carte référente de cette association (le résultat prédit de la transition) devient active et que les préconditions ont été activées à l'instant précédent (observation simultanée des causes et des effets). Alors, la valeur de gain induit $\mathcal{G}(X)$ par l'association X est calculée en soustrayant à la récompense globale perçue la récompense perçue à l'instant précédent l'apprentissage très semblable à celui présenté

dans 3.9 est donné par l'équation suivante :

$$\mathcal{G}_X(t+1) = \alpha_r(\mathcal{R}(t) - \mathcal{R}(t-1)) + (1 - \alpha_r)\mathcal{G}_X(t) \quad (4.1)$$

où $\mathcal{R}(t)$ est le niveau de récompense global à l'instant t et α_r ($0 \leq \alpha_r \leq 1$) le coefficient d'apprentissage de la récompense. À la différence des représentations de vues, de position ou d'action, les associations ne participe pas au calcul de la valeur globale de récompense. La valeur relative de récompense associée n'aura pas pour but de décrire l'état subjectif présent du système mais servira de base au choix des futures actions du robot par l'exploration des gains possibles dans un horizon de prédiction et l'apprentissage des valeurs d'utilité. Ce dernier mécanisme sera le sujet du prochain chapitre.

4.7 Complexité du réseau, élagage

La création de cartes associative nécessite un grand nombre de connexions qui, du point de vue du simulateur, représente un important handicap pour le traitement en temps réel de l'information. Plusieurs techniques ont toutefois été mises en œuvre pour réduire le temps de traitement inhérent à l'échange de donnée dans la mémoire associative. Une première technique intégrée au cœur même du simulateur consiste à propager les activations des neurones uniquement lorsque ceux-ci ont déchargés (traitement event-driven) réduisant ainsi considérablement la quantité d'information échangée. Le nombre de neurones actifs à un instant donné est en effet très petit comparé au nombre total de neurones simulés (représentations creuses). Il est également possible d'obtenir une amélioration notable des performances en procédant à une simplification du réseau associatif par la suppression de connexions inutiles. L'intérêt d'une connexion ne peut cependant être déterminée qu'à l'issue de la convergence de l'apprentissage, lorsque son poids synaptique associé se retrouve sensiblement égal à 0. Nous considérons qu'une carte associative a atteint un seuil de spécialisation autorisant l'élagage des connexions lorsqu'elle exhibe une fonction de transfert caractéristique d'un comportement binaire.

4.8 Conclusion

Nous venons de décrire dans ce chapitre l'apprentissage incrémental d'associations sensori-motrices. Ces associations constituent un modèle de la dynamique des interactions entre le robot et l'environnement et permettent d'anticiper l'activation de trois types de représentations : les vues, les positions et les actions. Les cartes temporisées fournissent un support à l'apprentissage non supervisé en conservant en mémoire la trace des séquences d'activation passés tout en autorisant la compétition au sein des groupes associatifs. Nous avons en outre présenté l'apprentissage des variations de récompense ayant lieu lors des transitions d'états. Ces valeurs de gains seront exploitées par la suite pour évaluer l'espérance de gain associée à chaque action (valeur d'utilité). Nous montrerons dans le prochain chapitre comment exploiter l'ensemble des associations pour construire en ligne l'horizon des états atteignables par le robot et évaluer l'utilité de chaque action.

PRÉVOIR ET AGIR

5.1	Neurones et connexions pour la prévision	100
I.	Cartes temporisées étendues	100
II.	Stabilité perceptive et diffusion de la prévision	100
A.	Déplacements et transformations	101
B.	Réflexes	101
III.	Connexions de maintien d'activation	102
IV.	Cohérence de l'activation	104
A.	Réflexe	105
B.	Déplacement	105
C.	Transformation	105
5.2	Propagation des gains et valeur d'utilité des association <i>réflexes</i>	106
D.	Réflexe	106
E.	Déplacement	107
F.	Transformation	107
5.3	Propagation de la prévision : représentation perspective	107
5.4	Le réseau associatif résultat de l'apprentissage	109
5.5	Choix de l'action, cohérence de la prévision	110
5.6	Résolution d'une action, validité de la prévision	110
5.7	Aspects temporels	114
5.8	Conclusion	114

Nous allons présenter dans ce chapitre l'utilisation des associations de *déplacement*, de *transformation* de vue ainsi que les associations *réflexes* dans le but de construire un horizon de prévision pour l'évaluation des valeurs d'utilité et la prise de décision. L'idée générale de ces travaux consiste à considérer la prévision comme une diffusion d'activation dans la mémoire associative. La représentation classique en arbre d'état et de transition ne sera pas adoptée en tant que telle pour définir la structure de donnée et l'algorithme de décision bien qu'il soit possible de représenter l'activité du réseau par ce biais. La création incrémentale des connexions associatives doit donc se

substituer aux algorithmes d'inférence utilisés dans des systèmes classiques. Ainsi, nous souhaitons exploiter pleinement les caractéristiques des neurones formels ainsi que celles du simulateur comme l'utilisation d'algorithmes locaux et l'aspect event-driven.

5.1 Neurones et connexions pour la prévision

I. Cartes temporisées étendues

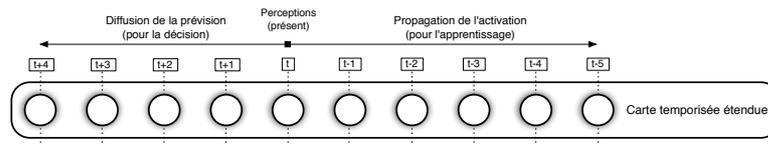


FIG. 5.1 – Les cartes temporisées étendues sont des cartes temporisées auxquelles sont ajouté des neurones dédiés à la prévision d'activation

Les cartes temporisées utilisées jusqu'à présent permettent de mémoriser une séquence d'activation. Nous avons montré comment cette mémoire du passé peut servir de support aux cartes associatives pour l'apprentissage de transitions ou d'actions *réflexes*. Nous proposons à présent d'étendre les cartes temporisées afin de représenter un horizon fini de prévision pour chaque représentation. À chaque neurone de prévision est associé un rang (de $t+1$ à $t+4$ dans la Figure 5.1) symbolisant le nombre d'actions consécutives nécessaire à la réalisation de cette prévision. Ainsi, l'activation d'un neurone au rang $t+2$ d'une carte de position P1 signifie que le système est en mesure d'atteindre la position P1 par l'intermédiaire de deux actions consécutives. Nous insistons sur le fait que le rang d'un neurone reflète un nombre de transition et non un nombre de pas de temps du simulateur. Cet aspect est indispensable pour permettre la diffusion des signaux de prévision dans le réseau.

II. Stabilité perceptive et diffusion de la prévision

L'état courant perçu par le robot est défini par un couple (vue, position) matérialisé par l'activation de deux neurones situées dans des cartes de positions et de vue au rang t . Chacun de ces neurones est de plus associé à un champ d'activation (respectivement un champ de vue ou un champ de lieu) au sein duquel l'activation est maintenue malgré le déplacement du robot. Il existe donc, pendant le déplacement du robot, une période de temps durant laquelle le contexte perceptif demeure stable. Cette durée dépend bien sûr de la surface des champs d'activation et de la vitesse de déplacement du robot, du terrain, mais son ordre de grandeur peut toutefois être évalué à quelques secondes. Étant donnée l'implémentation actuelle du simulateur, nous obtenons une vitesse de traitement de l'ordre de 15 phases d'intégration et de tir par seconde. Ainsi, plusieurs dizaines de phases d'intégrations et de tir peuvent avoir lieu lors du déplacement du robot en zone de stabilité perceptive. La vitesse de propagation de l'information de

prévision dans la mémoire associative est donc nettement supérieure à la vitesse de "consommation" des prévisions par la prise de décision.

A. DÉPLACEMENTS ET TRANSFORMATIONS

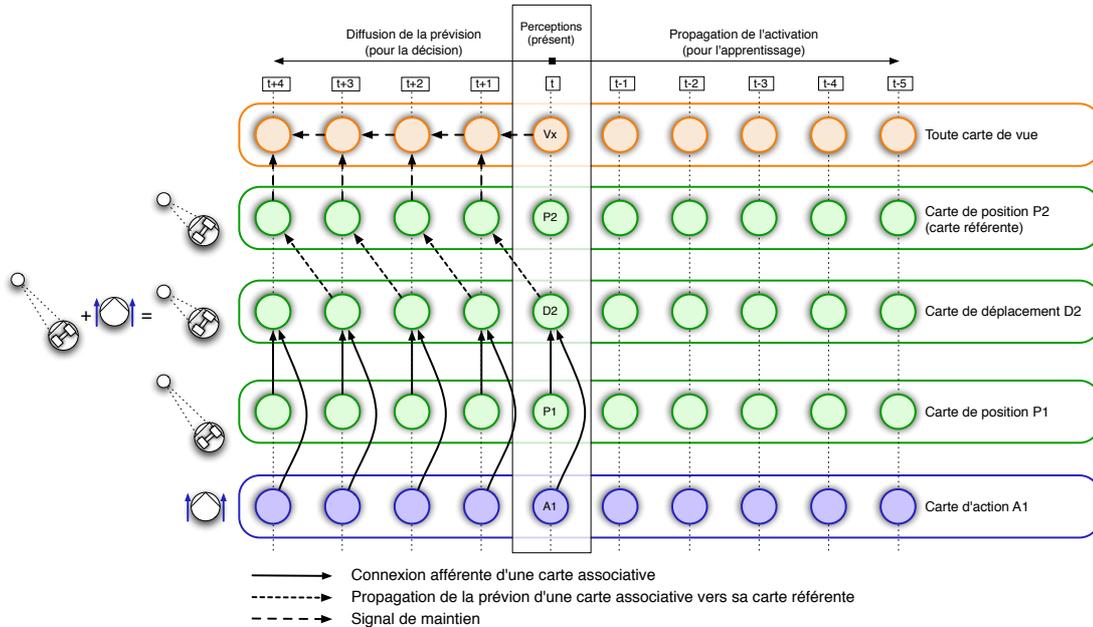


FIG. 5.2 – Représentation plane des connexions pour la prévision d'un *déplacement*.

La propagation des activations neuronales de prévision est obtenue par la création de connexions reliant toutes les cartes d'un groupe associatif à leur carte référente. Ces connexions sont décalées d'un rang de telle sorte que l'activation d'une association à l'instant t provoque l'excitation d'un neurone dans la carte référente au rang $t+1$ (Figures 5.2 et 5.3). Ces connexions sont construites lors de la création de chaque carte associative autorisant par la suite une propagation passive des activations sans que le parcours d'une base de règle ne soit nécessaire. Ces connexions sont représentées dans les Figures 5.2 et 5.3 par les flèches reliant respectivement les neurones des cartes D2 et P2 d'une part et T2 et V2 d'autre part. Les flèches en trait plein symbolisent les connexions subsistantes à l'issue de la phase de simplification du réseau présenté dans la section 4.7.

B. RÉFLEXES

À la différence des associations de *déplacement* et de *transformation*, les associations *réflexes* ne nécessitent pas la mise en place d'un décalage de rang pour les connexions à destination des cartes référentes (Figure 5.4). L'activation induite par une carte *réflexe* dans une carte d'action référente doit en effet se situer au même rang que les activations des neurones des cartes de vue et de positions ayant engendré l'activation du *réflexe*

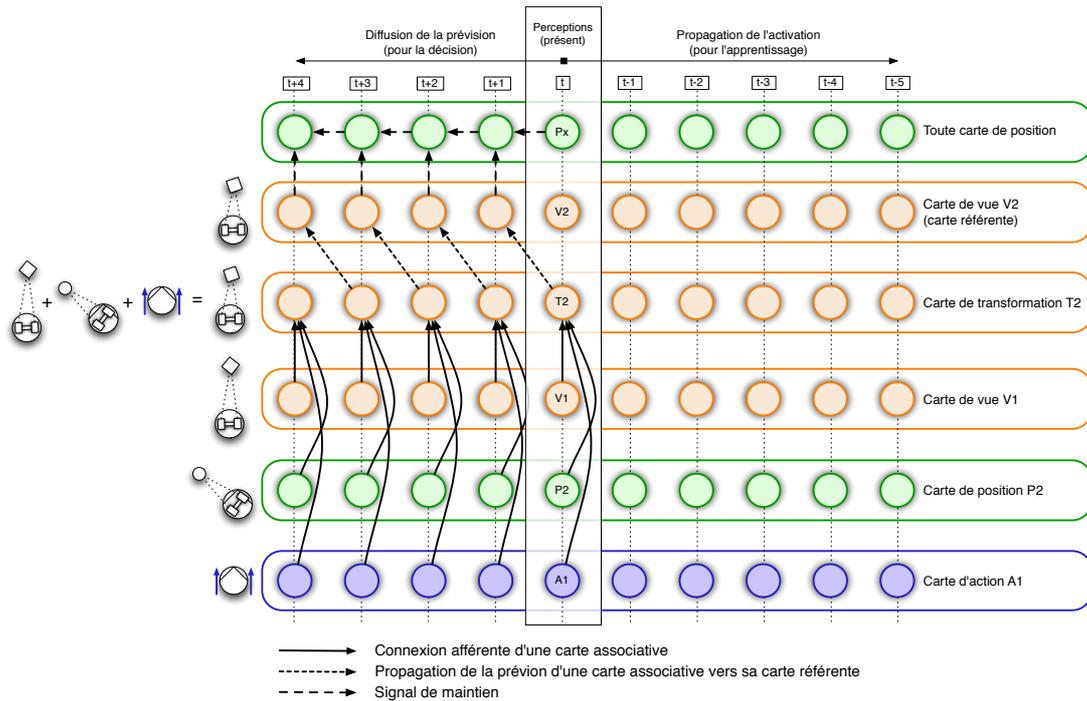


FIG. 5.3 – Représentation plane des connexions pour la prévision d'une *transformation*.

de sorte que les associations de *transformation* et de *déplacement* puissent par la suite s'appliquer. Les Figures 5.2 et 5.3 illustrent cette absence de décalage de rang pour les connexions reliant les neurones associatifs D2 et T2 et l'ensemble de leurs neurones afférents.

III. Connexions de maintien d'activation

Pour qu'une association *réflexe* puisse être activée, il est nécessaire que les préconditions de l'action correspondante soient actives simultanément au même rang. Lorsqu'il s'agit des perceptions (rang t), il suffit que ces préconditions soient effectivement perçues à l'instant présent, activant ainsi l'association *réflexe* correspondante et, par propagation, l'action associée au rang t . Cette co-activation n'est cependant pas aussi directe pour les rangs supérieurs. Prenons l'exemple de la prévision d'une transition de position au rang $t+1$ (Figure 5.5). L'activation d'une carte de position à ce rang signifie que le système prévoit que le déplacement du robot est sur le point d'engendrer une transition dans les champs de position. Or cette transition ne concernera, si elle se révèle exacte, que la position, la vue actuelle de l'objet restant inchangée. Il est donc nécessaire de reporter la vue qui restera inchangée (ou la position dans le cas d'une *transformation*). Pour les neurones de la carte de vue, cela revient à maintenir l'activation d'un rang donné à un rang supérieur par le biais de connexions de maintien d'activation. Ces connexions sont de deux types : un premier type relie les cartes de vue ou de position aux cartes susceptibles de nécessiter un maintien d'activation, les cartes

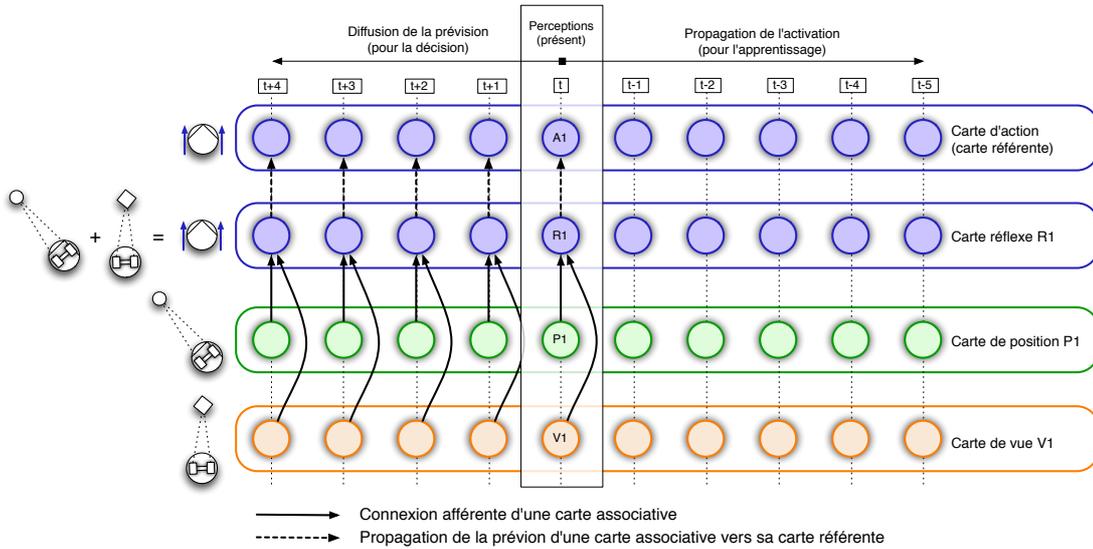


FIG. 5.4 – Représentation plane des connexions pour la prévision d'un *réflexe*.

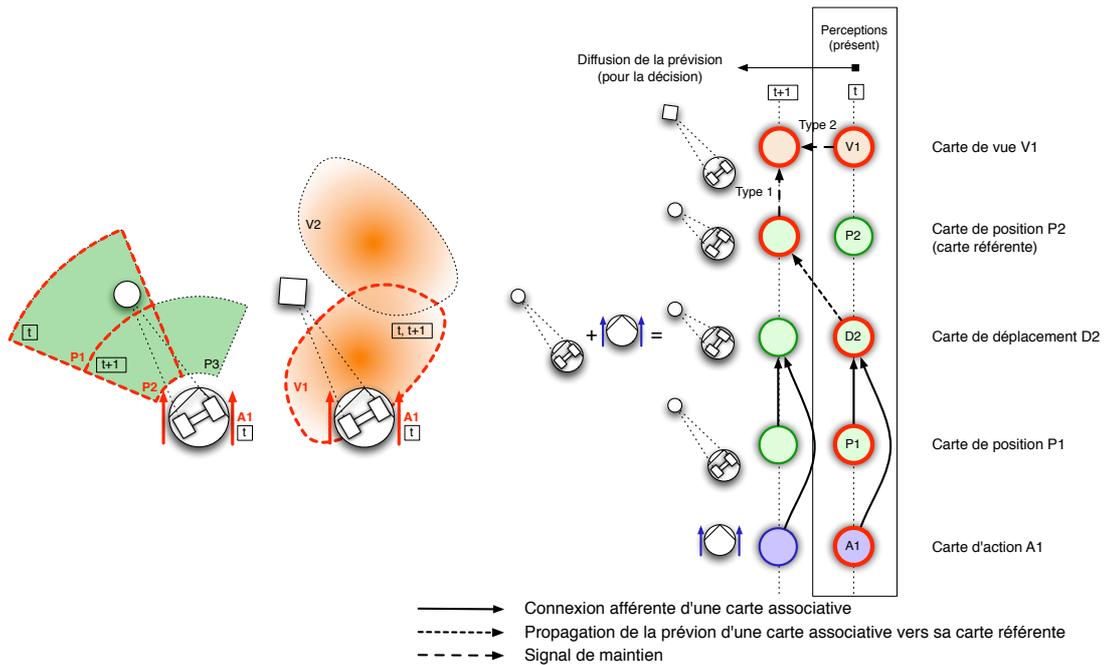


FIG. 5.5 – Maintien de l'activation d'une carte de vue V1 par la prévision d'un *déplacement* D2 vers une position P2

de position seront donc connectées à toutes les cartes de vue existantes et les cartes de vue à toutes les cartes de positions. Un deuxième type de connexions relie dans chaque carte de position ou de vue à partir du rang t chaque neurone au neurone de rang

immédiatement supérieur afin de propager d'un rang l'activation si une connexion du premier type a été activée. Nous montrons dans la Figure 5.5 le maintien d'activation d'une carte de vue V1 au rang $t + 1$ par l'activation simultanée de cette vue au rang t (connexion de type 2) et l'activation de la carte de position P2 au rang $t + 1$ (connexion de type 1).

IV. Cohérence de l'activation

La représentation distribuée de l'horizon de prévision sous la forme de cartes temporisées étendues ne permet pas de représenter les relations de dépendance entre états et actions comme le pourrait un arbre de décision classique. En effet, l'activation simultanée d'une représentation de lieu et d'une représentation de vue aux mêmes rangs dans leurs cartes temporisées respectives ne signifie pas que ces deux états perceptifs peuvent être atteints simultanément par une même séquence d'action mais que ces deux états peuvent être atteints de façon indépendante en un même nombre d'actions, celles-ci pouvant être différentes. Or, pour qu'une action puisse s'appliquer il est nécessaire de vérifier avant toute propagation, la cohérence des préconditions, c'est-à-dire vérifier que le couple (position,vue) est le résultat d'une même séquence d'action. Le maintien de la cohérence est obtenu en mémorisant la liste des états et des actions successives ayant mené à l'activation d'un neurone associatif. Il s'agit donc de mémoriser le chemin suivi par les signaux de prévision dans la mémoire associative pour, dans un second temps comparer ces chemins afin de respecter la cohérence des états prédits. La liste état-action est construite incrémentalement à partir d'une liste initialement vide au rang t et selon le type de carte de la façon suivante :

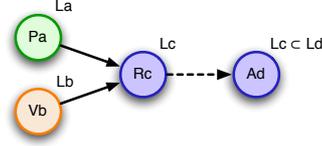
- cartes de position, de vue ou d'action : diffusion passive de la liste entrante.
- cartes *réflexes* : ajout de l'Action référente et diffusion de la liste.
- cartes de *déplacement* : ajout de la Position référente et diffusion de la liste.
- cartes de *transformation* : ajout de la Vue référente et diffusion de la liste.

Chaque vue, position ou action pouvant être activée par de multiples cartes associatives, c'est un ensemble de listes et non pas une liste unique qu'un neurone peut être amené à construire et/ou propager. Lors de l'activation d'une carte associative de *réflexe*, de *transformation* ou de *déplacement*, le maintien de la cohérence est obtenu en comparant les ensembles de listes de chaque neurone présynaptique au sein de deux phases distinctes de contrôle de la cohérence et de construction de l'ensemble des listes sortantes L_x .

Nous allons à présent détailler la construction et la propagation des ensembles de listes état-action pour chaque type de carte associative en introduisant quelques définitions et notations :

- \mathcal{M} : ensembles des cartes du réseau
- \mathcal{I} : ensemble des identifiants de cartes
- l : ensemble des listes l_i
- $Id : \mathcal{M} \rightarrow \mathcal{I}$: opérateur permettant d'obtenir l'identifiant d'une carte (avec $\forall(A, B) \in \mathcal{M} \times \mathcal{M}, A = B \Leftrightarrow Id(A) = Id(B)$)
- $add : \mathcal{I} \times l \rightarrow l$: opérateur d'ajout d'un identifiant à une liste

A. RÉFLEXE



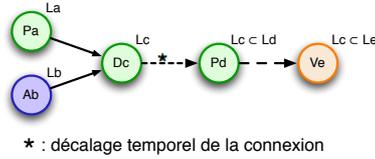
Condition d'activation

$$\exists(l_i, l_j) \in L_a \times L_b, l_i = l_j \tag{5.1}$$

Construction des ensembles de liste d'activation

$$\forall(l_i, l_j) \in L_a \times L_b, l_i = l_j \Leftrightarrow add(Id(A_d), l_i) \in L_c \tag{5.2}$$

B. DÉPLACEMENT



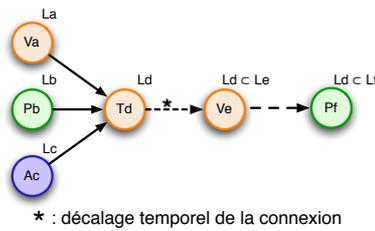
Condition d'activation

$$\exists(l_i, l_j) \in L_a \times L_b, l_j = add(Id(A_b), l_i) \tag{5.3}$$

Construction des ensembles de liste d'activation

$$\forall(l_i, l_j) \in L_a \times L_b, l_j = add(Id(A_b), l_i) \Leftrightarrow add(Id(P_d), l_j) \in L_c \tag{5.4}$$

C. TRANSFORMATION



Condition d'activation

$$\exists(l_i, l_j, l_k) \in L_a \times L_b \times L_c, l_k = \text{add}(\text{Id}(Ac), l_i) = \text{add}(\text{Id}(Ac), l_j) \quad (5.5)$$

Construction des ensembles de liste d'activation

$$\forall(l_i, l_j, l_k) \in L_a \times L_b \times L_c, l_k = \text{add}(\text{Id}(Ac), l_i) = \text{add}(\text{Id}(Ac), l_j) \Leftrightarrow \text{add}(\text{Id}(V_e), l_k) \in L_d \quad (5.6)$$

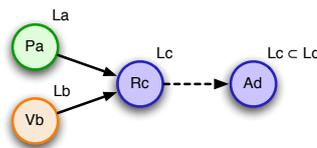
5.2 Propagation des gains et valeur d'utilité des association *réflexes*

À l'image de l'apprentissage des représentations, la construction de l'horizon de prévision est constituée d'une phase de propagation objective : la propagation des états et actions possibles, et d'une phase de propagation subjective : celle des valeurs de gain le long des chemins état-action prédits. Cette deuxième phase est rendue possible grâce au calcul des gains relatifs présentés lors du chapitre précédent. Nous avons en effet montré comment chaque association est capable d'apprendre la récompense induite lors de son activation. Nous souhaitons à présent exploiter ces valeurs de gain pour évaluer les valeurs d'utilité des associations *réflexes*. Soit Q_X la valeur d'utilité associée à l'association *réflexe* X qui, lors de la création de cette association est initialisée à la valeur 0. Nous notons G_X la valeur de gain associée à l'association X et $G_l(l)$ la valeur de gain cumulé associée à la liste d'activation l . A l'image de la méthode des différences temporelles les valeurs Q_X seront évaluées par une mise à jour locale. Toutefois nous allons exploiter pour cette mise à jour, la valeur des gains cumulés sur l'ensemble de la séquence prédite et non pas sur le seul état successeur :

$$Q_X(t+1) \leftarrow Q_X(t) + \alpha_Q [G_l(l_f) + Q_F(t) - Q_X(t)] \quad (5.7)$$

où α_Q est le coefficient d'apprentissage de la valeur d'utilité, l_f est une liste d'activation initiée par l'association *réflexe* X et F la dernière association *réflexe* prédite. La valeur de gain cumulé pour chaque chemin construit lors de la phase de propagation objective est donc égal à la somme de tous les gains relatifs de chaque association rencontrée lors de la prévision. Le calcul des gains cumulés est incrémental et a lieu pendant la construction des listes d'activation par les neurones associatifs *réflexes*, de *déplacement* et de *transformation*.

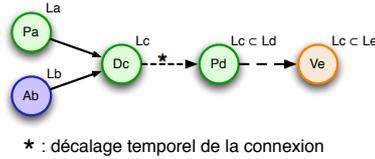
D. RÉFLEXE



Calcul du gain de la liste d'activation L_c

$$\forall l_i \in L_c \times L_b, l_i = l_j, \text{add}(Id(A_d), l_i) \in L_c \Rightarrow \mathcal{G}_l(\text{add}(Id(A_d), l_i)) = \mathcal{G}(R_c) + \mathcal{G}_l(l_i) \quad (5.8)$$

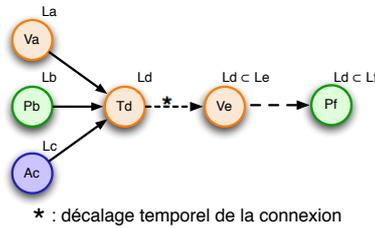
E. DÉPLACEMENT



Calcul du gain de la liste d'activation L_c

$$\forall (l_i, l_j) \in L_a \times L_b, \text{add}(Id(P_d), l_j) \in L_c \Rightarrow \mathcal{G}_l(\text{add}(Id(P_d), l_j)) = \mathcal{G}(D_c) + \mathcal{G}_l(l_j) \quad (5.9)$$

F. TRANSFORMATION



Calcul du gain de la liste d'activation L_d

$$\forall (l_i, l_j, l_k) \in L_a \times L_b \times L_c, \text{add}(Id(V_e), l_k) \in L_d \Rightarrow \mathcal{G}_l(\text{add}(Id(V_e), l_k)) = \mathcal{G}(T_d) + \mathcal{G}_l(l_k) \quad (5.10)$$

5.3 Propagation de la prévision : représentation perspective

Afin de permettre une meilleure compréhension des principes de propagation de la prévision au sein du réseau associatif, nous proposons de modifier le type de représentation graphique utilisée jusqu'à présent. Une représentation plane des cartes temporisées et de leurs connexions aboutit en effet à une représentation graphique difficilement compréhensible. C'est pourquoi nous avons choisi d'utiliser pour la suite de

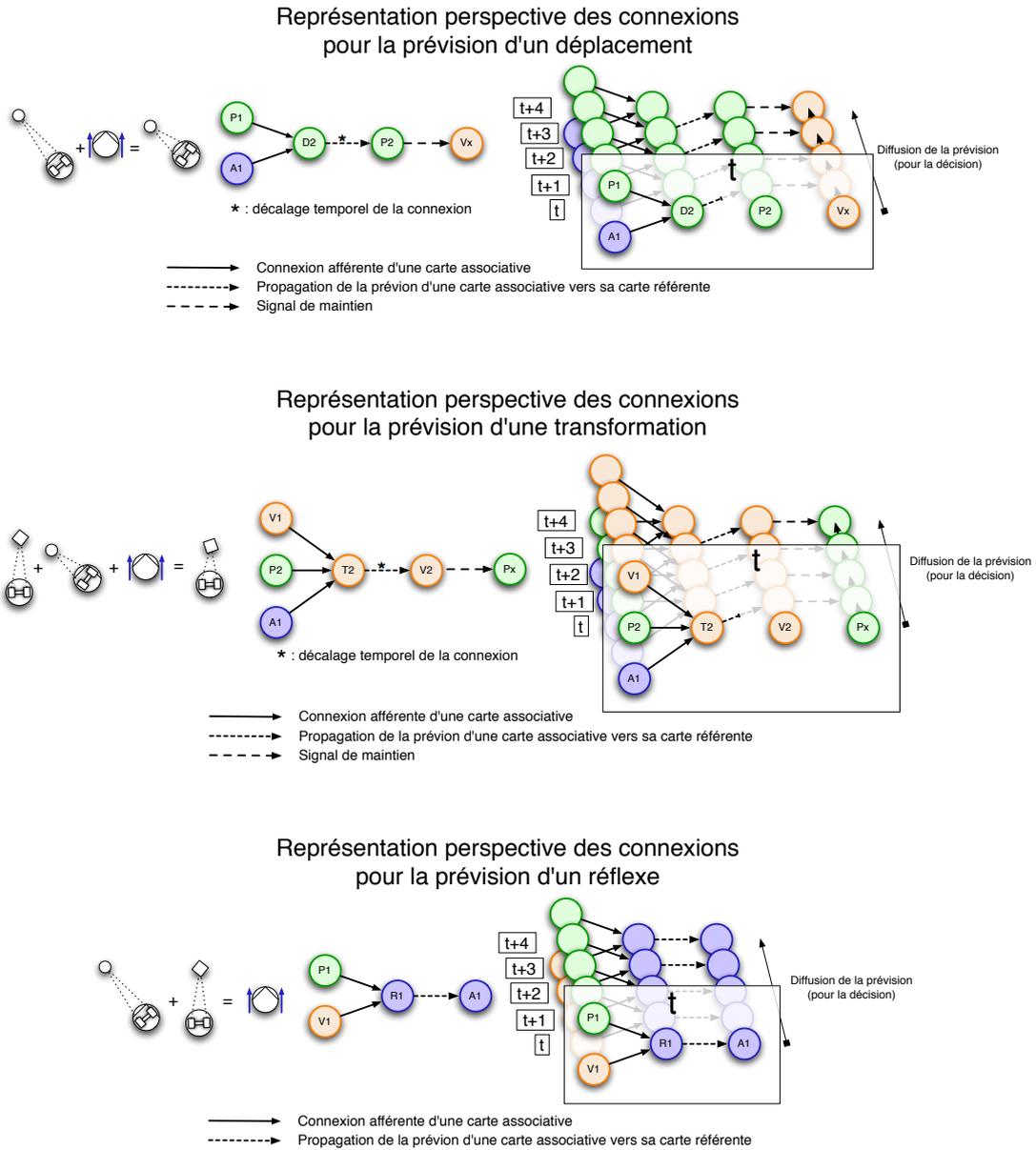


FIG. 5.6 – Représentation perspective des cartes temporisée étendue.

cet exposé une représentation perspective permettant de mieux distinguer les différents rangs de prévision ainsi que les connexions entre cartes. De plus, cette représentation ne rendra compte que de l'activité des neurones dédiés à la prévision c'est-à-dire ceux dont les rangs sont supérieurs ou égaux à t , les neurones restants n'étant utilisés que pendant l'apprentissage. La propagation de la prévision dans sa dimension temporelle est rendue possible grâce aux décalages de rangs des connexions entre les cartes de *transformation* et de *déplacement* et leur carte référente (Figure 5.3 et 5.2). Ce décalage temporel qui aura un rôle décisif lors de la propagation est symbolisé dans la représentation perspective par un décalage des connexions dans le sens de la profondeur (Figure 5.6). Dans cette représentation, les neurones sont en effet organisés en plans parallèles de profondeurs différentes symbolisant les différents rangs de prévision. Dans le cas des connexions entre une carte référente et une carte associative ou d'une carte *réflexe* vers son action référente les neurones présynaptiques et postsynaptiques sont situés à une même profondeur (neurones de même rang). Dans le cas des connexions entre une carte de *déplacement* ou de *transformation* vers une carte *réflexe*, l'activation sera propagée vers le plan de profondeur supérieure. De plus, chaque connexion de maintien d'activation (voir III.) reliant des neurones de même rang, sera supportée par un plan situé sur ce rang.

5.4 Le réseau associatif résultat de l'apprentissage

Nous allons considérer dans cette section un réseau associatif obtenu à l'issue de l'apprentissage d'un ensemble de représentations de vues, de position et d'action ainsi qu'un ensemble d'associations formées à l'aide de ces représentations. Nous notons :

- $V1$ et $V2$: deux représentations de vues.
- $P1$, $P2$ et $P3$: trois représentations de positions.
- $A1$ et $A2$: deux représentations d'actions.
- $T2$: une association de *transformation* de vue référente $V2$.
- $D2$ et $D3$: deux associations de *déplacement* de position référente respectives $P2$ et $P3$.
- $R1$ et $R2$: deux associations *réflexe* d'actions référentes respectives $A2$ et $A1$.

La Figure 5.7 illustre les cinq schémas d'associations apprises ainsi que le réseau associatif complet. Afin d'alléger la représentation graphique nous avons choisi de représenter les connexions de maintien d'activation dans une représentation perspective séparée.

5.5 Choix de l'action, cohérence de la prévision

Au moment de la prise de décision et grâce aux listes d'activations propagées par le réseau associatif, le système est capable de choisir l'action dont l'utilité est maximale après une phase de mise-à-jour s'appuyant sur le champ de prévision actuel. La propagation des activations dans la mémoire associative offre en effet une estimation de gain qui apparaît dans la mise à jour de la valeur d'utilité de l'action considérée : $\mathcal{G}_l(l_f) + \mathcal{Q}_F(t)$ dans l'équation 5.7. Ce gain correspond à la valeur de récompense cumulée maximale parmi les états prédit ayant atteint le rang le plus élevé à laquelle est ajouté l'utilité de la dernière association *réflexe* rencontrée. Notons que la profondeur du champ de prévision peut varier et dépend de différents facteurs qui seront décrits dans 5.7. Au moment où l'action courante a produit une transition, c'est-à-dire au moment de la prise de décision, toutes les associations *réflexes* candidates (celles qui sont actives au rang t) effectuent une mise à jour de leur valeur d'utilité Q_X selon l'équation 5.7 en exploitant les valeurs de gains des listes d'activation. Puis, l'association *réflexe* candidate ayant la plus forte valeur d'utilité est choisie et son action référente exécutée. Lorsque l'action est choisie, deux types d'opération doivent modifier la mémoire associative. Dans un premier temps, celle-ci doit extraire de l'horizon de prévision toute activité incohérente avec l'action choisie et ainsi conserver une partie de l'horizon actuel pour la prochaine phase de prévision. Cette étape est rendue possible par la simple lecture des listes d'activation de chaque neurone actif en ne conservant que les neurones dont les chemins états-actions commencent par l'action envisagée. Dans une deuxième phase, toutes les activations qui ont été conservées sont décalé d'un rang (vers la droite), ce qui traduit la consommation par le processus décisionnel d'un rang de prévision. Ainsi, toute prévision de rang supérieur à $t+1$ sera préservée et la construction de l'horizon de prévision pourra se poursuivre. Ce décalage de rang est associé à une modification des listes d'activations de sorte que l'action choisie est effacée de la première position de ces listes et que la valeur de gain associé à cette action soit soustraite de l'ensemble des gains cumulés.

5.6 Résolution d'une action, validité de la prévision

Dès lors qu'une action a été choisie, le système est en attente d'une transition d'état symbolisée par le déplacement du robot d'un champ perceptif (position ou vue) à un autre. Pendant ce temps d'attente, la mémoire associative est en mesure de poursuivre la construction de l'horizon de prévision sur la base de l'ensemble des neurones associatifs activés ou préservés lors de la phase d'action précédente. Cette construction incrémentale n'est possible que si 1) la fréquence de production de la prévision des actions est supérieure à celle de la décision et 2) si l'état prédit au moment de la résolution de l'action courante correspond à l'état perçu. Le premier point sera discuté par la suite. Quant à la vérification des données de la prévision, elle a lieu dès lors qu'une transition d'état est détectée (rang 0). Il est alors possible de comparer l'activation prévue (rang $t+1$) à l'activation présente parmi les cartes de position et de vues. Ainsi, si la prévision se révèle exacte une phase de décision classique (5.5) peut s'appliquer

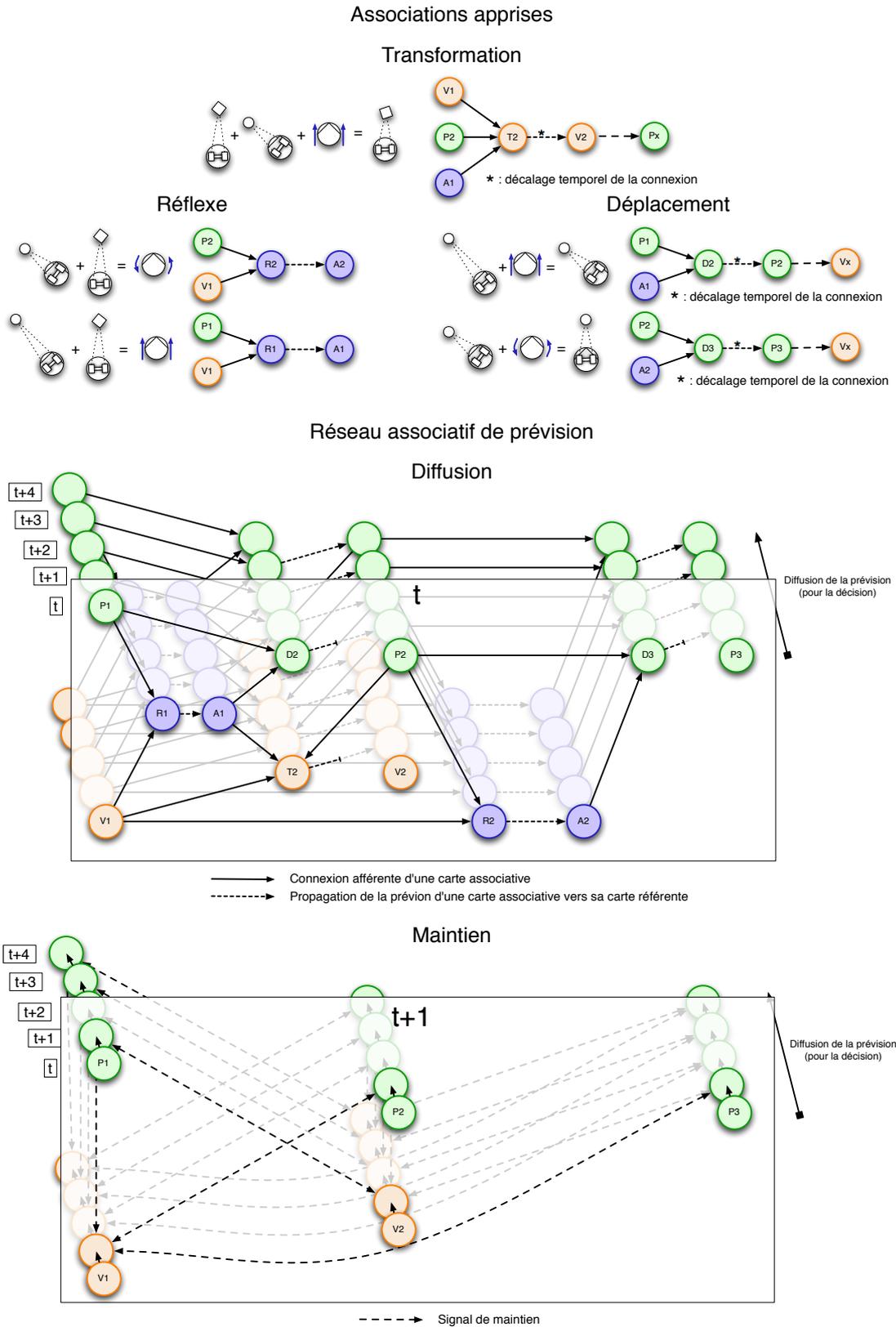


FIG. 5.7 – Réseau associatif complet issu de l'apprentissage de cinq associations.

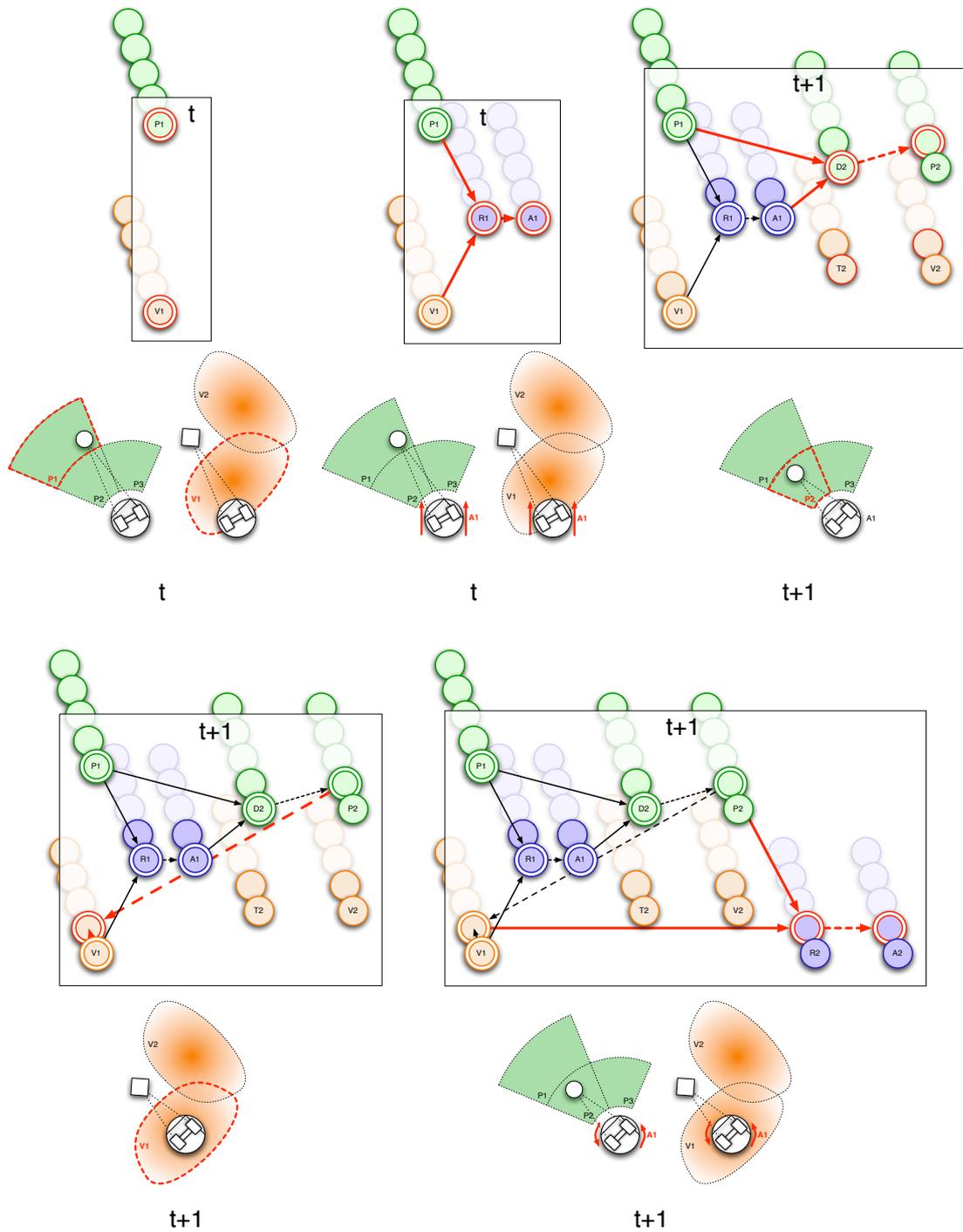


FIG. 5.8 – Propagation de la prévision dans la mémoire associative. Cette propagation utilise le réseau associatif illustré par la figure 5.7. Les neurones ainsi que les connexions affectées par la propagation à chaque pas de temps sont représentés en rouge. Les connexions de maintien d'activation sont représentées par des flèches en pointillés.

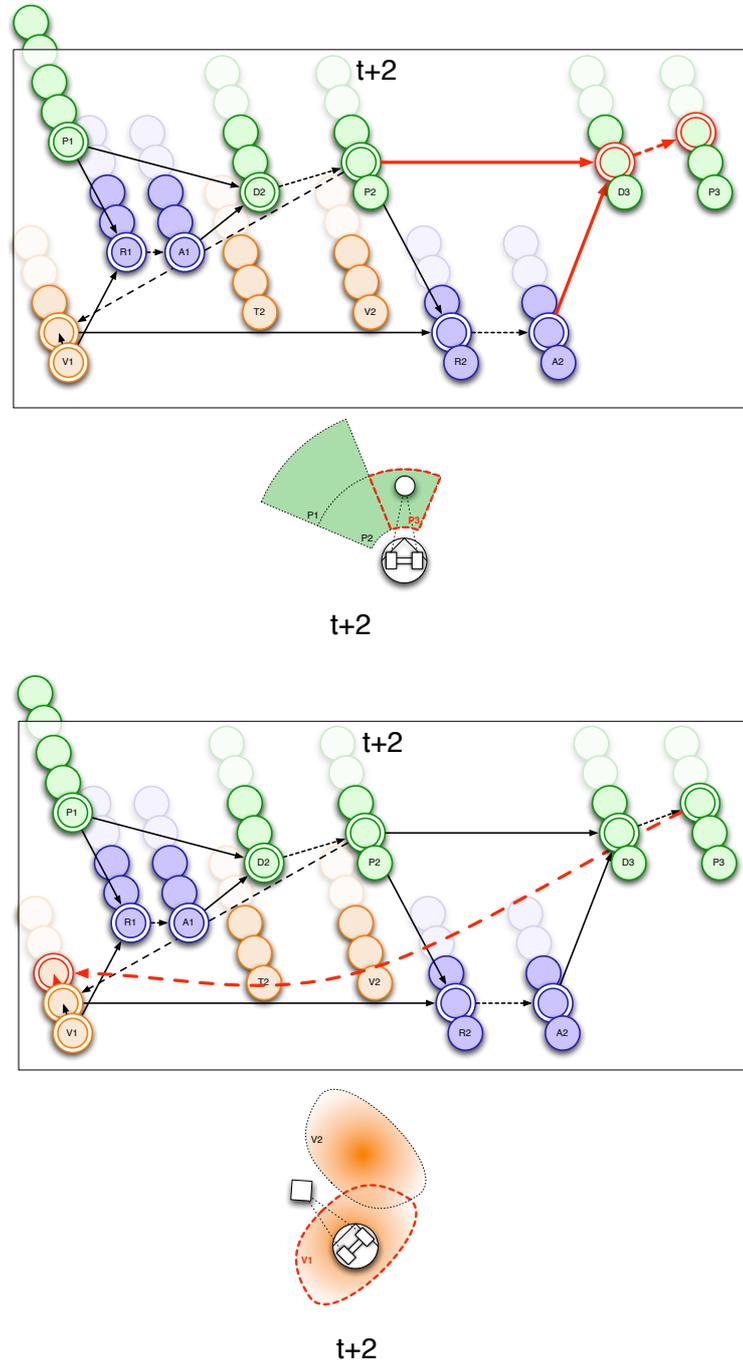


FIG. 5.9 – Propagation de la prévision dans la mémoire associative (suite).

sinon, l'ensemble des activations de la mémoire associative est invalidé ce qui se traduit par la désactivation de tous les neurones prédictifs. Cette situation peut être due par exemple à une association erronée par manque d'apprentissage ou bien à l'action de facteurs externes sur l'environnement (objets mobiles...).

5.7 Aspects temporels

Afin de rendre la prise de décision efficace, il est nécessaire de disposer de prévisions suffisamment riches lorsque l'action courante arrive à son terme. Du point de vue du système, l'activation d'une action à pour objectif de modifier le contexte perceptif courant en générant la transition entre deux champs de positions et/ou de vue voisins. Or, il existe un intervalle de temps durant lequel l'action demeure active et le contexte perceptif reste inchangé (5.6). Ce temps "d'attente" est le résultat de la formation de catégories perceptives stables et sa durée dépend de la taille des champs de lieu et de vues appris par le système ainsi que de la vitesse de déplacement du robot au sein de ces champs (ici la vitesse réelle de déplacement du robot). C'est en fait grâce à l'existence de ce temps d'attente que la mémoire associative est en mesure de construire un horizon de prévision suffisamment riche au moment du choix de la prochaine action. Le temps d'attente observé lors d'expérience étant de l'ordre de la seconde, et notre simulateur disposant d'une fréquence de traitement d'environ 20 cycles d'intégration et de tir par seconde, il se produit en moyenne une vingtaine de propagation par seconde au sein du réseau associatif. Étant donnée l'implémentation actuelle du simulateur, la prévision d'une transition nécessite l'activation séquentielle d'un neurone associatif suivi de l'activation du neurone référent (de rang supérieur pour les associations de *déplacement* et de *transformations* et de même rang pour les associations *réflexes*) suivi si besoin (*déplacements* et *transformations*) de l'activation de la connexion de maintien permettant la "recopie" de la partie du contexte perceptif resté inchangé. Il en résulte que le nombre de cycles nécessaires à la prévision d'un nouveau contexte à partir d'un contexte initial est égal à 2 (association *réflexe*) + 3 (*déplacement* ou *transformation*) = 5 cycles d'intégration et de tir. Si l'on considère que la fréquence de traitement demeure constante quel que soit le nombre d'activations simultanées, le système peut donc en moyenne disposer à l'issue du temps d'attente d'un horizon de prévision dont la profondeur aura été augmentée de 4 niveaux en moyenne. Dans le cas où la prévision se sera révélée exacte, le système est en mesure de conserver les prévisions construites pendant le déroulement de l'action précédente permettant ainsi une construction incrémentale de l'horizon de prévision.

5.8 Conclusion

Nous avons vu comment sont exploitées, dans notre système, les associations sensori-motrices pour créer de façon incrémentale un réseau capable, par diffusion, d'anticiper les états accessibles par le robot. Cette diffusion exploite l'ensemble des algorithmes locaux de nos unités de traitement d'une part mais elle est d'autre part à l'origine du

calcul des valeurs d'utilité associée à chaque association *réflexes*. En effet, l'objectif n'est pas ici d'explorer l'horizon des états possibles pour en extraire un but mais de mettre à jour les valeurs d'utilité pour fournir au processus décisionnel les moyens d'un choix rapide de la meilleure action à accomplir. L'originalité de notre approche est donc de mêler l'évaluation subjective (des gains et des utilités) à l'apprentissage des représentations qui, dans le cadre d'un apprentissage par renforcement classique, font partie des données du problème.

CONCLUSION ET PERSPECTIVES

Les travaux menés durant cette thèse ont eu pour objectif de faire évoluer un système d'apprentissage essentiellement perceptif vers un système de contrôle global du comportement d'un robot mobile. Pour mener à bien ces travaux nous nous sommes donc dans un premier temps plongés dans les travaux préalablement entrepris par William Paquier dans sa thèse intitulée "Apprentissage ouvert de représentations et de fonctionnalités en robotique : analyse, modèles et implémentation" [Paquier 04]. Ces travaux constituent d'une part une analyse des composantes nécessaires à un apprentissage ouvert ainsi qu'une version initiale du simulateur de réseaux de neurones appelé NeuSter, ayant conduit aux premiers résultats de reconnaissance de motifs visuels. Ce manuscrit décrit d'autre part l'utilisation des techniques d'apprentissage par renforcement au sein d'une boucle sensori-motrice dans laquelle la mémoire associative occupe une place centrale. Tout en faisant évoluer le système existant, nous nous sommes donc attachés à développer puis à mettre en œuvre les mécanismes associatifs et décisionnels de ce système en l'intégrant à la plate-forme robotique.

Le premier développement apporté au simulateur, situé au cœur du fonctionnement des neurones, a consisté au passage d'une phase de tir déterministe à une phase de tir stochastique où la valeur de seuil suit une distribution de probabilités qui évolue dynamiquement lors de l'apprentissage. Cette modification a permis de réduire les phénomènes de sur-apprentissage en permettant à l'apprentissage d'avoir lieu en présence de motifs plus éloigné du motif idéal que dans le cas déterministe.

Lors de l'intégration du système à la plate-forme robotique, nos travaux se sont portés sur la reconnaissance de formes au sein d'images naturelles. Les expériences menées dans ce domaine ont montré la nécessité d'adapter la taille des matrices de poids synaptiques à celles des motifs observés. Nous avons ainsi développé les mécanismes d'expansion et de rétraction de ces matrices en exploitant la stéréovision pour identifier les contours des objets.

Ces premières modifications du modèle ont été par la suite exploitées pour permettre un apprentissage autonome de représentations d'objets. Ces travaux présentés dans [Do Huu 05], introduisent les notions de cartes de vues et de carte tracker ainsi que les

mécanismes d'interaction les reliant afin de permettre un apprentissage non-supervisé et le regroupement des vues par cohérence spatio-temporelle : deux vues voisines d'un objet tendent à apparaître proche dans l'espace et dans le temps.

Afin de permettre l'apprentissage des motifs de position et d'action, nous avons introduit la notion de cartes neuronales temporisées. Ces cartes, qui conservent la trace des activations passées en les propageant à chaque pas du simulateur, sont devenues dans un second temps le support de l'apprentissage de séquences d'activations pour l'apprentissage associatif. La mémoire associative, telle que nous l'avons développée est constituée de trois types d'associations : les *déplacements*, les *transformations* et les *réflexes*. L'apprentissage de chaque association est piloté par l'activité de sa carte référente de position, de vue ou d'action. Au terme de l'apprentissage associatif, le système possède une représentation de la dynamique de ses interactions avec l'environnement sous la forme d'un réseau associatif construit de façon incrémental.

Enfin, l'apprentissage de la composante subjective, c'est-à-dire l'association d'une variation de récompense avec chaque association, couplé à la propagation de la prévision au sein des cartes temporisées étendues, constitue une méthode originale d'évaluation des valeurs d'utilité pour les actions du robot. Cette approche, largement inspirée des techniques d'apprentissage par renforcement, offre la possibilité d'une prise de décision rapide sur la base de l'évaluation d'un gain cumulé espéré.

Les principales perspectives de recherche qui apparaissent à l'issue de cette thèse concernent tout d'abord l'amélioration de la phase de décomposition des images pour la reconnaissance d'objets. Afin de fournir une représentation plus riche des motifs visuels, il serait intéressant d'adjoindre à la détection d'orientation locales celle d'angles orientés, de terminaisons de traits et de contrastes de couleur.

La prise en compte de multiples objets dans la scène visuelle grâce à un moteur de saccades permettrait en outre de construire une représentation plus évoluée de l'espace sous la forme de cellule de lieux sensibles à un environnement d'objets particuliers. Cette approche nécessiterait cependant de conserver en mémoire les associations entre un objet et sa position dans la scène. Une piste de réflexion déjà proposée dans la thèse de W. Paquier serait d'adjoindre au fonctionnement du neurone une valeur supplémentaire : la phase, qui permettrait de relier l'objet et la position grâce au partage d'une valeur commune.

L'apprentissage des représentations d'actions tel que nous l'avons envisagé est directement issu de l'observation de données proprioceptives. Grâce à l'utilisation des cartes neuronales temporisées, il serait envisageable d'apprendre des représentations de séquences d'actions élémentaires exhibant des valeurs d'utilité satisfaisantes. Cette démarche permettrait en outre de limiter le nombre de neurones nécessaires à la prédiction et d'accélérer l'évaluation des valeurs d'utilité en propageant en un seul pas de temps un ensemble d'action et les valeurs de gains cumulés correspondants.

Pour conclure, nous avons souhaité dans cette thèse montrer que l'apprentissage, plus qu'une fonction isolée de l'architecture robotique, peut être exploité pour former des représentations mais aussi construire et structurer le processus de contrôle.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [Arleo 99] A. Arleo, J. del R. Millán & D. Floreano. *Efficient learning of variable-resolution cognitive maps for autonomous indoor navigation*. IEEE Trans. Robot. Automat., vol. 15, pages 990–1000, 1999.
- [Arleo 00a] A. Arleo & W. Gerstner. *Modeling Rodent Head-direction Cells and Place Cells for Spatial Learning in Bio-mimetic Robotics*. Dans Floreano D. Roitblat H.L. Meyer J.A. Berthoz A. & Wilson S.W., éditeurs, *From Animals to Animats VI*, pages 236–245. MIT Press, Cambridge MA, 2000.
- [Arleo 00b] A. Arleo & W. Gerstner. *Spatial cognition and neuro-mimetic navigation : A model of hippocampal place cell activity*. Dans Biological Cybernetics, 2000.
- [Arleo 01] A. Arleo & W. Gerstner. *Spatial Orientation in Navigating Agents : Modeling Head-direction Cells*. Neurocomputing, vol. 38-40, pages 1059–1065, 2001.
- [Arleo 04] A. Arleo, F. Smeraldi & W. Gerstner. *Cognitive Navigation Based on Nonuniform Gabor Space Sampling, Unsupervised Growing Networks, and Reinforcement Learning*. Dans IEEE Trans Neural Networks, volume 15, pages 639–52, 2004.
- [Bellman 57] R. Bellman. *Dynamic programming*. Princeton University Press, 1957.
- [Biederman 87] I. Biederman. *Recognition-by-components : a theory of human image understanding*. Psychological Review, vol. 94, pages 115–147, 1987.
- [Biederman 93] I Biederman & P. C. Gerhardstein. *Recognizing depth-rotated objects : evidence and conditions for three-dimentionnal viewpoint*

- invariance*. Journal of Experimental Psychology : Human Perception and Performance, vol. 19, no. 6, pages 1162–1182, 1993.
- [Biederman 00] I. Biederman. *Recognizing depth-rotated objects : A review of recent research and theory*. Spatial Vision, vol. 13, no. 2-3, pages 241–253(13), 2000.
- [Bliss 93] T.V. Bliss & G.L. Collingridge. *A synaptic model of memory : long-term potentiation in the hippocampus*. Nature, vol. 361, no. 6407, pages 31–39, 1993.
- [Bülthoff 94] H. Bülthoff, S. Edelman & M. Tarr. *How Are Three-Dimensional Objects Represented in the Brain ?* Rapport technique, Massachusetts Institute of Technology Artificial Intelligence Laboratory and Center for Biological and Computational Learning Department of Brain and Cognitive Sciences, 1994.
- [Chatila 85] R. Chatila & J. P. Laumond. *Position referencing and consistent world modeling for mobile robots*. Dans IEEE Int. Conf. Robotics Automation, 1985.
- [Cuperlier 06] N. Cuperlier, M. Quoy, C. Giovannangeli, Ph. Gaussier & Ph. Laroque. *Transition Cells for navigation and planning in an unknown environment*. Dans The Society For Adaptive Behavior SAB'2006, pages 286–297, [http ://publi-etis.ensea.fr/2006/CQGGL06](http://publi-etis.ensea.fr/2006/CQGGL06).
- [Do Huu 05] N. Do Huu, W. Paquier & R. Chatila. *Combining structural descriptions and image-based representations for image object and scene recognition*. Dans International Joint Conference on Artificial Intelligence (IJCAI), 2005.
- [Elfes 87] A. Elfes. *Sonar-based real-world mapping and navigation*. Dans IEEE Journal of Robotics and Automation, volume 3, pages 249–265, 1987.
- [Elfes 89] A. Elfes. *Occupancy Grids : A Probabilistic Framework for Robot Perception and Navigation*. Thèse de doctorat, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1989.
- [Fregnac 88] Y. Fregnac, D. Shulz, S. Thorpe & E. Bienenstock. *A cellular analogue of visual cortical plasticity*. Nature, vol. 333, no. 6171, pages 367–370, 1988.
- [Fukushima 03] K. Fukushima. *Neocognitron for handwritten digit recognition*. Neurocomputing, vol. 51, pages 161–180, April 2003.
- [Gaussier 02] P. Gaussier, A. Revel, J.P. Banquet & V. Badeau. *From view cells and place cells to cognitive map learning : processing stages of the hippocampal system*. Biological cybernetics, vol. 86, no. 1, pages 15–28, 2002.

- [Georges-François 99] P. Georges-François, E.T. Rolls & R.G. Robertson. *Spatial view cells in the primate hippocampus : allocentric view not head direction or eye position or place*. Cerebral Cortex, vol. 9, pages 197–212, 1999.
- [Gerstner 95] W. Gerstner. *Time structure of the activity in neural network models*. Phys. Rev. E, vol. 51, pages 738–758, 1995.
- [Gibson 79] James J. Gibson. *The ecological approach to visual perception*. Lawrence Erlbaum Associates, 1979.
- [Hebb 49] Donald O. Hebb. *The organization of behavior : A neuropsychological theory*. Wiley, New York, June 1949.
- [Ito 93] M. Ito. *Synaptic plasticity in the cerebellar cortex and its role in motor learning*. The Canadian journal of neurological sciences, vol. 20, no. 3, pages S70–S74, 1993.
- [Koch 92] c. Koch & T. Poggio. *Multiplying with synapses and neurons*. Dans Academic Press, éditeur, Single neuron computation, pages 315–345. Academic Press Professionnal, Inc., 1992.
- [Kohonen 77] T. Kohonen. *Associative memory : A system-theoretical approach*. Springer-Verlag, Berlin, 1977.
- [Kohonen 89] T. Kohonen. *Self-organization and associative memory*. Springer-Verlag New York, Inc., 1989.
- [Kortenkamp 94] D. Kortenkamp & T. Weymouth. *Topological mapping for mobile robots using a combination of sonar and vision sensing*. Dans the Twelfth National Conference on Artificial Intelligence, pages 979–984, 1994.
- [Kuipers 78] B. Kuipers. *Modeling Spatial Knowledge*. Cognitive Science, vol. 2, pages 129–153, 1978.
- [Kuipers 91] B. Kuipers & Y.T. Byun. *A Robot Exploration and Mapping Strategy Based on a Semantic Hierarchy of Spatial Representations*. Dans Robot. Autonomous Syst, volume 8, pages 47–63, 1991.
- [Lapicque 07] L. Lapicque. *Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation*. J. Physiol. Pathol. Gen, no. 9, pages 620–635, 1907.
- [Leprêtre 00] S. Leprêtre, P. Gaussier & J.P. Cocquerez. *From Navigation to Active Object Recognition*. Dans Simulation of Adaptive Behavior, SAB'2000, 2000.
- [Maass 96] W. Maass. *Lower Bounds for the Computational Power of Networks of Spiking Neurons*. Neural Computation, vol. 8, no. 1, pages 1–40, 1996.
- [McCulloch 43] W.S. McCulloch & W. Pitts. *A Logical Calculus of the Ideas Immanent in Nervous Activity*. Bulletin of Mathematical Biophysics, vol. 5, pages 115–133, 1943.

- [Minsky 69] M. Minsky & S. Papert. *Perceptrons : An introduction to computational geometry*. MIT Press, 1969.
- [Moravec 85] H. P. Moravec & A. Elfes. *High resolution maps from wide angle sonar*. Dans IEEE Int. Conf. Robotics Automation, pages 116–121, 1985.
- [Moravec 88] H. P. Moravec. *Sensor fusion in certainty grids for mobile robots*. A.I. Magazine, 1988.
- [O’Keefe 71] J. O’Keefe & J. Dostrovsky. *The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat*. Brain Research, vol. 34, no. 1, pages 171–175, 1971.
- [O’Keefe 78] J. O’Keefe & L. Nadel. *The hippocampus as a cognitive map*. Oxford University Press, 1978.
- [O’Keefe 96] J. O’Keefe & N. Burgess. *Geometric determinants of the place fields of hippocampal neurons*. Nature, vol. 381, pages 425–428, 1996.
- [Paquier 04] W. Paquier. *Apprentissage ouvert de représentations et de fonctionnalités en robotique : analyse, modèles et implémentation*. Thèse de doctorat, Université Paul Sabatier de Toulouse, 2004.
- [Poggio 90] T. Poggio & S. Edelman. *A Network That Learns to Recognize Three Dimensional Objects*. Letters to Nature, vol. 343, pages 263–266, 1990.
- [Redish 97a] A.D. Redish & D.S. Touretzky. *Cognitive maps beyond the hippocampus*. Hippocampus, vol. 7, pages 15–35, 1997.
- [Redish 97b] A.D. Redish & D.S. Touretzky. *Navigating with landmarks : Computing goal locations from place codes*. Dans Ikeuchi K. & Veloso M., éditeurs, Symbolic Visual Learning, chapitre 12, pages 325–351. Oxford University Press, 1997.
- [Riesenhuber 99] M. Riesenhuber & T. Poggio. *Hierarchical models of object recognition in cortex*. Nature Neuroscience, vol. 28, pages 1019–1025, 1999.
- [Riesenhuber 00] M. Riesenhuber & T. Poggio. *Models of Object Recognition*, 2000.
- [Ritz 97] R. Ritz & T.J. Sejnowski. *Synchronous oscillatory activity in sensory systems : new vistas on mechanisms*. Current Opinion in Neurobiology, vol. 7, no. 4, pages 536–546, 1997.
- [Rolls 95] E.T. Rolls & S.M. O’Mara. *View-responsive neurons in the primate hippocampal complex*. Hippocampus, vol. 5, no. 409–424, 1995.
- [Rolls 98] E.T. Rolls & A. Treves. *Neural networks and brain function*. Oxford University Press, 1998.
- [Rolls 99] E. T. Rolls. *Spatial View Cells and the Representation of Place in the Primate Hippocampus*. Hippocampus, vol. 9, pages 467–480, 1999.

- [Rolls 04] E. T. Rolls, J. Xiang & L. Franco. *Object, Space, and Object-Space Representations in the Primate Hippocampus*. Journal of Neurophysiology, vol. 94, pages 833–844, 2004.
- [Rolls 05] E. T. Rolls & J. Xiang. *Reward-Spatial View Representations and Learning in the Primate Hippocampus*. The Journal of Neuroscience, vol. 25, no. 26, pages 6167–6174, 2005.
- [Rolls 06] E.T. Rolls & J. Xiang. *Spatial view cells in the primate hippocampus and memory recall*. Reviews in neurosciences, vol. 17, no. 1-2, pages 175–200, 2006.
- [Rosenblatt 62] F. Rosenblatt. Principles of neurodynamics; perceptrons and the theory of brain mechanisms. Washington, Spartan Books, 1962.
- [Rumelhart 86] D. E. Rumelhart, G. E. Hinton & R. J. Williams. *Learning Internal Representations by Error Propagation*. Dans D. E. Rumelhart & J. L. McClelland, éditeurs, Parallel Distributed Processing, volume 1, pages 318–362. MIT Press, 1986.
- [Sharp 91] P.E. Sharp. *Computer simulation of hippocampal place cells*. Psychobiology, vol. 19, no. 2, pages 103–115, 1991.
- [Shawe-Taylor 91] J. Shawe-Taylor, P. Jeavons & M. Daalen. *Probabilistic Bit Stream Neural Chip : Theory*. Connection Science, vol. 3, no. 3, pages 317–328, 1991.
- [Soodak 91] R.E. Soodak. *Reverse-Hebb plasticity leads to optimization and association in a simulated visual cortex*. Vis Neurosci., vol. 6, no. 5, pages 507–518, 1991.
- [Srinivasan 76] M.V. Srinivasan & G.D. Bernard. *A proposed mechanism for multiplication of neural signals*. Biological Cybernetics, vol. 21, no. 4, pages 227–36, 1976.
- [Sutton 98] Richard S. Sutton & Andrew G. Barto. Reinforcement learning : An introduction (adaptive computation and machine learning). The MIT Press, March 1998.
- [Tarr 95] M. Tarr & H. Bülthoff. *Is human object recognition better described by geon structural description or by multiple views ? Comment on Biedermann and Gerhardstein (1993)*. Journal of Experimental Psychology : Human Perception and Performance, vol. 21, pages 1494–1505, 1995.
- [Tarr 98] M. Tarr & H. Bülthoff. *Image-based object recognition in man, monkey and machine*. Cognition, no. 67, pages 1–20, 1998.
- [Thrun 98] S. Thrun. *Learning maps for indoor mobile robot navigation*. Artificial Intelligence, vol. 89, pages 21–71, 1998.
- [Touretzky 96] D.S. Touretzky & A.D. Redish. *A theory of rodent navigation based on interacting representations of space*. Hippocampus, vol. 6, no. 3, pages 247–270, 1996.

- [Tuckwell 88] H. Tuckwell. Introduction to theoretical neurobiology, volume 1, 2. Cambridge University Press, 1988.
- [Wallis 96a] G. Wallis. *How neurons learn to associate 2D-views in invariant object recognition*. Rapport technique 37, Max-Planck-Institute für Biologische Kybernetik, Tübingen, Germany, 1996.
- [Wallis 96b] G. Wallis. *Using spatio-temporal correlations to learn invariant object recognition*. Neural Networks, vol. 9, no. 9, pages 1513–1519, 1996.
- [Wan 94] H.S. Wan, D.S. Touretzky & A.D. Redish. *Towards a Computational Theory of Rat Navigation*. Dans Proceedings of the 1993 Connectionist Models Summer School, pages 11–19, 1994.
- [y Cajal 11] Santiago Ramón y Cajal. *Histologie du système nerveux de l'homme et des vertébrés ; vol 1 et 2*. A. Maloine Paris, 1911.
- [Yamauchi 96] B. Yamauchi & R. Beer. *Spatial learning for navigation in dynamic environments*. IEEE Transactions on Systems, Man, and Cybernetics - Part B : Cybernetics. Special Issue on Learning Autonomous Robots, 1996.
- [Yuille 89] A.L. Yuille, D.M. Kammer & D.S. Cohen. *Quadrature and the development of orientation selective cortical cells by Hebb rules*. Biological Cybernetics, vol. 61, no. 3, pages 183–194, 1989.

ASSEMBLÉE : CRÉATION ET MONITORING DE RÉSEAUX

L'objectif de cette application est de permettre à l'expérimentateur de bénéficier d'un outil de construction et de monitoring de réseaux de neurones efficace, basés sur une représentation graphique simplifiée. Lors de la phase de construction, une bibliothèque de blocs préconstruits permet à l'utilisateur de glisser dans l'espace de travail un ensemble d'éléments du réseaux. L'utilisateur peut alors connecter dans l'espace de travail ces différents éléments tandis que la structure réelle du réseau est mémorisée en arrière plan. Pour faciliter le déploiement du réseau ainsi construit parmi les calculateurs du réseaux local, nous nous appuyons sur la technologie nommée "Zeroconf". Celle-ci désigne un ensemble de protocole réseaux permettant la découverte automatique de services, sans configuration préalable ni information centralisée au sein d'un réseau local. Les machines proposant le service de simulation neuronale sont ainsi visibles dans une fenêtre dédiée de sorte que l'utilisateur n'ait plus qu'à glisser-déposer une machine particulière sur un bloc afin de simuler sur celle-ci le réseau de neurone sous-jacent. Pendant le déroulement de la simulation, une fenêtre d'information permet de configurer localement les paramètres des cartes neuronales et des matrices de poids tout en visualisant l'ensemble des paramètres du modèle. Enfin, chaque réseau et l'ensemble de ses paramètres peuvent-être sauvegardés sous forme de document pour être réutilisé et/ou complété par la suite.

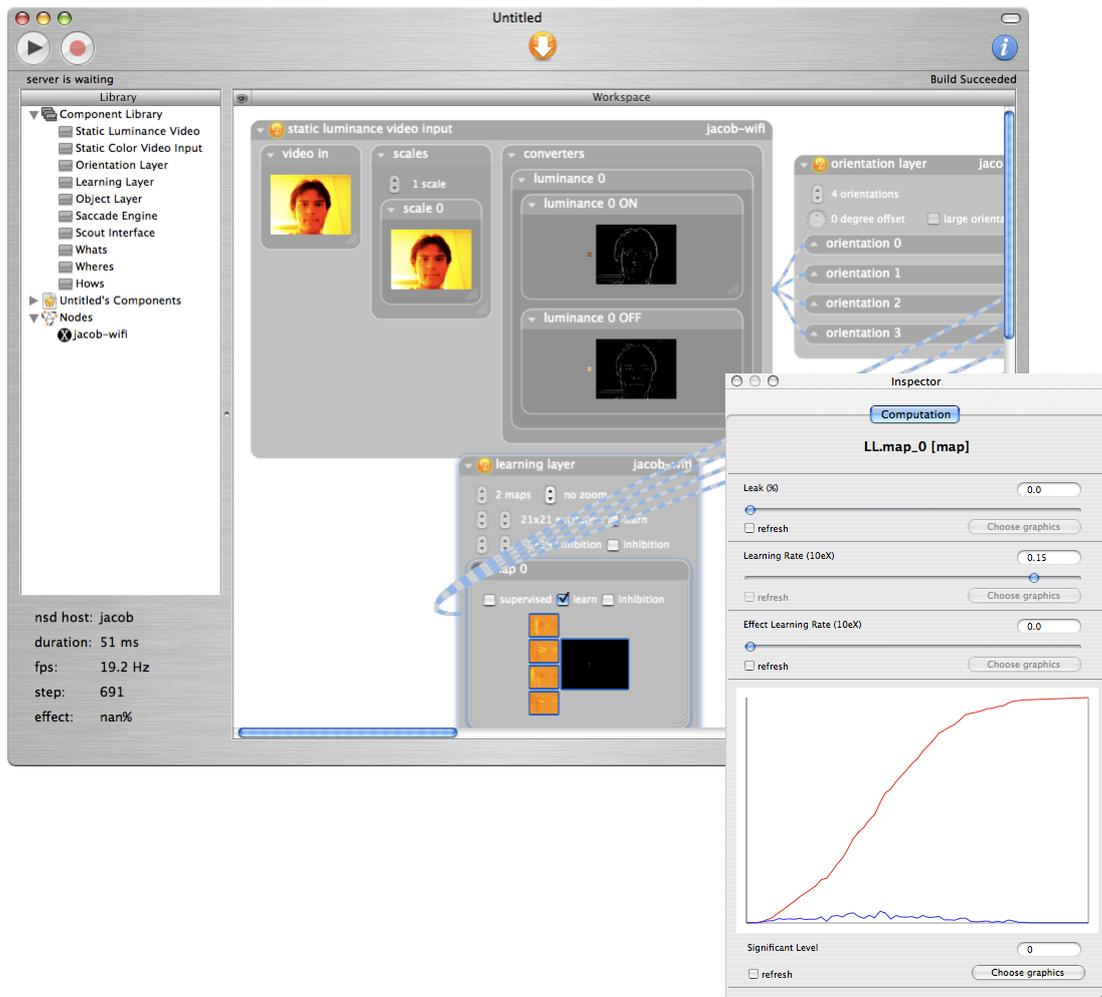


FIG. 1 – Assemblée : création et monitoring de réseaux.

NEUSTER : SIMULATION DE RÉSEAUX DE NEURONES ET CALCULS DISTRIBUÉS

Le simulateur NeuSter (NEUron-ClusTER), offre un ensemble de composants logiciels élaborés pour simuler l'activité d'un réseau de neurones à impulsion au sein d'un ensemble de calculateurs d'architecture hétérogène (POSIX) dans un réseau local. L'ensemble des informations de construction et de configuration du réseau simulé transite par une machine particulière (le serveur) à destination des machines de calcul (les esclaves). Chaque machine esclave prend en charge le calcul d'un ensemble de cartes neuronales et communique directement avec les autres esclaves si les cartes efférentes sont distantes. Pour limiter la quantité d'information transmise sur le réseau, des données sont transmises uniquement lorsqu'un événement se produit (propagation event-driven). L'acquisition des données issues des capteurs ainsi que la réalisation des actions sont prises en charge par une machine esclave embarquée sur le robot et munie d'une connexion sans-fil.

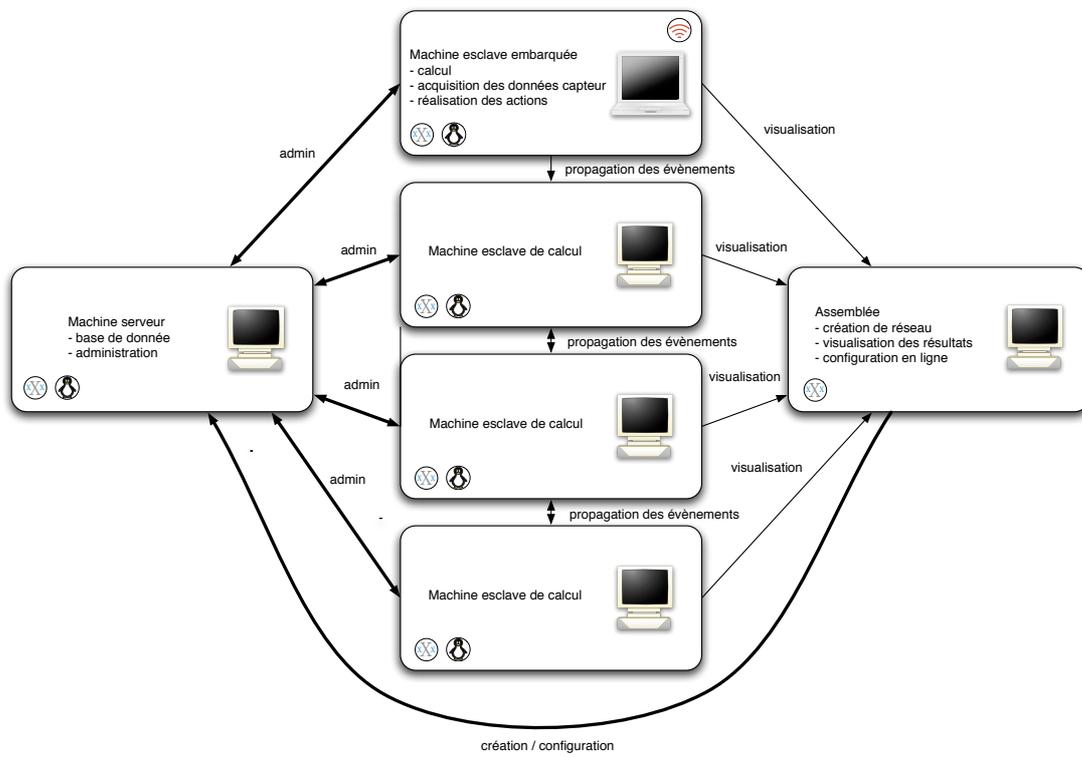


FIG. 2 – NeuSter : simulation de réseaux de neurones et calculs distribués.

RÉSUMÉ

Depuis plusieurs années, la robotique mobile tente de s'extraire de l'espace amniotique des laboratoires de recherche afin d'explorer l'univers imprévisible, voire hostile, de nos lieux de vie, de travail, pour nous servir ou nous divertir. Or, les méthodes classiques, symboliques et logiques de l'intelligence artificielle nécessitent des modèles du robot, de ses actions et de ses perceptions, conçus a priori. Elles demeurent de ce fait peu adaptées vis-à-vis de l'inattendu et de la nouveauté. D'autre part, les systèmes d'apprentissage artificiel, souvent d'inspiration biologique, semblent à présent en voie de fournir les capacités d'adaptation manquantes à ces premières approches tout en affrontant des environnements de plus en plus riches et complexes.

Nous envisageons dans cette thèse l'apprentissage comme un mécanisme central et constitutif de l'architecture robotique. Cette architecture peut être représentée sous les traits d'une boucle sensori-motrice où actions et perceptions se rejoignent au sein d'une structure associative. Dans cette architecture l'apprentissage permet, bien sûr, l'acquisition de connaissances nouvelles sur l'environnement. Se pose alors la question du type de représentation à mettre en œuvre afin d'identifier, de façon efficace et robuste, les données mémorisées dans le flux d'information provenant des capteurs. L'apprentissage intervient également dans la modélisation des actions du robot : sous une forme directe, en associant des combinaisons de consignes simples sur les moteurs, et sous une forme indirecte en mémorisant les effets de ces actions sur l'environnement ou sur le robot lui-même. Dans l'architecture étudiée, cette forme d'apprentissage (l'apprentissage objectif) a pour support un réseau de neurones à impulsions permettant un apprentissage en ligne non supervisé. L'architecture que nous étudions permet également d'exprimer les motivations et les objectifs du robot par le biais d'un second système d'apprentissage (l'apprentissage subjectif), en associant une valeur de récompense aux représentations des actions ou des perceptions, par un mécanisme d'apprentissage par renforcement. C'est donc l'utilité de chaque action, son espérance de gain, qui permettra finalement à un processus décisionnel d'avoir lieu.

Cette architecture conceptuelle, mais également son pendant algorithmique et logiciel, constitue le point de départ, la base de réflexion qui nous a été offerte afin de

permettre son intégration au sein d'une plateforme robotique. Nos contributions se situent d'une part dans l'amélioration du modèle de neurone formel qui est à la base de l'apprentissage, d'autre part au niveau de l'apprentissage de la représentation des objets et enfin dans une étude approfondie des mécanismes associatifs et décisionnels. Dans un premier temps, nous proposons de remplacer le modèle d'intégrateur à seuil adaptatif par un modèle utilisant une fonction d'activation stochastique. Ce modèle permet à la fois une plus grande robustesse de la reconnaissance et une simplification des communications entre neurones. Nous proposons d'autre part de représenter chaque objet par un ensemble de détecteurs sensibles aux images perçues selon différents points de vue. Cette représentation par vue (view-based) peut-être construite en exploitant une cohérence spatiale et temporelle qui assure que si cette perception est continue, alors les différentes vues de l'objet apparaîtront proches spatialement et temporellement et pourront donc être associées. La troisième partie de nos travaux est dédiée à l'étude des associations permettant de lier les représentations issues de la perception. Nous proposons de distinguer trois classes de représentations : d'une part les représentations mentionnées plus haut qui permettent l'identification des objets dans la scène (classe "Quoi"). D'autre part, des représentations issues de données proprioceptives permettant la localisation d'un objet dans la scène (classe "Où"). Finalement, les représentations des actions du robot sont obtenues par l'apprentissage de consignes aux moteurs (classe "Action"). Ces trois classes ainsi identifiées, nous proposons d'associer leurs éléments suivant trois schémas particuliers. Dans un premier temps, l'association sensori-motrice des éléments des classes Where et How aura pour but de produire une représentation des modifications de la position des objets, observées lors du déplacement du robot. Dans un second temps, le système devra associer les transitions entre les différentes vues d'un objet particulier ("Quoi") étant donnés sa position ("Où") et l'action ("Action") ayant généré le déplacement. Enfin, nous associerons à chaque action ("Action"), l'ensemble des contextes perceptifs ("Quoi" et/ou "Où") rendant cette action applicable, en d'autres termes les préconditions de l'action considérée. Ces associations ont pour objectif de construire un horizon de prédiction permettant de sélectionner l'action la mieux appropriée dans un contexte donné. Replacée dans un cadre connexionniste, la conception de cette structure adaptative a pour but une construction incrémentale du réseau mettant à profit les capacités adaptative des neurones tout en conservant l'uniformité des algorithmes de traitement locaux.

L'exposé de ces travaux sera précédé d'une présentation des concepts clés de l'apprentissage ouvert, tels qu'ils sont débattus dans la littérature. Nous joindrons à cette introduction, une présentation des principes de l'apprentissage par renforcement et une introduction aux réseaux de neurones formels qui permettront au lecteur de positionner nos travaux dans le cadre plus global de l'apprentissage artificiel.