



Université Paris Sud - Ecole doctorale d'Informatique

THÈSE

présentée pour obtenir le titre de DOCTEUR en Informatique

DE L'AUTO-ÉVALUATION AUX ÉMOTIONS - APPROCHE NEUROMIMÉTIQUE ET BAYESIENNE DE L'APPRENTISSAGE DE COMPORTEMENTS COMPLEXES IMPLIQUANT DES INFORMATIONS MULTIMODALES

par

Adrien Jauffret

LIMSI - CNRS / Université Paris Sud 11 / UPR3251
ETIS - ENSEA / Université de Cergy-Pontoise / CNRS UMR 8051

Soutenue le 11 juillet 2014 devant le jury composé de :

P. GAUSSIER,	ETIS, ENSEA - Université de Cergy-Pontoise - CNRS	Co-directeur de thèse
P. TARROUX,	LMSI-CNRS, Université Paris-Sud	Co-directeur de thèse
D. FILLIAT,	INRIA/ENSTA ParisTech FLOWERS	Rapporteur
N. ROUGIER,	Mnemosyne, INRIA Bordeaux	Rapporteur
P. BONNIFAIT,	Heudiasyc, Université Technologique de Compiègne	Examineur
B. POU CET,	LNC, CNRS - Université Aix-Marseille	Examineur
M. SEBAG,	LRI, CNRS - Université Paris-Sud	Examineur

Remerciements

Je tiens tout d'abord à remercier les membres de mon jury pour avoir accepté d'évaluer la valeur scientifique de mes travaux. J'aimerais exprimer toute mon estime à mes rapporteurs, David Filliat et Nicolas Rougier, pour avoir accepté de lire et commenter mon manuscrit ainsi que Philippe Bonnifait, Bruno Poucet et Michèle Sébag pour leur temps et pour les discussions intéressantes que nous avons eu lors de ma soutenance.

Je remercie également mon encadrant, Nicolas Cuperlier, ainsi que mes directeurs de thèse, Philippe Gaussier et Philippe Tarroux, pour m'avoir guidé tout au long de ces années. J'ai pris plaisir à partager avec eux ma passion pour les créatures artificielles, la cybernétique et les neurosciences. Leurs enseignements ont largement contribué à faire de moi le Jedi que je suis devenu ainsi qu'à transformer peu à peu cette passion en une vocation.

Je tiens à remercier Inbar Fijalkow et Mathias Quoy, respectivement ex-directrice et directeur du laboratoire ETIS, pour m'avoir accueilli pendant ces années de thèse.

Merci également à Astrid et Nelly pour leur humour et bonne humeur ainsi que pour avoir su brillamment faire face avec moi à l'incroyable paperasse administrative que requiert une thèse co-encadrée par deux universités très différentes.

Je remercie bien évidemment tous les membres du laboratoire pour l'indulgence dont ils ont fait preuve face à mes nombreuses pirouettes linguistiques ainsi que pour l'effervescence intellectuelle que nos débats éclectiques ont pu me procurer.

Merci à Pierre A., Cyril, Arnaud, Ghilès, Julien, Frederic, Christophe, Matthieu et Sofiane pour m'avoir permis de m'intégrer au sein du laboratoire et me frotter au monde de la recherche alors que je n'étais encore que stagiaire.

Merci également à la nouvelle génération : Abdelhak, Alexandre, David, Khursheed, Souheil, Ali, Raphael, Moeed et Artem ainsi qu'à mes jeunes padawans, Marwen et Caroline.

Merci à Pierre D., Antoine, Sylvain, Joanna, Quang, Kasia, Loïc, Abdel, Rahima et bien sur Ramesh pour l'esprit d'équipe dont ils font preuve et avec qui je construis jour après jour l'avenir de Partnering-Robotics.

Je tiens également à remercier mes collègues artificiels au même titre que mes collègues humains (bien que moins stimulants sur le plan conversationnel) : merci à Robulab, Roburoc, Berenson, Alvex et finalement Diya One pour leur patience et leur dévouement face aux nombreuses expériences que je leur ai fait subir.

Enfin, comment ne pas remercier mes parents, Chantal et Daniel, ainsi que ma soeur, Mathilde, pour le soutien qu'ils m'ont offert tout au long de ces 27 années à leur côté.

J'aimerais également remercier, et surtout m'excuser auprès de Fabien et Léo ainsi que toutes les personnes que je n'ai pas cité pour ces nombreuses soirées lors desquelles mon travail est passé au premier plan.

Pour terminer, j'aimerais souligner le fait que cette thèse n'aurait sûrement pas été ce qu'elle est sans la présence et le soutien inconditionnel de Charlotte, mon affordance émotionnelle et ma moitié.

Cette thèse a pour objectif la conception d'une architecture de contrôle bio-inspirée permettant à un robot de naviguer de manière autonome sur de grandes distances. Du point de vue des sciences cognitives, le modèle développé permet également d'améliorer la compréhension des mécanismes biologiques impliqués. De précédents travaux ont montré qu'un modèle de cellules de lieu, enregistrées chez le rat, permettait à un robot mobile d'apprendre des comportements de navigation robustes, tels qu'une ronde ou un retour au nid, à partir d'associations entre lieu et action. L'apprentissage et la reconnaissance d'un lieu ne reposaient alors que sur des informations visuelles. L'ambiguïté de certaines situations visuelles (e.g. un long couloir) ne permettait pas de naviguer dans de grands environnements.

L'ajout d'autres modalités constitue une solution efficace pour augmenter la robustesse de la localisation dans des environnements complexes. Cette solution nous a permis d'identifier les briques minimales nécessaires à la fusion d'informations multimodales, d'abord par le biais d'un conditionnement simple entre 2 modalités sensorielles, puis par la formalisation d'un modèle, plus générique, de prédictions inter-modales. Ce modèle générique permet à un ensemble de modalités d'apprendre à s'inter-prédire dans des situations usuelles. C'est un mécanisme bas niveau qui permet de générer une cohérence perceptive, en ce sens que l'ensemble des modalités sensorielles s'entraident pour ne renvoyer qu'une perception claire et cohérente aux mécanismes décisionnels de plus haut niveau. Les modalités les plus corrélées sont ainsi capables de combler les informations manquantes d'une modalité défaillante (cas pathologique). Ce modèle implique la mise en place d'un système de prédiction et donc une capacité à détecter de la nouveauté dans ses perceptions. Ainsi, le modèle est également capable de détecter une situation inattendue ou anormale et possède donc une capacité d'auto-évaluation : l'évaluation de ses propres perceptions. Partant de ce modèle de prédictions inter-modales, capable d'une première auto-évaluation, nous nous sommes ensuite mis à la recherche des propriétés fondamentales à tout système pour l'évaluation de ses comportements.

La première propriété essentielle a été de constater qu'évaluer un comportement sensorimoteur revient à reconnaître une dynamique entre sensation et action, plutôt que la simple reconnaissance d'un pattern sensoriel. La première brique du modèle encapsule donc ce couplage entre sensation et action, formant ainsi un modèle interne minimaliste des interactions du robot avec son environnement. Ce modèle interne de la dynamique du comportement est la base sur laquelle le système fera des prédictions.

La seconde propriété essentielle est la capacité à extraire l'information importante de ce modèle interne par le biais de calculs statistiques. Il est nécessaire que le robot apprenne à capturer les invariants statistiques en supprimant l'information incohérente. Nous avons donc montré qu'il était possible d'estimer une densité de probabilité par le biais d'une réduction des erreurs quadratiques moyennes, c'est à dire un simple conditionnement. Cet apprentissage statistique permet de réaliser l'équivalent d'une inférence bayésienne, puisque le système estime la probabilité de reconnaître un comportement à partir de la reconnaissance d'informations statistiques apprises sur ce comportement. C'est donc par la mise en cascade de simples conditionnements que le système peut apprendre à estimer les moments statistiques d'une dynamique comportementale (moyenne, variance, asymétrie, etc...). La non-reconnaissance de cette dynamique lui permet de détecter qu'une situation est anormale.

Mais détecter un comportement inhabituel ne nous renseigne pas pour autant sur son inefficacité. Le système doit également surveiller l'évolution de cette anomalie dans le temps pour pouvoir juger de la pertinence du comportement. Nous montrerons comment un contrôleur émotionnel

Synopsis

peut faire usage de cette détection de nouveauté pour réguler le comportement et ainsi permettre au robot d'utiliser la stratégie la plus adaptée à la situation rencontrée. Pour finir, nous avons mis en place une procédure de frustration permettant au robot de lancer un appel à l'aide lorsqu'il détecte qu'il se retrouve dans une impasse. Ce réseau de neurones permet au robot d'identifier les situations qu'il ne maîtrise pas dans le but d'affiner son apprentissage, à l'instar de certains processus développementaux.

Mots clés : robotique bio-inspirée, réseaux de neurones artificiels, navigation sensori-motrice, auto-évaluation, meta-apprentissage, sélection de l'action.

The goal of this thesis is to build a bio-inspired architecture allowing a robot to autonomously navigate over large distances. In a cognitive science point of view, the model also aim at improving the understanding of the underlying biological mechanisms. Previous works showed that a computational model of hippocampal place cells, based on neurobiological studies made on rodent, allows a robot to learn robust navigation behaviors. The robot can learn a round or a homing behavior from a few associations between places and actions. The learning and recognition of a place were only defined by visual information and shows limitations for navigating large environments.

Adding other sensorial modalities is an effective solution for improving the robustness of places recognition in complex environments. This solution led us to the elementary blocks required when trying to perform multimodal information merging. Such merging has been done, first, by a simple conditioning between 2 modalities and next improved by a more generic model of inter-modal prediction. In this model, each modality learns to predict the others in usual situations, in order to be able to detect abnormal situations and to compensate missing information of the others. Such a low level mechanism allows to keep a coherent perception even if one modality is wrong. Moreover, the model can detect unexpected situations and thus exhibit some self-assessment capabilities: the assessment of its own perception. Following this model of self-assessment, we focus on the fundamental properties of a system for evaluating its behaviors.

The first fundamental property that pops out is the statement that evaluating a behavior is an ability to recognize a dynamics between sensations and actions, rather than recognizing a simple sensorial pattern. A first step was thus to take into account the sensation/action coupling and build an internal minimalist model of the interaction between the agent and its environment. Such of model defines the basis on which the system will build predictions and expectations.

The second fundamental property of self-assessment is the ability to extract relevant information by the use of statistical processes to perform predictions. We show how a neural network can estimate probability density functions through a simple conditioning rule. This probabilistic learning allows to achieve bayesian inferences since the system estimates the probability of observing a particular behavior from statistical information it recognizes about this behavior. The robot estimates the different statistical momentums (mean, variance, skewness, etc...) of a behavior dynamics by cascading few simple conditioning. Then, the non-recognition of such a dynamics is interpreted as an abnormal behavior.

But detecting an abnormal behavior is not sufficient to conclude to its inefficiency. The system must also monitor the temporal evolution of such an abnormality to judge the relevance of the behavior. We show how an emotional meta-controller can use this novelty detection to regulate behaviors and so select the best appropriate strategy in a given context. Finally, we show how a simple frustration mechanism allows the robot to call for help when it detects potential deadlocks. Such a mechanism highlights situations where a skills improvement is possible, so as some developmental processes.

Keywords : bio-inspired robotics, artificial neural-networks, sensory-motor navigation, self-assessment, metalearning, action selection.

Table des matières

Introduction	13
1 Outils du cybernéticien	25
1.1 Une brève introduction aux réseaux de neurones artificiels	26
1.2 Conditionnement	26
1.3 Compétition	29
1.3.1 Winner Takes All	30
1.4 Classification	31
1.4.1 Champs de neurones dynamiques	33
1.5 Discussion	38
2 Modèles de navigation en robotique	41
2.1 Navigation nautique	42
2.1.1 Navigation côtière	42
2.1.2 Navigation à l'estime	43
2.1.3 Navigation astronomique	44
2.2 Navigation en robotique classique	44
2.3 Navigation animale	46
2.3.1 Intégration de chemin	48
2.3.2 Région hippocampique et cellules de lieu	49
2.4 Navigation robotique bio-inspirée	53
2.4.1 Modèles inspirés des insectes	53
2.4.2 Modèles inspirés des mammifères	55
2.5 Per-Ac : un modèle de couplage sensori-moteur pour la perception	56
2.5.1 Modèle de cellules de lieu visuelles	56
2.5.2 Modèle d'intégration de chemin	61
3 Un parallèle entre formalisme neuronal et bayésien	63
3.1 Limitations du modèle actuel	64
3.2 Approche bayésienne et notion d'inférence	65
3.3 Notre modèle de cellules de lieu d'un point de vue bayésien	68
3.3.1 Description générale du problème	69
3.3.2 Factorisation immédiate	69
3.3.3 Factorisation retardée	72

3.4	Pondération statistique des amers	73
3.4.1	Un modèle neuronal de l'estimation de densité de probabilité	73
3.4.2	Expérience : estimons la dispersion de l'azimut de chaque amer : $P(A/S)$	75
3.4.3	Expérience : estimons la dispersion globale de l'azimut : $P(A)$	76
3.5	Une approche statistique de la reconnaissance de lieu	79
3.5.1	Modèle de cellules de lieux probabilistes	79
3.5.2	Expérience : estimation de la probabilité d'être dans un lieu sachant les observations	82
3.6	Discussion	84
4	Localisation à partir d'informations idiothétiques	87
4.1	Les cellules de grilles	88
4.2	Un modèle plausible de cellules de grille	91
4.2.1	Description du modèle	91
4.2.2	Structure des expériences	94
4.3	Résultats	94
4.3.1	Expérience 1 - exploration aléatoire, sans recalibration	94
4.3.2	Expérience 2 - recalibration de l'intégration de chemin à l'aide d'un retour au nid périodique.	96
4.3.3	Expérience 3 - De la cellule de grille à la cellule de lieu	99
4.4	Discussion	101
5	Fusion d'informations multimodales	105
5.1	Cohérence et conflits sensoriels	106
5.1.1	Conflits bénins entre modalités externes	106
5.1.2	Conflits majeurs et ruptures perceptives	108
5.2	Un couplage bi-modale par simple conditionnement	109
5.3	Robustesse de la localisation : une expérience sur robot réel	110
5.3.1	A propos de la topologie des cellules de grille	112
5.3.2	Robustesse de la reconnaissance des lieux aux ambiguïtés visuelles	114
5.3.3	A propos du problème du kidnapping	114
5.4	Discussion	115
5.4.1	Vers un modèle générique de prédiction intermodale	117
6	Un modèle générique de détection de nouveauté pour l'auto-évaluation	119
6.1	Modèles computationnels de détection de nouveauté	121
6.2	Modèle des perceptions	123
6.3	Modèle de détection de nouveauté	125
6.4	Expériences et résultats	126
6.5	Discussion	130
7	De l'évaluation à la régulation	135
7.1	Un modèle de contrôleur émotionnel au second-ordre	137
7.2	Sélection et coopération de stratégie	138
7.3	Expérience : La frustration et son rôle communicatif dans le développement de l'autonomie	139

7.3.1	Résultats en simulations	139
7.3.2	Expérience sur robot réel	140
7.4	Expérience : Le stress et son rôle actif dans la régulation des comportements . .	143
7.4.1	Résultats en simulations	143
7.5	Discussion	146
	Conclusion	149
A	Strategie de suivi de route	181
A.1	Un algorithme minimaliste pour réaliser un comportement de suivi de route . .	181
B	Outils de simulation de réseaux de neurones	185
B.1	Coeos : une interface graphique de conception d'architectures neuronales . . .	185
B.2	Promethe : un simulateur temps-réel et distribué de réseaux de neurones	186
C	Plates-formes robotiques	189
C.1	Plate-forme d'intérieur	189
C.2	Plate-forme d'extérieur	190
C.3	Communication réseau	190

People say to me, "Are you looking for the ultimate laws of physics?" No, I'm not. I'm just looking to find out more about the world. If it turns out there is a simple ultimate law that explains everything, so be it, that would be very nice to discover. If it turns out it's like an onion with millions of layers... then that's the way it is. But either way, Nature is there and she's going to come out the way She is. So therefore when we go to investigate we shouldn't predecide what it is we're trying to do, except to find more about it.

– R. Feynman

Introduction

Les premiers jalons historiques de l'intelligence artificielle (IA) remontent à l'antiquité et commencent avec "le vieux souhait de jouer à Dieu" [McCorduck, 2004]. Ainsi, les concepts de machines pensantes et de créatures artificielles apparaissent déjà dans la mythologie grecque, dont les légendes de Talos de Crete, le robot de bronze d'Héphaïstos ou encore Galatée de Pygmalion sont de bons exemples. Entre le 19ème et le 20ème siècle, les êtres artificiels sont progressivement devenus des lieux communs de fiction comme en témoigne l'oeuvre "Frankenstein" de [Shelley, 1823] ou "R.U.R." (Rossum's Universal Robots) de [Apek and Wylie, 2010] puis les romans précurseurs d'Asimov [Asimov, 2004] ou d'Arthur C. Clarke [Clarke and Kubrick, 1999].

L'étude du raisonnement formel a lui aussi été développé par de nombreux philosophes et mathématiciens depuis l'antiquité (parmi les plus notables : Aristote, Euclide, Al-Khwarizmi puis Leibniz). En se basant sur ces études, les fondateurs de l'IA ont proposé des algorithmes capables d'imiter, étape par étape, le raisonnement permettant à l'être humain de résoudre des puzzles et réaliser des déductions logiques [Hart et al., 1968; Poole et al., 1998; Nilsson, 1998]. Le champ de recherche de l'IA a été fondé en 1956 lors d'une conférence incluant notamment John McCarthy, Marvin Minsky, Allen Newell et Herbert Simon qui devinrent leaders de ce domaine plusieurs dizaines d'années durant [Crevier, 1993]. Ces pionniers étaient profondément optimistes à propos de l'avenir de ce nouveau domaine. Simon avaient même prédit : *"les machines seront capables, d'ici une vingtaine d'années, d'accomplir toutes tâches qu'un humain peut réaliser"* [Simon, 1965] et Minsky écrivit : *"dans une génération ... le problème de l'intelligence artificielle sera globalement résolu"* [Minsky, 1967].

Cette vision de l'intelligence a permis quelques avancées majeures dans la résolution automatique de problèmes d'algèbre, la preuve de théorèmes logiques, la réalisation de systèmes experts ou encore la création d'agents conversationnels. La victoire de Deep Blue contre le champion d'échec Garry Kasparov en 1997 a marqué les esprits [Newborn and Newborn, 1997]. Mais la capacité de cette approche a vite trouvé une limite dès lors qu'elle se confrontait au monde réel continu, dynamique et incertain. La puissance de calcul limitée des machines et l'explosion combinatoire engendrée par la résolution de tâches même triviales constituent les premières limitations de l'approche [Moravec, 1988]. La compréhension par la machine du sens commun associé aux concepts manipulés en est une autre. Ces programmes ne savent pas plus qu'ils effectuent des démonstrations géométriques ou algébriques que Deep Blue ne savait qu'il jouait aux échecs. Ces cas représentent plus des opérations intellectuelles assistées par ordinateur que de l'intelligence à proprement parler.

La raison de cet échec est que la majeure partie des travaux effectués dans le domaine de l'IA

reposent sur une représentation symbolique du monde. Les symboles représentent des étiquettes que l'on associe à des concepts, tel qu'une chaise ou un ballon. L'agent se voit attribuer une base prédéfinies de symboles avec lesquels il peut jongler. Selon ce point de vue, l'intelligence d'un agent se résume donc à la manipulation de symboles. Mais l'approche fait face à un problème fondamental (le "Symbol Grounding Problem") [Harnad, 1990] puisque manipuler des symboles ne permet pas de comprendre pour autant leur sens dans le monde réel.

[Searle, 1999] eu l'idée d'une expérience de pensée démontrant l'échec de l'approche symbolique : "Supposons que l'on soit dans une pièce fermée avec la possibilité de recevoir et d'envoyer des symboles (via un clavier et un écran, par exemple). On dispose de caractères chinois et de règles de travail (instructions) permettant de produire certaines suites de caractères en fonction des caractères introduits dans la pièce, sur le modèle des organismes de vente par correspondance qui traitent leur courrier client en plaçant des réponses préparées à l'avance, et déjà imprimées, dans des enveloppes. Si l'on fournit une histoire suivie d'une question, toutes deux écrites en chinois, l'application des règles ne peut que conduire à donner la bonne réponse, mais sans que l'opérateur ait compris quoi que ce soit, puisqu'il ne connaît pas le chinois. Il aura juste manipulé des symboles qui n'ont pour lui aucune signification."

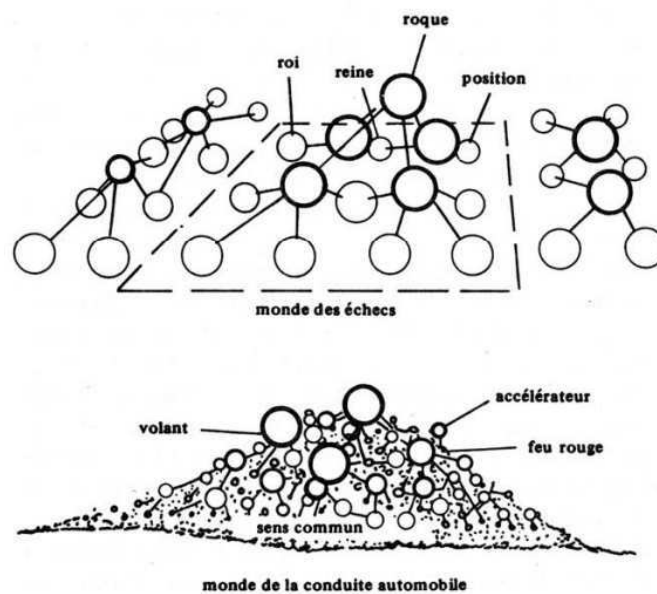


FIGURE 1 – Schéma extrait de [Bierre, 1985] repris par [Varela, 1996], illustrant les différentes complexités de représentation du monde en fonction des limites et des règles de ce monde : "Pour l'espace « échecs », il est possible de dessiner un réseau de relations dont les noeuds représentent chaque élément pertinent. Dans le cas de l'espace « conduite automobile », une semblable tentative montre bien qu'au-delà de quelques éléments isolés, le réseau évolue rapidement vers une masse non-circonscrite de sens commun."

Comment définir, en des termes interprétables par un programme, le concept communément associé au mot "chaise" ? De toute évidence, définir ce mot en associant le symbole "Chaise" à un ensemble d'images représentant toutes les chaises existantes n'est pas une bonne solution. En plus d'un apprentissage laborieux, l'agent n'est alors pas capable de reconnaître une chaise qui ne lui a pas été présentée auparavant puisqu'il n'a pas saisi le sens de ce mot.

Robotique comportementale

À la fin des années 1980, plusieurs chercheurs plaident pour une approche de l'intelligence artificielle complètement inédite, centrée sur la robotique. La théorie de l'énaction, proposée par [Mataruna and Varela, 1980] met ainsi en avant l'importance du corps physique pour acquérir une compréhension sensée du monde. La machine doit percevoir, bouger, survivre et évoluer dans un environnement réel pour parvenir à le comprendre. L'intelligence n'est plus vue comme un processus de réflexion détaché de l'environnement mais bien un processus de cognition "incarné" au sein d'un corps mobile à la morphologie particulière. On parle alors d'*embodiment* [J. Haugeland, 1995; Scheier and Pfeifer, 1997]. Ces capacités sensori-motrices sont essentielles aux capacités de plus haut niveau et le raisonnement abstrait est en fait la capacité humaine la moins importante (cf. le paradoxe de Moravec [Russell and Cohn, 2012]).

Selon ce point de vue, reconnaître une chaise c'est avant tout reconnaître un objet sur lequel on peut s'asseoir. Le sens naît de l'interaction de l'agent avec son monde. Reconnaître un objet c'est reconnaître ce que l'on peut faire avec. On notera que le sens que l'on donne au mot "chaise" n'est valable que pour l'être humain et sa morphologie spécifique. Ce même objet n'a certainement pas le même sens pour l'éléphant ou pour la mouche par exemple.

Dans le même temps, quelques chercheurs tentaient de reproduire les comportements élémentaires de certains animaux [Walter, 1953; Braitenberg, 1984]. Suivant la mouvance bio-inspirée, les roboticiens s'intéressent alors aux modèles que pouvaient représenter les insectes dont la morphologie est particulièrement adaptée à une réalisation robotique. L'un des premiers à s'intéresser aux insectes en tant que modèle fut Rodney Brooks qui suggéra que les robots devaient d'abord pouvoir se comporter comme le plus primitif des animaux, avant de prétendre à des capacités cognitives plus élaborées [Brooks, 1986; Arkin, 1998]. Ces études donnent naissance à l'approche *animat* introduite par [Wilson, 1991] et dont le but est la réalisation d'animaux artificiels en étudiant la cognition animale au niveau le plus simple pour grimper progressivement en complexité et atteindre un jour l'intelligence humaine par le bas. L'approche ravive des concepts nés de la cybernétique et de la régulation qui avaient perdu de leur impact depuis les années soixante. La **cybernétique**, popularisée en 1948 par [Wiener, 1948], est une science des systèmes autorégulés, qui étudie le concept de "feedback", ou boucle de retour. La discipline s'intéresse moins aux composantes d'un système qu'à leurs interactions. Le comportement global du système est pris en compte avant tout le reste.

Cette thèse s'inscrit dans cette démarche et considère le comportement émergent du système comme un *miroir de son intellect*. Aussi, nous nous éloignerons d'une représentation symbolique pour adopter une approche dite "sensori-motrice". C'est le couplage entre sensation et action qui permettra au robot de se construire une représentation de son environnement.

Nous implémenterons nos solutions sous la forme de réseaux de neurones artificiels. Ces derniers représentent un outil cadrant parfaitement avec l'approche de la cybernétique. L'interaction d'un grand nombre de neurones modélisés par des équations simples peut faire émerger des capacités de catégorisation et des comportements complexes. L'une des propriétés essentielles de tels réseaux est la possibilité d'apprendre i.e. de modifier le comportement du système au fur et à mesure de ses expériences. Ces propriétés permettent d'obtenir de bonnes capacités de généralisation et facilite également l'adaptation du comportement du robot lorsque l'environnement change.

Les solutions industrielles classiques s'appuient généralement sur la programmation spéci-

fique du robot pour réaliser les tâches désirées. Bien souvent, les robots ne sont programmés que pour réaliser une seule tâche de manière optimale. Les ingénieurs sont forcés de programmer entièrement le comportement du robot pour l'adapter à chaque nouvelle tâche. Cette opération nécessite des humains experts et se révèle très coûteuse en temps de développement.

Selon l'approche que nous adopterons ici, la programmation spécifique de tâches est remplacée par l'apprentissage de celles-ci. Cependant, nous irons plus loin encore en concevant le robot non pas comme une machine introvertie mais comme un système capable d'interagir naturellement avec un utilisateur pour affiner son apprentissage avec le temps. L'utilisateur n'a alors plus besoin d'être expert ni de programmer le système pour chaque nouvelle tâche à résoudre.

Pour aboutir à ce type de solution, une démarche ascendante (bottom-up) et constructive est nécessaire [Mataruna and Varela, 1980; Varela et al., 1993]. Suivant cette démarche, il convient de développer progressivement la complexité du robot pour avoir la possibilité d'étudier ses capacités lorsque ce dernier est plongé dans son environnement. Cette démarche tente de faire émerger des théories de la pratique, et non l'inverse. Les modèles proposés s'inspirent de principes minimalistes que l'on trouve en biologie animale. A l'inverse d'approches descendantes (top-down) contraignant l'espace des possibles pour aboutir à la solution désirée, nous laisserons apparaître des propriétés émergentes, que l'on ne saurait imaginer si le robot et l'environnement étaient considérés indépendamment.

Problématique

Cette thèse s'inscrit dans le cadre de la navigation robotique bio-inspirée en environnement réel et implique la capacité pour un système mobile à se déplacer de manière autonome dans un monde a priori dynamique et inconnu.

Pour se localiser et naviguer, un robot autonome peut utiliser de multiples solutions. Mais que ce soit pour la modélisation des processus visuels ou pour la localisation, la cartographie, la navigation et la planification, les mesures nécessaires sont fortement bruitées, ce qui justifie le recours à des méthodes probabilistes. Les approches bayésiennes ont la faveur de la communauté robotique pour conserver le contrôle sur les incertitudes pesant sur les données. Les approches neuromimétiques fournissent des solutions robustes et originales pour résoudre le même problème. Des travaux récents suggèrent que des processus statistiques sont à l'oeuvre dans le cerveau et montrent comment certains codages par population de neurones implémentent les principes de bases de l'approche bayésienne [Deneve S, 2004]. L'équipe Neurocybernétique d'ETIS et l'équipe "Cognition, Perception et Usages" du LIMSI ont travaillé à la mise au point de modèles bio-inspirés neuronaux et de modèles bayésiens pour la localisation et la navigation en robotique autonome.

Les travaux de [P. Gaussier, 1997] ont montré il y a maintenant plus d'une dizaine d'années qu'un système bio-inspiré pouvait par exemple apprendre à retourner vers un lieu en apprenant un nombre très réduit de couples lieu-action dans le voisinage de ce lieu-but. L'emploi de mécanismes de compétition entre ces couples lieu-action permet la création d'un bassin d'attraction qui attire inmanquablement le robot sur le lieu cible pour peu que le point de départ appartienne au même voisinage visuel. Le comportement du robot démontre ainsi sa reconnaissance du lieu cible alors même que ce lieu n'a pas été codé en temps que tel dans le réseau.

D'autres travaux ont permis de montrer que ce type d'approche pouvait se généraliser à l'apprentissage de rondes [Giovannangeli and Gaussier, 2010] ou à la reconnaissance active d'objets

[S. Leprêtre, 1999]. Cependant, même si dans ces tâches la capacité de généralisation tenait la première place, on peut se poser la question de la capacité de ces approches à traiter des problèmes de très grande taille. Que se passe-t-il lorsque le nombre d'amers visuels utilisés ne se compte plus en centaines mais en centaines de milliers voire millions ? Il est clair que la capacité d'un simple mécanisme de compétition à proposer une interprétation correcte s'effondre avec l'augmentation du nombre de vues apprises.

De récents travaux [Cuperlier et al., 2008] ont montré qu'une information de contexte pouvait être apprise en ligne et utilisée à la fois pour limiter la compétition (augmenter le rapport signal à bruit) mais aussi pour limiter la puissance de calcul nécessaire en ne calculant que l'activité des neurones associés au bon contexte. Mais même avec ce genre de mécanisme, le robot peut se trouver trop loin des lieux appris pour être attiré dans le bassin d'attraction désiré. Il peut alors se mettre à errer indéfiniment sans rejoindre le lieu cible.

Cependant, l'observation interne de l'évolution du niveau de reconnaissance des lieux traduit clairement cet échec. Un "*observateur*" interne pourrait donc qualifier le comportement du système.

L'enjeu dans la construction d'un tel système complexe doté d'une autonomie comportementale est double. D'une part il est nécessaire que les futurs systèmes de traitement de l'information soient à la fois fiables et flexibles, capable d'extraire de manière active les informations de l'environnement qu'ils jugent pertinentes pour adapter leur comportement. D'autre part, ces systèmes devront être capables d'auto-évaluer leurs performances afin de détecter les situations d'échec ou de dead-lock (boucle sans fin). Ces situations peuvent se produire au cours d'interactions dynamiques complexes, alors même que les comportements élémentaires du système sont corrects.

Pour un robot, l'autonomie peut alors se traduire par l'expression de comportements adaptatifs et non supervisés, lui permettant de détecter mais aussi de corriger des situations d'échec afin de mener à bien les tâches nécessaires à sa mission dans l'environnement dans lequel il se trouve. Rendre réellement utilisable les systèmes robotiques sophistiqués que l'on voit apparaître ces dernières années (cf la création de Cap robotique pour fédérer les nombreuses initiatives industrielles concernant les robots de services, les robots ludiques, les robots interactifs voire humanoïdes comme Asimo de Honda, HRP2, ICub ou Nao d'Aldebaran robotics à Paris) passe par la maîtrise de systèmes informatiques complexes relevant le défi de l'autonomie et de l'adaptation comportementale nécessitant la capacité à détecter des situations critiques et la capacité à les résoudre. Ces compétences sont cruciales pour le développement de nouveaux produits robotiques destinés à évoluer de manière autonome dans des environnements dynamiques a priori inconnus.

La robotique autonome se développe en particulier en direction de deux types d'applications : le véhicule autonome et la robotique d'assistance. Une amélioration du transport urbain pourrait passer par la mise au point de véhicules capables de se diriger de façon automatique d'un point à un autre en gérant les situations complexes et inattendues qu'il rencontrera lors de son parcours. Le fait de naviguer sur de longues distances est susceptible d'induire une instabilité de localisation si des informations de type GPS ne sont pas accessibles sur l'ensemble du trajet. Le système devra identifier ces situations d'erreur et y remédier. De la même façon, la robotique d'assistance supposera une capacité de navigation autonome, d'apprentissage des situations nouvelles et de reconnaissance de lieux et d'objets mais également la capacité à reconnaître les situations d'échec et à y remédier. Enfin, la maîtrise du passage à l'échelle des mécanismes per-

mettant de reconnaître des milliers voire des millions de formes en temps réel et d'auto-adapter le comportement du robot en cas d'échec est un défi incontournable pour l'ingénierie des systèmes intelligents de demain.

Ces questions sont aussi au centre des sciences cognitives avec notamment un nombre de plus en plus important de recherches sur le rôle des émotions dans nos processus cognitifs. Ces processus peuvent être assimilés à des contrôleurs au second ordre, capables à la fois de remettre en cause la priorité de nos stratégies, ou de réaliser la fusion d'informations multimodales pour fabriquer une perception stable et robuste du monde.

L'autonomie globale d'un système passe par 4 axes principaux que nous développerons progressivement tout au long de ces travaux, à savoir :

- La sélection d'informations pertinentes par l'extraction de régularités statistiques.
- L'intégration d'information multimodale pour obtenir une perception robuste et stable.
- L'auto-évaluation des comportements sensori-moteurs exhibés par le robot, par l'ajout d'un méta-prédicteur générique.
- La sélection et la régulation de ces comportements par le biais de processus émotionnels.

Extraction de régularités statistiques

Afin de se construire une perception robuste de son environnement, un système nécessite la capacité à extraire de celui-ci l'information importante pour la tâche à réaliser. Les approches statistiques fournissent des solutions efficaces pour filtrer l'espace des données que le système perçoit en mettant en avant les informations les plus stables. Il est nécessaire que le robot apprenne à capturer ces invariants statistiques en supprimant l'information instable ou trop bruitée. Dans l'architecture actuelle, tous les amers visuels sont traités de manière identique. Pourtant ces amers peuvent avoir des caractéristiques différentes (proximaux ou distaux) et être plus ou moins pertinents selon la tâche envisagée. Il est donc nécessaire d'introduire une capacité à prédire le déplacement de ceux-ci afin de les caractériser. Ceux dont la position angulaire reste invariante (sur une certaine distance) et dont la signature est unique peuvent servir à reconnaître des lieux de manière robuste. Au contraire, un même amer perçu sous différentes positions angulaires reflète clairement la présence d'un objet ambigu, redondant ou encore mobile, et ne devrait pas être pris en compte par le système.

Un mécanisme de méta-contrôle sera donc nécessaire ici pour apprendre à caractériser les amers visuels et moduler leur influence dans la reconnaissance des lieux. La constitution d'un tel mécanisme sera abordée selon les deux paradigmes : neuronal et bayésien. L'exploitation de ces prédictions devrait permettre d'une part d'améliorer les performances du système (adaptation) et d'autre part de définir un détecteur de nouveauté (curiosité, surprise...) réagissant à des données sortant de l'épure déterminée par un prédicteur neuronal (après apprentissage de la distribution des données).

Intégration d'information multimodale

Dans une deuxième phase, nous chercherons à utiliser ces mêmes techniques pour maîtriser des tâches plus complexes telles que l'intégration d'informations proprioceptives dans le codage des lieux. La nature même de cette intégration étant méconnue, différentes stratégies de fusion entre ces informations seront explorées.

D'un point de vue modèle neurobiologique, la fusion d'informations visuelles avec des informa-

tions idiothétiques devrait conduire à une amélioration notable de la stabilité des champs de lieu multimodaux en comparaison de champs issus d'une seule modalité. Cette fusion devra permettre la création de cellules de lieux plus robustes à des changements de l'environnement (variation de luminosité, perte d'amers lointains, fonctionnement épisodique de la modalité visuelle ...). La recalibration de l'odométrie par des informations visuelles devrait également permettre d'améliorer la fiabilité de ces informations lors de navigations dans de grands environnements (limitation de la dérive idiothétique, robustesse aux fermetures de boucles et au kidnapping...). L'impact des mécanismes de fusion/compétition entre modalités sur la stabilité et l'exactitude de l'information de position (lieux) sera étudiée dans le cadre de tâches de navigation sensorimotrices (ronde, retour à un lieu appris).

Auto-évaluation

Les travaux présentés dans cette thèse visent à doter nos systèmes robotiques de propriétés leur permettant de sélectionner les informations les plus pertinentes et d'adapter jusqu'à leurs comportements et leurs buts en fonction de l'évaluation de la situation rencontrée. Nous ferons l'hypothèse que les mécanismes de fusion impliquent la nécessité de prédire l'apport d'une modalité pour la reconnaissance d'un état de la même manière que juger le succès ou l'échec d'un comportement nécessite la prédiction de l'évolution d'un certain nombre de variables liées à la reconnaissance des états cibles. Il est donc légitime de supposer que ces différents problèmes pourraient reposer sur un système de prédiction générique capable de capturer rapidement les informations statistiques caractérisant une situation donnée pour faire des prédictions sur l'évolution de cet état. L'incapacité à prédire l'état courant pourrait alors être vue comme un signal de renforcement négatif impliquant la remise en cause du comportement en cours et l'apprentissage de nouveaux états [P. Andry, 2002].

Processus émotionnels

Dans ce cadre, la modélisation des émotions semble indispensable à la fois au méta-contrôle de nos mécanismes cognitifs [Damasio, 1995; LeDoux, 1992; Panksepp, 1998] mais aussi pour communiquer de manière pré-verbale nos états internes. Les travaux en psychologie et neurobiologie montrent l'importance d'introduire des mécanismes émotionnels au coeur même des systèmes d'Intelligence Artificielle [Damasio, 1995; Picard, 1997; Canamero and Gaussier, 2004]. On retrouve ainsi toute une gamme de modèles permettant aux robots de détecter de la nouveauté, de prédire le rythme d'une interaction [Andry et al., 2001], de devenir curieux [Oudeyer et al., 2007] afin qu'ils puissent développer de manière autonome de nouvelles compétences dans un cadre "open ended" (à horizon temporel ouvert). L'approche bio-inspirée basée sur des modèles émotionnels pour le méta-contrôle est encore peu explorée [Balkenius et al., 2009; Canamero and Gaussier, 2004; Mannella et al., 2008]. Nous nous intéresserons ainsi à comprendre comment les émotions peuvent moduler les motivations ou les «drives» contrôlant un robot.

Chez les mammifères, la régulation des comportements est liée aux interactions entre différentes structures dites limbiques. Nos émotions régulent nos comportements (ex. en cas de danger, arrêt des tâches en cours et déclenchement d'un comportement de fuite ou d'agression) mais elles sont elles même régulées par d'autres processus cognitifs (la surprise liée à la détection d'une nouveauté par exemple). Différents courants théoriques s'affrontent et opposent des modèles d'émotions élémentaires [Ekman, 1984] à des modèles dimensionnels (passage continu

d'une émotion à une autre) [Spencer, 1870; Wundt and Judd, 1897] et discutent de la primauté des mécanismes cognitifs ou viscéraux dans la genèse des émotions.

Dans ces travaux, nous ne tenterons pas de prendre parti au départ sur ces questions mais chercherons à utiliser dans un cadre restreint les données et modèles computationnels disponibles sur la détection de nouveauté et la gestion des comportements (modèles des ganglions de la bases, de l'hippocampe, de l'amygdale et du cortex préfrontal) pour en déduire les règles de régulation les plus simples possibles et tester leurs effets comportementaux dans des tâches de navigation autonome.

Portrait-Robot

Les travaux réalisés dans cette thèse ont été menés sur 2 robots mobiles : la plateforme mobile d'intérieur *Robulab*, et celle adaptée à la navigation en environnement extérieur *Roburoc* de Robosoft.

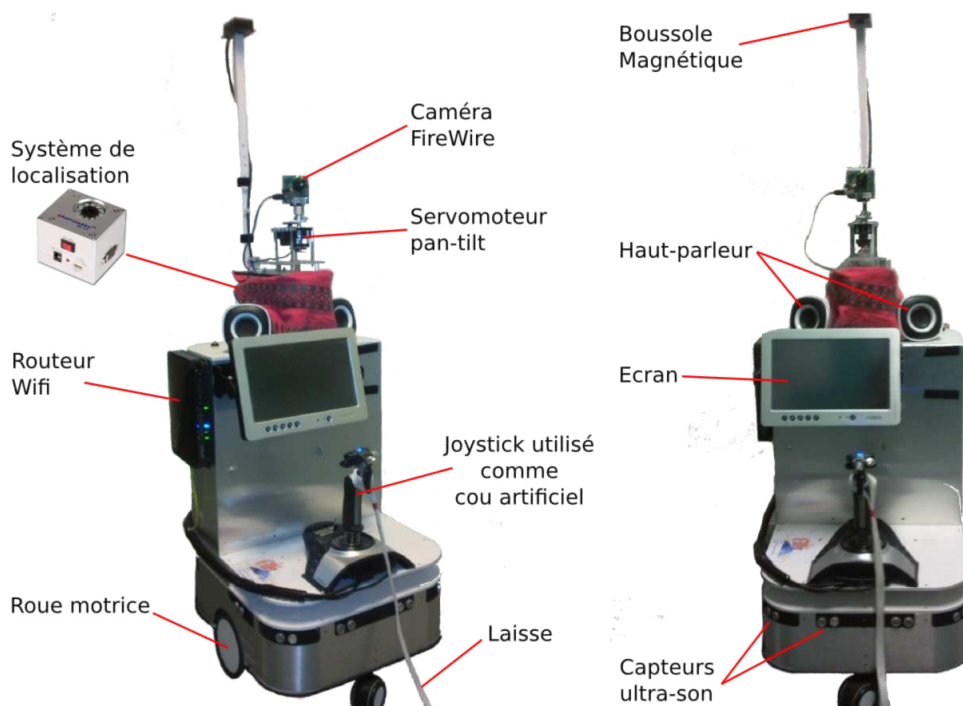


FIGURE 2 – Plateforme robotique mobile Robulab de Robosoft. Le robot est composé d'une base mobile, d'un corps contenant le pc embarqué et d'une tête supportant la caméra sur servomoteur. La boussole est positionnée au sommet d'un mat afin d'éviter d'éventuelles interférences par l'électronique embarquée. Le système de localisation n'est jamais utilisé par le robot. Il permet seulement d'enregistrer la trajectoire de celui-ci.

Les plateformes sont composées de 3 parties distinctes, à savoir une base mobile (Robulab PC/104) qui permet au robot de se mouvoir, un corps comprenant un coffre au sein duquel se trouve un ordinateur embarqué (Carte mini ITX i7), puis une "tête" composée d'une caméra montée sur servomoteur. Les robots sont également munies de microcontrôleurs (Arduino) pour contrôler le matériel embarqué ainsi que d'un routeur WiFi pour communiquer avec une ma-

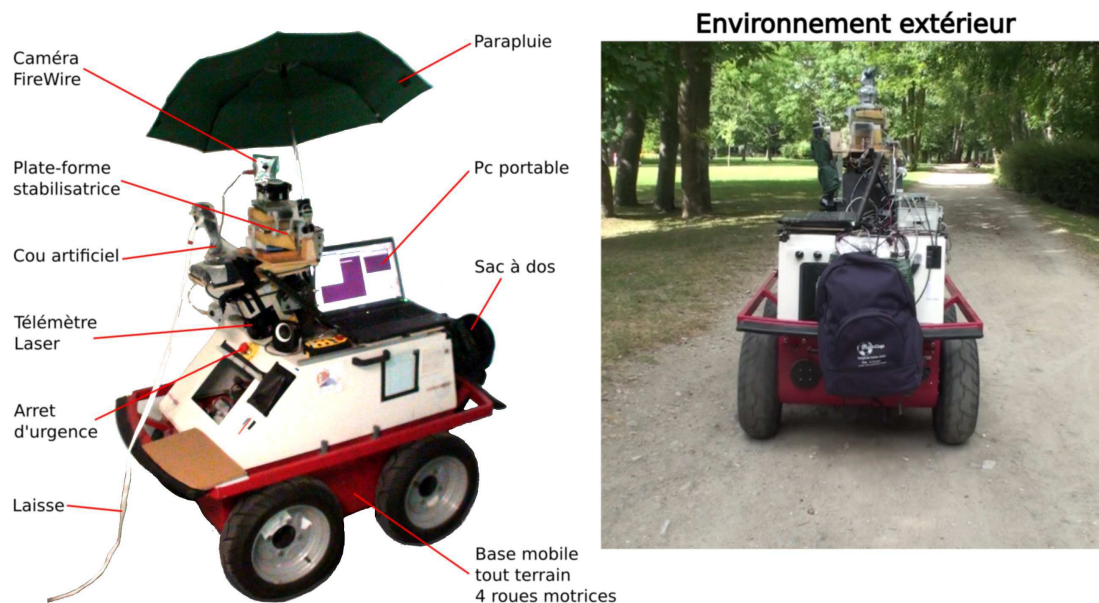


FIGURE 3 – Plateforme robotique mobile Roburoc de Robosoft. La base mobile est munie de 4 roues motrices tout terrain adaptées aux environnements extérieurs. Le corps contient le pc embarqué protégé à l’intérieur d’un coffre, un pc portable permettant un contrôle du robot. La caméra est soutenue par une plateforme stabilisatrice. L’électronique est recouverte d’un film transparent et l’ensemble est surmonté d’un parapluie déployable afin de résister à l’humidité. Une bâche étanche est transporté dans le sac à dos du robot en cas d’averse.

chine distante si besoin. Les 2 plateformes sont munies de capteurs de proximité utilisés pour l’évitement d’obstacle, d’une boussole magnétique ainsi que d’autres capteurs plus spécifiques. Nous classerons ces capteurs en deux catégories selon qu’ils mesurent un état interne au robot ou bien un état externe relatif à son environnement. Dans le premier cas, nous parlerons de *proprioception* ou de modalité sensorielle idiothétique (e.g odométrie, accéléromètre, gyroscope, indicateur de charge de la batterie). Dans le second cas, nous parlerons d’*extéroception* ou de modalité sensorielle allothétique (e.g. caméra, capteurs ultra-son, capteurs infra-rouges, capteur thermique, boussole magnétique). Pour une liste complète du matériel embarqué, on pourra se référer à l’annexe C.

Les 2 plateformes que nous utiliserons sont des robots mobiles à roues non holonômes. Cette contrainte implique que le robot ne peut effectuer de translation omnidirectionnelle. La locomotion à l’aide de roues exploite la friction de contact, dont la nature (régularité, matériaux) à une forte influence sur les propriétés du mouvement.

Dans le cas du robot d’intérieur (fig. 2), les 2 roues sont pleines et disposées de part et d’autre de la base. Le centre instantané de rotation est donc situé au milieu des 2 roues ce qui permet un rayon de braquage nul (rotation sur place à vitesse nulle). En condition intérieure (sol plan), nous considérerons qu’il y a roulement sans glissement de la roue sur le sol. Pour vérifier cette condition, il est nécessaire que le contact soit ponctuel et que les roues soit indéformables.

Dans le cas du robot d’extérieur (fig. 3), les 4 roues tout terrain sont parallèles 2 à 2, ce qui ne permet pas un rayon de braquage nul. De plus, alors qu’il est raisonnable de dire que des roues pleines sont indéformables, cette hypothèse est fautive si l’on considère des roues équipées de pneus. Une telle contrainte, ajoutée à celle d’un sol accidenté en environnement extérieur, im-

plique des glissements qui ne sont cette fois plus négligeables.

Plan de la thèse

Ce manuscrit se décompose en 7 chapitres distincts.

Dans le chapitre 1, j'essaierai d'aborder de manière didactique les outils nécessaires à la compréhension des modèles et méthodes que nous utiliserons tout au long de ces travaux. Nous aborderons les bases du formalisme neuronal, de l'apprentissage par conditionnement, de la catégorisation, de la décision ainsi que de l'importance de l'aspect dynamique dans les processus neuronaux.

Dans le chapitre 2, nous passerons en revue un état de l'art des modèles permettant la navigation autonome d'un robot mobile. Après un bref rapprochement entre les méthodes historiques utilisées pour la navigation maritimes et celles utilisées en robotique mobile, nous comparerons les méthodes dites "classiques" avec d'autres modèles inspirés du comportement animal. Nous verrons que les techniques de SLAM (Simultaneous Localisation And Mapping) se basent principalement sur l'idée de la construction d'une carte interne détaillée de l'environnement. Cette carte doit permettre au système de naviguer de manière précise dans cet environnement. Ces méthodes classiques montrent leurs limites face à des environnements dynamiques et/ou de grandes tailles. Nous verrons comment des solutions simples et robustes, abordant le problème sous un autre angle, ont pu émerger de la nature. Dans le cadre de l'approche bio-inspiré, je présenterais le modèle de cellules de lieux hippocampiques que nous exploiterons comme base du modèle de navigation visuelle dans la suite de ces travaux.

Dans le chapitre 3, je présenterai un parallèle entre l'approche neuronale et l'approche bayésienne en appliquant cette comparaison au modèle de navigation visuelle présenté précédemment. Après une description des limitations du modèle d'apprentissage figé précédent, nous procéderons à une ré-écriture du modèle sous une forme probabiliste. Je proposerai ensuite une réflexion et des solutions au problème de l'extraction de l'information pertinente pour la tâche. A partir de cette comparaison, je montrerai comment un robot est capable d'apprendre à prédire un amer visuel à partir de la perception de son azimut (sa position angulaire). De la même manière, je montrerai comment construire des cellules de lieu "probabilistes" en apprenant à prédire un lieu à partir de la perception d'une constellation d'amers-azimuts donnée. Pour cela, je proposerai un modèle neuronal capable d'estimer des densités de probabilités à partir d'observations statistiques de l'environnement. Ce modèle de prédiction uni-modal constituera la première brique d'auto-évaluation du système (l'évaluation de la pertinence des informations qu'il reçoit).

Dans le chapitre 4, nous implémenterons une nouvelle stratégie de navigation basée sur une localisation idiothétique, c'est à dire une représentation des lieux construite à partir d'informations proprioceptives. En partant d'un mécanisme simple d'intégration de chemin (présenté au chapitre 2), nous proposerons un modèle de cellules de grille, enregistrées dans le cortex entorhinal dorso-médian des rats [Hafting et al., 2005]. Nous analyserons les possibilités et limitations de ces cellules pour la navigation. Nous proposerons un modèle permettant d'obtenir des cellules de lieu idiothétiques uniquement alimentées par l'information de ces cellules de grilles.

Après avoir analysé en détails ce nouveau modèle de cellules de lieu idiothétiques, nous nous intéresserons dans le chapitre 5, à la manière dont peuvent être fusionnées les représentations issues du modèle visuel et du modèle idiothétique. Nous proposerons un modèle de fusion simple afin de générer des cellules de lieu multimodales, c'est à dire des cellules dont l'activation

dépend à la fois d'informations allothétiques et idiothétiques. Plusieurs expériences sur robot réel rendront compte de la capacité du modèle à conserver une reconnaissance cohérentes lors de perturbations extérieures (ambiguïtés visuelles, kidnapping et fermeture de boucle).

Pour finir, nous ouvrirons la discussion sur la fusion de modalités en proposant un modèle générique pour l'intégration d'informations multi-modales (ne se limitant pas à 2 modalités). Ce modèle de prédictions intermodales constituera la seconde brique d'auto-évaluation du système (l'évaluation de l'adéquation existant entre ces différentes sensations).

Le chapitre 6 se focalisera sur l'auto-évaluation des comportements exhibés par le système. En partant d'un état de l'art sur les modèles de détection de nouveauté et les théories de la perception, je proposerai un modèle capable d'apprendre à reconnaître la dynamique résultante du couplage entre sensations et actions. L'apprentissage de cette dynamique forme un modèle interne des interactions du robot avec son environnement : un modèle de ses perceptions. La surprise (erreur de prédiction) engendrée par une dynamique sensori-motrice inhabituelle renseigne alors le système sur d'éventuelles anomalies de son comportement. Ce modèle de détection de nouveauté constituera la troisième brique d'auto-évaluation du système, l'évaluation de son comportement.

Dans le chapitre 7, nous verrons que le fait de détecter un comportement inhabituel ne nous renseigne pas pour autant sur son inefficacité. Le système doit également surveiller l'évolution de la surprise dans le temps pour pouvoir juger de la pertinence de celle-ci. Je montrerais comment un contrôleur émotionnel au second ordre peut faire usage de cette détection de nouveauté pour réguler le comportement et ainsi permettre au robot d'utiliser la stratégie la plus adaptée à la situation rencontrée. Par ailleurs, un modèle simpliste de frustration poussera le robot à initier une interaction avec son professeur lorsqu'il détecte qu'il se retrouve dans une impasse. Ce mécanisme permet au robot d'identifier les situations qu'il ne maîtrise pas dans le but d'affiner son apprentissage par le biais d'une interaction qu'il à lui même initié, à l'instar de certains processus développementaux.

Pour finir, nous discuterons des différentes problématiques abordées ainsi que des apports de ces travaux relatifs à l'autonomie en robotique mobile.

Une chose est particulièrement frappante dans le cas de la colonie d'insectes : contrairement à ce qui se passe avec le cerveau, nous n'avons aucun mal à admettre deux choses : a) la colonie est composée d'individus ; b) il n'y a pas de centre ou de "moi" localisé. Pourtant, l'ensemble se comporte comme un tout unitaire et, vu de l'extérieur, c'est comme si un agent coordinateur était "virtuellement" présent au centre. C'est ce que j'entends lorsque je parle d'un moi dénudé de moi : une configuration globale et cohérente qui émerge grâce à de simples constituants locaux, qui semble avoir un centre alors qu'il n'y en a aucun [...] pp. 85-87

– F. Varela. "Quel savoir pour l'Éthique"

CHAPITRE 1

Outils du cybernéticien

Le but de ce chapitre est de fournir une vue d'ensemble des outils nécessaires à la compréhension des modèles et méthodes développé(e)s tout au long de ce manuscrit. Les notions présentées seront abordées de manière volontairement didactique afin de rendre la lecture de cette thèse accessible au lecteur non-initié.

Nous aborderons brièvement les bases du formalisme neuronal en présentant le fonctionnement d'un neurone biologique et de son homologue, le neurone formel. Nous parlerons de *groupe* neuronal pour désigner une population de neurones partageant des caractéristiques communes : une fonction et/ou une règle d'apprentissage identique.

Nous introduirons ensuite le concept d'apprentissage, concept clé pour l'ensemble des applications utilisant des réseaux de neurones artificiels. Un réseau de neurones peut apprendre en modifiant la valeur de ses poids synaptiques en fonction des entrées qu'il reçoit. La modification de ces poids autorise l'apprentissage d'associations et la catégorisation des signaux d'entrées en différentes classes. Un signal de *neuromodulation* permet de moduler la vitesse d'apprentissage du réseau.

Ces capacités d'apprentissages sont cruciales pour rendre un agent autonome et capable de faire face aux aspects dynamiques, changeants et complexes du monde réel. Elles permettent de capturer les régularités statistiques du monde en filtrant naturellement le bruit présent. Elles permettent aussi à l'animal de s'adapter au mieux à son environnement, par la capacité à reconnaître des lieux pour se déplacer efficacement, la capacité d'apprendre de ses erreurs ou encore celle de détecter une situation anormale et fuir en cas de danger.

Le signal de neuromodulation peut provenir de stimuli externes (renforcement positif ou négatif de la part de l'environnement ou d'un autre animal), mais il peut aussi être généré en interne par certains mécanismes compétitifs et l'usage de seuils liés aux niveaux physiologiques de l'animal (faim, soif, émotions, etc...). Ces mécanismes générant des signaux de modulation de l'apprentissage et/ou du comportement ne sont pas directement liés à l'environnement et sont principalement issus de décisions internes. Cela pose la question de comment fonctionnent ces processus décisionnels à l'oeuvre dans le cerveau. De tels mécanismes sont utilisés dans le cadre de *l'apprentissage non-supervisé*, c'est à dire un apprentissage autonome. L'agent prend la décision d'apprendre ou non en fonction de ses propres motivations et non de facteurs extérieurs.

L'apprentissage n'est pas la seule caractéristique intéressante de l'approche neuronale. Nous aborderons succinctement l'intérêt du codage en population et celui des réseaux récurrents. Plus particulièrement, nous nous attarderons sur les mécanismes de compétition, très utilisés dans

mes travaux. L'intérêt d'une compétition réside dans le fait que des neurones en rivalité réalisent un processus décisionnel doté de propriétés intéressantes. La décision, nous le verrons, est également une caractéristique essentielle pour la survie et l'autonomie d'un agent.

1.1 Une brève introduction aux réseaux de neurones artificiels

Un réseau de neurones artificiels est un modèle computationnel utilisé pour réaliser un ensemble de traitements en certains points semblables à son homologue biologique. Le premier modèle mathématique du neurone biologique a été proposé par McCulloch et Pitts en 1943 [McCulloch et al., 1943] et a été repris en 1958 par Rosenblatt [Rosenblatt, 1958] dans son célèbre modèle du *Perceptron*. Le perceptron est un classifieur linéaire composé de couche de neurones formant un réseau. Chaque neurone du réseau est une unité computationnel effectuant une somme pondérée de ses entrées. L'information en entrée peut provenir d'un capteur (e.g. rétine) ou bien de la sortie d'autres neurones. Chaque entrée e est pondérée d'un poids synaptique w , modulant l'importance de certaines entrées par rapport à d'autres dans le résultat en sortie (voir fig. 1.1).

Pour n entrées, le calcul de l'activité (ou potentiel) P_i d'un neurone i est donc défini tel que :

$$P_i = \sum_1^n x_n \cdot w_n \quad (1.1)$$

Ce potentiel de sortie peut ensuite subir une transformation non-linéaire par le biais d'une fonction d'activation, usuellement un seuillage ou bien une base radiale (e.g. base sigmoïdale), prodiguant ainsi nombre de propriétés intéressantes. Il est important de bien choisir la fonction d'activation pour obtenir un modèle utile en pratique. Dans la suite de ces travaux, nous utiliserons principalement une fonction d'activation linéaire ou bien un seuillage binaire tel que :

$$\text{Seuillage binaire : } S_b(P) = \begin{cases} 1 & \text{si } P > S \\ 0 & \text{sinon.} \end{cases} \quad (1.2)$$

Avec S , un seuil particulier.

Bien que la réponse réelle d'un neurone biologique soit souvent impulsionnelle (spike), le modèle approxime l'activité de sortie par un potentiel d'activité, ou fréquence moyenne de décharge. Certains modèles plus fins prennent en compte cet aspect impulsionnel (spiking neurons) dont le premier a été introduit par [Hodgkin and Huxley, 1952].

Nous allons maintenant nous concentrer plus précisément sur les règles qui permettent à ces réseaux d'apprendre par association.

1.2 Conditionnement

L'apprentissage associatif, ou apprentissage par conditionnement est une des techniques d'apprentissage les plus simples observée chez l'animal. C'est le physiologiste Pavlov qui, le premier, introduisit la notion de conditionnement classique [Pavlov, 1927] illustré dans sa fameuse expérience de conditionnement du chien (voir fig. 1.2). Pavlov avait remarqué que la vue ou l'odeur de la nourriture (stimulus inconditionnel (US)) activait de manière réflexe une

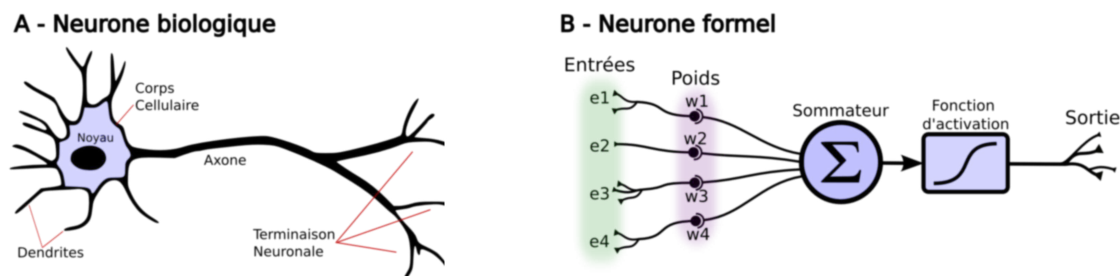


FIGURE 1.1 – Comparaison entre neurone biologique et neurone formel. **A** - Le neurone biologique intègre l'activité présente sur ses dendrites (entrées). La fréquence (ou potentiel) de décharge de l'axone (sortie) dépend de cette activité dendritique. Les terminaisons neuronales viennent se connecter aux dendrites d'autres neurones via des synapses. **B** - Le fonctionnement du neurone formel est assez similaire : la sortie est fonction de la somme des entrées par les poids synaptiques. L'utilisation d'une fonction d'activation radiale permet de traiter des problèmes non-linéaires.

salivation de la part du chien. Pavlov présentait ensuite, de manière répétée, le son d'une cloche (stimulus conditionnel (CS)) à chaque fois que la nourriture est présentée au chien. Au bout d'un certain temps, le son de la cloche est associé à la nourriture et active donc la salivation du chien, et ce même lorsque la nourriture n'est pas présentée. L'apprentissage consiste donc en l'association d'un stimulus nouveau à un stimulus réflexe (e.g. si une réaction réflexe R survient suite à la perception d'un stimulus particulier (e.g. une odeur), il est possible d'apprendre à déclencher cette même réaction R suite à la perception d'un son en faisant systématiquement précéder l'odeur par le son).

Quelques années plus tard, Thorndike et Skinner ont introduit la notion de conditionnement instrumental (ou opérant) [Skinner, 1938] qui rend compte de l'aptitude des animaux à apprendre le lien entre 2 événements et ce quel que soit la nature de ces événements (stimulus externe, action, etc..). Skinner distingue le conditionnement opérant du conditionnement classique par le fait que la réponse ne soit pas une réaction réflexe de l'organisme.

L'animal est donc capable d'apprendre une association entre 2 stimuli présentant un lien apparent. Ce lien apparent amène au concept clé de **corrélation**. 2 stimuli sont corrélés s'ils sont liés d'une manière ou d'une autre. Le type le plus simple de liaison est la relation affine. Dans le cas de deux variables, elle se calcule au travers d'une régression linéaire. La mesure de la corrélation linéaire entre les deux se fait alors par le calcul du coefficient de corrélation linéaire. Ce coefficient est égal au rapport de leur covariance et du produit de leurs écarts types. Le coefficient de corrélation est compris entre -1 (anti-corrélation) et 1 (corrélation maximum). Attention toutefois à ne pas confondre corrélation et causalité. Deux événements peuvent être corrélés sans pour autant impliquer un rapport de cause à effet. Par exemple, la probabilité de mourir n'augmente pas lorsque l'on se trouve dans un lit d'hôpital, mais on y est généralement pour soigner une maladie, et lorsque l'on est malade la probabilité de mourir est plus grande. Confondre ces 2 concepts peut amener à des raisonnements biaisés comme en joue l'humoriste Coluche :

« 1/3 des accidents de la route étant dus à des conducteurs alcooliques, qu'attend-on pour punir les 2/3 de conducteurs sobres responsables de la majorité des accidents ? »

L'un des algorithmes applicables pour modéliser un conditionnement classique est le principe de réduction des erreurs quadratiques moyennes (*Least Mean Square (LMS)*) proposé en 1960 par [Widrow and Hoff, 1960]. Cet algorithme permet un apprentissage supervisé d'un flux d'ac-

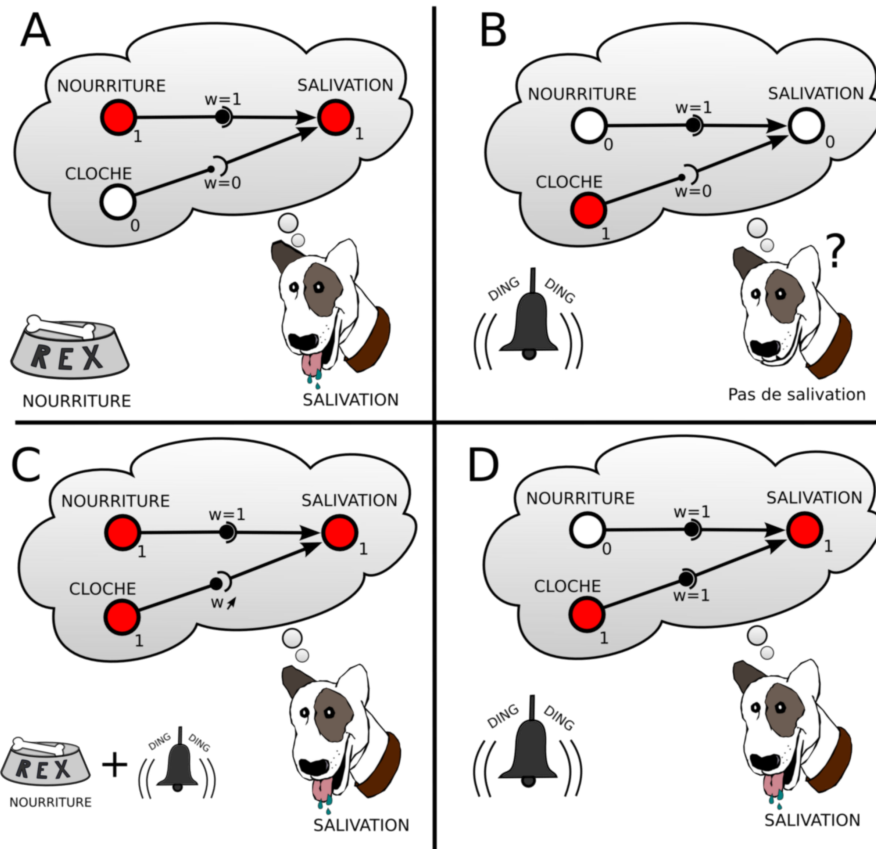


FIGURE 1.2 – Conditionnement classique : **A** - La vue de la nourriture (stimulus inconditionnel) fait saliver le chien de façon réflexe (poids synaptique = 1). **B** - Le son d'une cloche (stimulus neutre) n'a au départ aucun effet sur la salivation (poids synaptique nul). **C** - Pendant la phase de conditionnement, on présente de façon répétée la nourriture et le son de cloche simultanément. Le poids synaptique liant la cloche et la salivation se renforce progressivement. **D** - Une fois le conditionnement terminé, le son de la cloche activera la salivation même en l'absence de nourriture.

tivité d'entrée. C'est une méthode d'optimisation qui consiste en la modification des poids synaptiques w jusqu'à trouver l'erreur quadratique moyenne minimale entre l'entrée e et la sortie désirée Sd .

Dans l'exemple suivant, nous considérons un réseau à 3 neurones. Un neurone code pour le signal d'entrée E_i , un autre pour la sortie désirée Sd_j et enfin un troisième représente la sortie S_j du système. Ici il n'y a donc qu'un seul poids synaptique w_{ij} , reliant l'entrée E_i à la sortie S_j . Le système commence à apprendre lorsque qu'il existe une différence de potentiel entre la sortie réelle S et la sortie désirée Sd . Son unique but est de faire diminuer cette différence en modifiant la valeur du poids w_{ij} . L'adaptation du poids w_{ij} est d'autant plus forte que la différence $Sd_j - S_j$ est grande. Voici l'équation décrivant cette règle d'apprentissage :

$$\Delta w_{ij} = \lambda E_i \cdot (Sd_j - S_j) \quad (1.3)$$

Le paramètre λ permet de contrôler la vitesse d'apprentissage ($\lambda < 1$). A chaque itération, l'algorithme calcule également la nouvelle valeur du poids w_{ij} et de la sortie S_j :

$$w_{ij}(t + 1) = w_{ij}(t) + \Delta w_{ij}(t) \quad (1.4)$$

Ces apprentissages par conditionnement ont été largement utilisés dans le cadre de recherches impliquant des réseaux de neurones artificiels [Grossberg and Morahan, 1971; Grossberg, 1972a; Rescorla, 1967]. Il existe également d'autres méthodes d'apprentissage associatif comme la règle de Hebb [Hebb et al., 1949] ou encore la classe des apprentissages par renforcement initiée par [Klopf and Laboratory, 1972; Klopf, 1975] puis adapté par [Anderson et al., 1983]. L'apprentissage par renforcement fait référence à une classe de problèmes dont le but est d'apprendre à partir d'expériences, ce qu'il convient de faire en différentes situations de façon à optimiser une récompense au fil du temps. On peut par exemple considérer un agent autonome, plongé au sein d'un environnement, devant prendre des décisions en fonction de son état courant. En retour, l'environnement procure à l'agent des récompenses, qui peuvent être positives ou négatives. L'agent cherche un comportement décisionnel optimal appelé stratégie ou politique : une fonction associant à l'état courant l'action à exécuter. Cette politique est optimale si elle permet à l'agent de maximiser la somme des récompenses au cours du temps.

Parmi les premiers algorithmes d'apprentissage par renforcement, on compte le *TD-learning* (Temporal Difference learning) initié par [Sutton, 1988] et une de ses variantes célèbre : le *Q-learning* (Quality learning) proposé par [Watkins and Dayan, 1992]. On pourra se référer à [Wiering and van Otterlo, 2012] pour un état de l'art sur les différentes méthodes d'apprentissage par renforcement.

Formellement, la base du modèle d'apprentissage par renforcement consiste en :

1. un ensemble d'états S de l'agent dans l'environnement.
2. un ensemble d'actions A que l'agent peut effectuer.
3. un ensemble de valeurs scalaires "récompenses" R que l'agent peut obtenir.

À chaque pas de temps t , l'agent perçoit son état S_t et l'ensemble des actions possibles dans cet état $A(s_t)$. Il choisit une action qui implique un changement d'état et parfois une récompense. L'algorithme doit permettre à l'agent de développer une politique $\Pi : S \rightarrow A$ qui lui permette de maximiser le cumul de récompenses $R = \sum_t \lambda_t \cdot r_t$ (λ_t est un facteur compris entre 0 et 1 permettant de prendre en compte les récompenses plus ou moins loin dans le futur pour le choix des actions de l'agent).

L'apprentissage par renforcement est particulièrement adaptée aux problèmes nécessitant un compromis entre la quête de récompenses à court terme et celle de récompenses à long terme.

Nous allons nous intéresser maintenant aux mécanismes génériques de compétition qui seront utilisés tout au long de ces travaux.

1.3 Compétition

Les mécanismes de compétition sont un type de réseaux dont la topologie caractéristique (la façon dont les neurones sont connectés à leur voisins) permet à chaque neurone de se mesurer aux autres pour le partage de l'activité neuronale. La compétition est le résultat d'une dynamique d'inhibition réalisée par le biais de connections latérales inhibitrices. Ces processus naturels de compétition permettent une prise de décision, à l'instar d'un système de vote. Prendre des décisions s'avère bien souvent être une caractéristique indispensable pour assurer la survie de l'animal.

1.3.1 Winner Takes All

L'idée de base du *Winner Takes All* (WTA) provient des modèles connexionnistes [Grossberg, 1976, 1988; Kohonen, 1984; Feldman and Ballard, 1982]. Ils permettent de simuler les mécanismes de compétition existant au sein d'une population de neurones. Des liaisons inhibitrices latérales permettent de simuler le processus de compétition. Après convergence, seul le neurone ayant la plus grande activité reste actif et inhibe tous les autres [Rumelhart and Zipser, 1986; Lippmann, 1987; Carpenter and Grossberg, 1988; Lee et al., 1999].

Considérons deux neurones i et j possédant chacun une seule entrée et une seule sortie (voir fig. 1.3). Une excitation en entrée activera le neurone concerné qui déchargera à son tour (potentiel d'activité sur sa sortie). Supposons que les 2 neurones soit connecté entre eux par des connexions inhibitrices, de telle sorte que lorsque i est activé, il tende à inhiber j , et réciproquement. Une liaison inhibitrice sera représentée par un poids synaptique négatif. Supposons maintenant

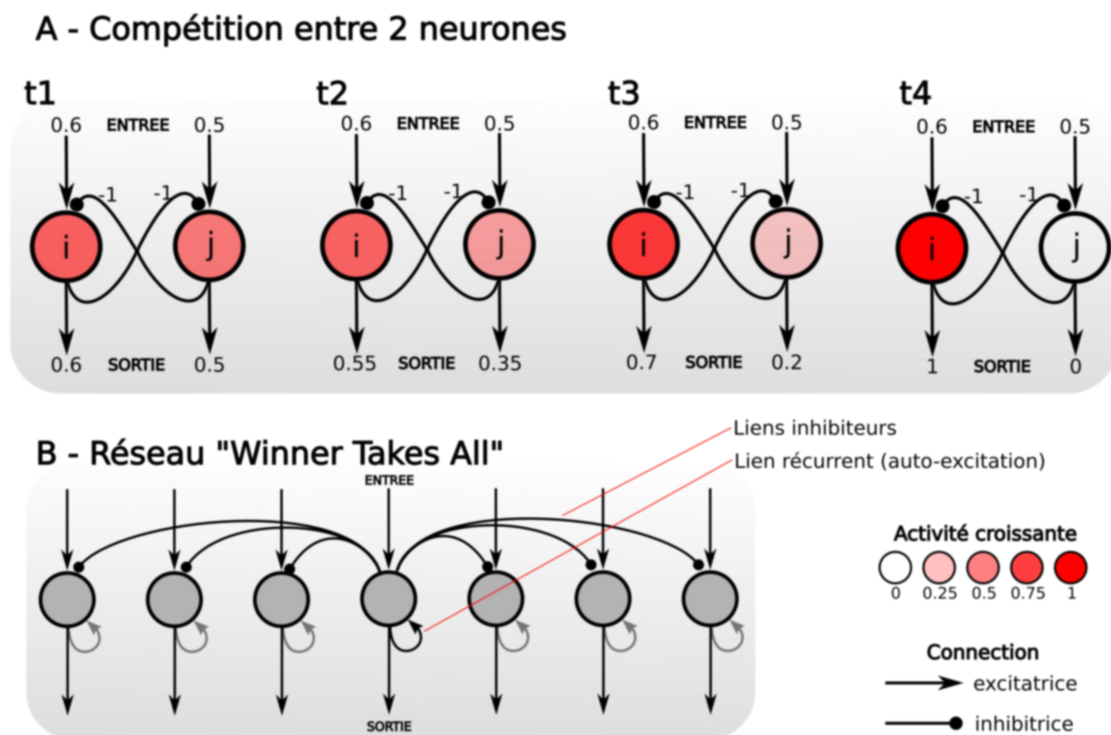


FIGURE 1.3 – Principe d'un mécanisme de compétition. **A** - Réseau minimaliste de compétition entre 2 neurones i et j . Chaque neurone possède une seule entrée et une seule sortie. Chaque neurone est relié à l'autre par une connexion inhibitrice (-1). Au temps t_1 , i et j reçoivent une excitation très similaire (respectivement 0.6 et 0.5). Au fil du temps, le neurone le plus actif prend le dessus sur l'autre (t_2 , t_3 puis t_4). A régime permanent (t_4), le réseau est parvenu à prendre une décision. **B** - Un réseau de compétition comportant plus de neurones fonctionne selon le même principe : chaque neurone tente d'inhiber les autres via des connexions inhibitrices. Pour une prise de décision plus rapide, les neurones peuvent s'auto-exciter via une connexion récurrente.

que les 2 neurones soient excités par une activité très similaire (e.g. 0.5 et 0.6), alors au fil du temps le réseau s'éloignera de cet état instable. L'unité la plus active au départ tendra vers une activité maximum et la moins active vers une activité minimum. Ce principe peut se généraliser à un plus grand nombre d'unités. Il suffit que chaque neurone inhibe l'ensemble des autres neu-

rones par des connections latérales négatives/inhibitrices. En plus des connections inhibitrices, chaque neurone peut également s'auto-exciter par le biais d'un lien récurrent, ce qui accélère le processus de déséquilibre.

La variation du potentiel d'action d'un neurone x_i du groupe est définie par :

$$\frac{dx_i}{dt} = \alpha x_i - \beta \sum_{k \neq i}^K x_k + I_i \quad (1.5)$$

Avec K le nombre de neurones du groupe, I_i l'activité entrante du neurone i , $\alpha \in 0, 1$ le poids du lien récurrent autoexcitateur du neurone i et $\beta \in 0, 1$ le poids des liens inhibiteurs provenant des autres neurones du groupe.

Un réseau possédant ce type de topologie inhibitrice assure qu'un seul et unique neurone possède tout le potentiel d'activité lorsque le réseau a atteint un régime permanent.

Ce mécanisme très simple permet de faire **émerger** une fonction de prise de décision en utilisant un nombre minimum de règles rudimentaires impliquant un équilibre instable, sans besoin de prescrire une procédure complète.

Le fait de jouer sur la fonction d'activation des neurones permet d'obtenir une compétition plus ou moins stricte. Plusieurs gagnants peuvent ainsi être conservés ("*Soft WTA*" ou "*k-WTA*") [Maass, 2000].

1.4 Classification

Pour reconnaître, apprendre et agir dans un environnement aussi complexe que celui que nous connaissons, un des aspects nécessaires est de recourir à des méthodes de *classification*. Ces méthodes permettent de capturer les caractéristiques importantes des signaux sensoriels pour en former des abstractions, c'est à dire des informations de plus haut niveau permettant à l'animal de généraliser.

La classification est un domaine vaste qui propose des méthodes pour réaliser un partitionnement de données (data clustering). Elle permet de diviser un ensemble de données en différents ensembles homogènes, en ce sens que les données de chaque sous-ensemble partagent des caractéristiques communes, qui correspondent le plus souvent à des critères de proximité que l'on définit en introduisant des mesures de distances ou de similarités entre objets. La classification peut être supervisée ou non. Ici nous nous limiterons au cas non-supervisé.

L'algorithme des **k-moyennes** (k-means) [MacQueen, 1967] est l'une des méthodes d'apprentissage non-supervisé les plus simples qui illustre bien la façon dont il est possible de classifier des données. Cette méthode propose de catégoriser un ensemble de données au sein d'un certain nombre de classes (clusters), dont le nombre est fixé a priori. L'idée maîtresse est de définir k centroïdes, un par classe. Ces centroïdes sont placés aléatoirement dans l'espace des données tout en s'assurant qu'ils soient autant que possible loin les uns des autres (voir fig. 1.4). Chaque donnée est associée au centroïde le plus proche. A chaque itération, la position de chaque centroïde est mise à jour en calculant le barycentre des données qui lui sont associées. L'opération est répétée jusqu'à atteindre un état de convergence à partir duquel les positions des centroïdes se sont stabilisées. Le but de l'algorithme est de minimiser la fonction de l'erreur quadratique J .

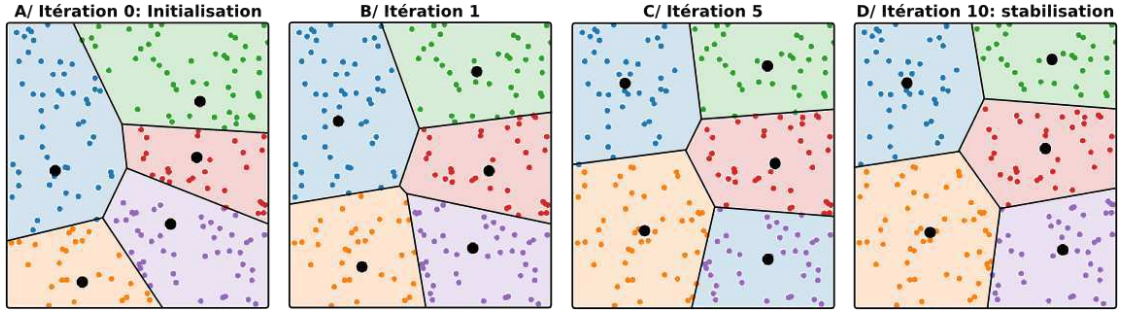


FIGURE 1.4 – Classification par la méthode des k-moyennes (k-means) : illustration du principe pour 5 classes ($k=5$). **A** - A l'initialisation, $k=5$ classes (points noirs) sont réparties de manière aléatoire dans l'espace des données (points colorés). 5 catégories (zones colorés) sont créées en associant chaque donnée à la classe la plus proche. Les partitions représentent un diagramme de Voronoï [Voronoi, 1908]. **B, C** - A chaque itération, le centroïde de chaque catégorie devient la nouvelle moyenne de la classe. Cette opération est répétée jusqu'à atteindre convergence. **D** - Après convergence, les moyennes de chaque classe ne sont plus modifiées. Le système a atteint une catégorisation optimale des données.

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (1.6)$$

Où $\|x_i^{(j)} - c_j\|^2$ est une mesure de distance entre une donnée $x_i^{(j)}$ et le centroïde c_j . Le résultat est un partitionnement de l'espace des données en cellule de Voronoï [Voronoi, 1908]. Cet algorithme réalise la discrétisation de l'espace d'entrée en ne modifiant à chaque cycle d'adaptation qu'un seul vecteur référent. Le processus d'apprentissage est donc très long. L'algorithme de [Kohonen, 1984] propose une amélioration de la méthode des k-moyennes en tirant profit des relations de voisinage pour réaliser une discrétisation dans un temps beaucoup plus court.

Dans nos travaux, certains groupes de neurones seront utilisés en tant que groupe de catégorisation, au sein desquels chaque neurone peut représenter une catégorie. Au départ, les poids synaptiques sont initialisés avec des valeurs aléatoires permettant l'émergence d'un neurone gagnant, élu représentant de la première catégorie codant pour l'entrée perçue. Un paramètre de vigilance permet de définir la granularité des catégories. Lorsque l'entrée change, le groupe peut être amené à recruter un nouveau neurone (nouvelle catégorie) si l'activité du neurone gagnant devient inférieure à un paramètre de "vigilance". La modulation du seuil de vigilance par un signal de neuromodulation permettra de modifier, à la volée, la granularité des catégories apprises. Une vigilance faible impliquera une granularité grossière, et réciproquement. L'apprentissage est là aussi réalisé au sein des poids, dont la modification au fil du temps est définie par :

$$\Delta w_{ij} = \delta_j^k (a_j(t) \cdot E_i) \quad (1.7)$$

avec k l'indice du neurone le plus actif ($k = \text{ArgMax}(C_j)$), δ_j^k la fonction de Kronecker¹

1. Le symbole de Kronecker est une fonction à deux variables égale à 1 si celles-ci sont égales, et 0 sinon.

$$\delta_i^k = \begin{cases} 1 & \text{si } i = k \\ 0 & \text{sinon.} \end{cases} \quad (1.8)$$

utilisée pour ne modifier les poids des liens que vers la catégorie gagnante. $a_j(t)$ est une fonction qui vaut 1 lorsque le neurone j est recruté et 0 sinon. Afin d'éviter un désapprentissage, le neurone est recruté aléatoirement mais seulement s'il ne code pas déjà pour une catégorie. Au moment du recrutement, on suppose qu'un mécanisme de compétition permet de mettre à 1 son activité (et les autres à 0), ce qui permet de ne modifier ensuite que les poids synaptiques des liens entre la couche d'entrée et le neurone recruté. Les poids des liens entrant du neurone recruté sont modifiés pour correspondre au motif présent en entrée (voir fig. 1.5). Intuitivement, la forme d'entrée est *imprimée* (copiée) dans les poids, ce qui permet au neurone recruté de "*se souvenir*" de cette forme ultérieurement. L'activité de ce neurone est ensuite directement dépendante de la nouvelle entrée qu'il perçoit. Si la forme perçue (en entrée) est identique à la forme apprise précédemment (dans les poids), alors l'activité du neurone est maximale ($C_k = 1$) : le neurone reconnaît la forme. Au contraire, si la forme perçue est complètement différente de celle apprise, l'activité du neurone est minimale ($C_j = 0$) : le neurone ne reconnaît en rien cette nouvelle forme.

A chaque itération, l'activité C_j du neurone j d'un groupe de catégorisation est définie par :

$$C_j = 1 - \frac{1}{N} \sum_{i=1}^N |w_{ij} - E_i| \quad (1.9)$$

avec N le nombre de neurones du groupe, E_i l'entrée et w_{ij} le poids synaptique du lien entre E_i et C_j .

1.4.1 Champs de neurones dynamiques

S'inspirant des mêmes principes, beaucoup de recherches se sont consacrées à l'étude d'une population de neurones (champ de neurones) plutôt qu'à celle du neurone particulier. Le nombre de neurones présents dans ne serait-ce qu'une petite portion du cortex est immense. Partant de ce constat, une nouvelle approche est apparue pour étudier des réseaux de neurones au sein desquels l'espace est considéré comme continu. La topologie est telle que des neurones voisins partagent des caractéristiques et activités semblables et donc que l'*espace neuronal* possède une certaine continuité. Cette notion de continuité peut être approximée par un **codage en population**, par opposition au concept de **neurone grand mère**. Le neurone grand mère est un concept populaire qui consiste à considérer chaque neurone comme susceptible d'encoder une idée abstraite à lui tout seul. Selon ce principe, il existerait un neurone précis dans notre cerveau qui s'active uniquement lorsque nous pensons à notre grand mère. De la même manière, un autre neurone s'activerait lorsque l'on pense à un objet particulier ou un concept donné. Au vue du caractère extrêmement distribué et redondant du cerveau, cette théorie est en fait peu plausible même si elle présente l'intérêt certain d'être intuitive et accessible.

Le codage en population est, quant à lui, un moyen de coder l'information de manière distribuée justement, au sein d'un groupe de neurones. Dans un tel codage, chaque neurone possède une portion de l'information, mais l'information complète est uniquement accessible lorsque l'on considère la réponse combinée d'un grand nombre de ces neurones. Un avantage de ce type de codage pour la survie d'un animal est son caractère extrêmement redondant. Le réseau peut alors se permettre la perte de quelques neurones sans que cela n'influe de manière dramatique sur l'information globale traitée. Dans les chapitres suivants, nous utiliserons l'une ou l'autre de

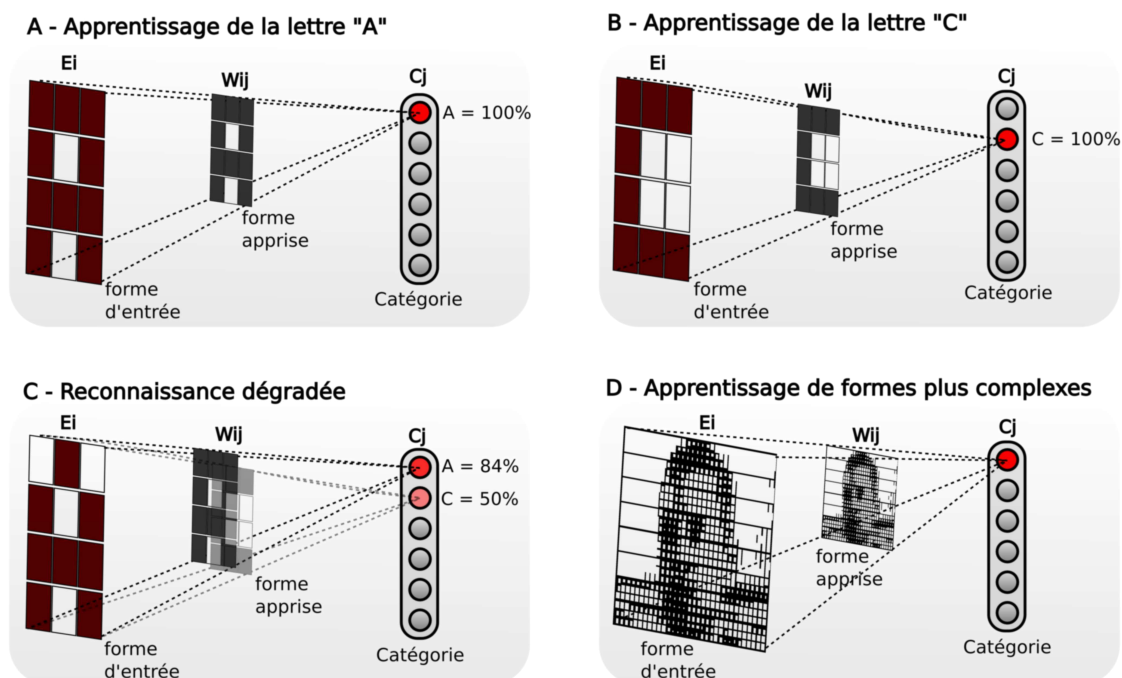


FIGURE 1.5 – Principe d'apprentissage et reconnaissance de formes. **A** - Suite à un signal de neuromodulation, un neurone est choisi pour coder une nouvelle catégorie. La forme présente en entrée (la lettre A) est apprise en étant mémorisée dans les poids synaptiques des liens entre la couche d'entrée et le neurone recruté. **B** - Un nouveau motif est présent en entrée (la lettre C). De la même façon, la catégorie est apprise par un nouveau neurone dont les poids se modifient. **C** - Après apprentissage, chaque neurone de sortie tente de reconnaître sa catégorie dans le nouveau motif présent en entrée. Ici, le motif présenté est une version dégradée de la lettre A. Le neurone codant pour la catégorie A est donc plus fortement activé que celui codant pour C.

ces représentations en fonction du problème à traiter.

A priori, la première tentative pour développer une théorie du continuum dans un champ neuronal peut être attribuée à Beurle en 1956 [Beurle, 1956] qui observe des **bulles d'activité** dans un tissu cérébrale. La théorie est reprise plus tard par Griffith [Griffith, 1963], puis par Wilson et Cowan [Wilson and Cowan, 1973] qui ajoutent les concepts d'inhibitions et excitations du voisinage ainsi que la notion de récurrence. Finalement, le modèle le plus utilisé actuellement nous vient d'Amari [Amari, 1977] qui proposa sa théorie des **champs de neurones dynamiques** quelques années plus tard. Selon cette théorie, les excitations sont locales tandis que les inhibitions ne sont que distales. Le noyau d'interaction est un laplacien de gaussienne dont la forme particulière rappelle celle d'un *Chapeau Mexicain*.

Le modèle d'Amari

L'intérêt du modèle d'Amari est qu'il possède quelques propriétés intéressantes de filtrage, de stabilité ainsi qu'un compromis entre compétition et coopération, là où un réseau de type WTA reste sensible et seulement compétitif.

Considérons un champ de neurones (une population de neurones liés de proche en proche, formant une carte de dimension x). Chaque neurone est connecté à ses voisins suivant une règle définie par le **noyau d'interaction**. Ce noyau est ici construit à l'aide d'une différence de

gaussiennes (DoG), ce qui lui confère cet aspect reconnaissable de chapeau mexicain. Le centre du noyau est excitateur, tandis que les bords sont inhibiteurs (voir fig. 1.6).

La dynamique du champ est modélisée par l'équation suivante :

Champs de neurones Mono-dimensionnel

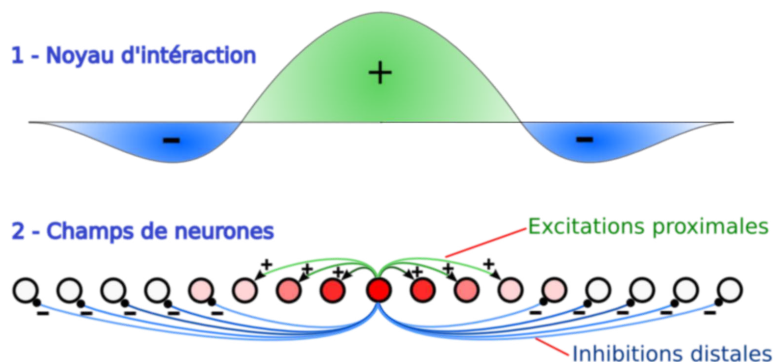


FIGURE 1.6 – Principe du codage en population. 1 : Le noyau d'interaction est obtenu à partir d'une *différence de gaussiennes* (DoG) centre-ON, donnant un profil en chapeau mexicain excitateur (+) en son centre et inhibiteur (-) sur ses bords. 2 : Champ de neurones avec interactions latérales correspondantes. Chaque neurone excite ses voisins proches et inhibe le voisinage lointain. L'information n'est pas encodée dans un unique neurone, mais est distribuée au sein d'une population, formant ainsi une **bulle d'activité**.

$$\tau \cdot \frac{u(x, t)}{dt} = -u(x, t) + I(x, t) + h + \int_{z \in V_x} w(z) \cdot f(x - z, t) dz \quad (1.10)$$

Où $u(x, t)$ représente l'activité d'un neurone x au temps t . $I(x, t)$ est l'entrée du système. h est une constante négative qui assure la stabilité du réseau. τ est le taux de relaxation du système. w est le noyau d'interaction utilisé pour l'activation du champ. V_x est l'intervalle d'interaction qui définit le voisinage. Cette équation confère à l'ensemble de la population des propriétés d'attracteurs qui correspondent à des maxima locaux du champ d'activité. Le réseau réalise également un compromis entre compétition et coopération qui nous sera utile par la suite. Par exemple, imaginons que 2 bulles d'activité soient présentes en entrée (voir fig. 1.7). Si les 2 bulles sont suffisamment éloignées, la dynamique du champ tendra à en sélectionner une seule (mode compétition). Par contre, si les 2 entrées sont suffisamment proches (dépendant de la largeur du noyau), le champ convergera vers une seule et unique bulle en sortie. Cette bulle résulte de la moyenne des 2 entrées (mode coopération), et son amplitude est supérieure à celle des entrées (se référer à [Schöner et al., 1995] pour une démonstration mathématique).

La forme du noyau d'interaction est importante et doit être paramétrée en fonction de la tâche désirée. La largeur de la bulle d'excitation (partie positive du noyau) peut être modulée par exemple. Notons qu'un réseau de type "Winner Takes All" est un cas particulier de champ de neurones pour lequel la bulle d'excitation du noyau est réduite au minimum (1 seul neurone). Le réseau comprend également des liens récurrents (un pour chaque neurone), qui ne sont pas illustrés sur les figures. Cette récurrence permet un lissage temporel et donc un filtrage efficace contre le bruit ou les oscillations présentes en entrée. Par exemple, si une entrée apparaît de manière épisodique suite au dysfonctionnement d'un capteur, la dynamique temporelle du réseau ne lui laissera pas le temps de s'imposer. Il est nécessaire qu'une entrée reste stable et immobile

pendant un certain temps pour être considérée comme *digne de confiance* et ainsi faire émerger une bulle d'activité en sortie.

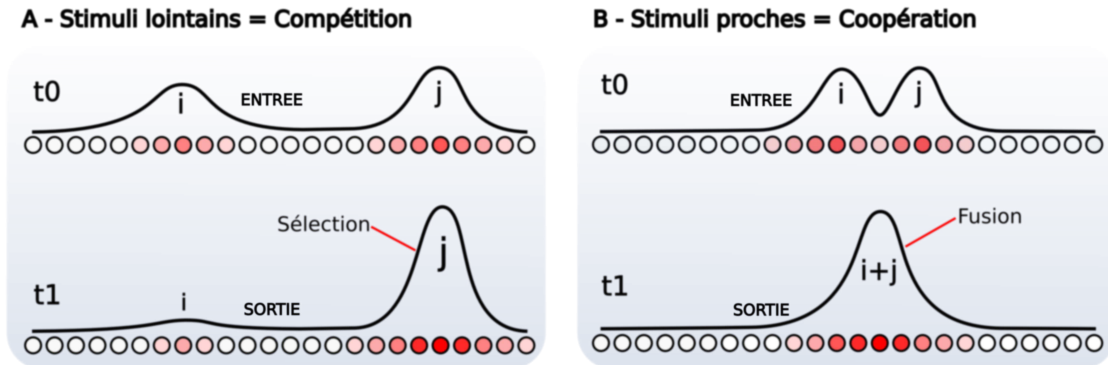


FIGURE 1.7 – Compromis sélection/coopération. **A** - 2 entrées distantes à t_0 tentent de s'inhiber l'une l'autre puisque les bords du noyau d'interaction sont inhibiteurs. Le réseau converge vers une unique bulle d'activité à t_1 . Le champ se comporte alors comme un réseau de compétition (WTA). **B** - 2 entrées proches viennent à fusionner grâce au centre exciteur du noyau d'interaction. La bulle résultante est une moyenne des bulles d'entrée.

Le contrôle moteur dynamique de Schöner

Les champs de neurones d'Amari ont été utilisés pour résoudre le problème du contrôle moteur, notamment grâce à la théorie des champs dynamiques (*Dynamic Neural Fields - DNF*) de [Schöner et al., 1995]. En conservant les propriétés intéressantes des champs de neurones cités plus haut, Schöner propose un moyen d'utiliser la continuité des bulles d'activités comme un gradient que le système peut remonter, en considérant comme sortie motrice la dérivée locale du champ d'activité. Nous utiliserons cette théorie comme référence pour le contrôle moteur de nos robots dans la suite de ces travaux, c'est pourquoi il convient d'entrer plus en détails dans la manière dont nous faisons usage de celle-ci.

Prenons l'exemple d'un robot, à la morphologie minimaliste, constitué d'un corps circulaire et de 2 roues motrices centrées de part et d'autre de ce corps. Les rotations sur place sont donc possibles. Le robot est muni de capteurs divers (caméra(s), détecteurs ultra-son, capteurs infrarouge, etc...) lui permettant de percevoir l'environnement tout autour de lui.

Nous nous intéressons à la manière idéale de représenter les informations provenant de ses capteurs dans le but de réaliser une tâche de navigation. On partira de l'a priori que le robot est non-holonome, naviguant dans un espace à 2 dimensions seulement (translation avant-arrière, rotation gauche-droite). La troisième dimension spatiale sera négligée. Le comportement de notre robot est donc soumis à 2 variables, qui sont la **vitesse linéaire** et la **vitesse de rotation**.

La vitesse linéaire pourra être constante (dans le cas le plus simple), ou bien modulée par la proximité des obstacles alentours (plus l'obstacle est proche, plus le robot ralentit).

La vitesse de rotation devra être modulée par les motivations du robot. Par motivations nous entendons les objectifs internes du robot permettant de contrôler son comportement (drives). Ceux-ci pourront être, par exemple, un cap à suivre pour rejoindre un but ou bien un obstacle à éviter. L'utilisation de champs de neurones monodimensionnels semble tout indiquée pour encoder les différentes motivations dans l'espace pericorporel du robot (voir fig.1.8).

Par soucis de simplicité, considérons un champ de 360 neurones encodant les motivations à va-

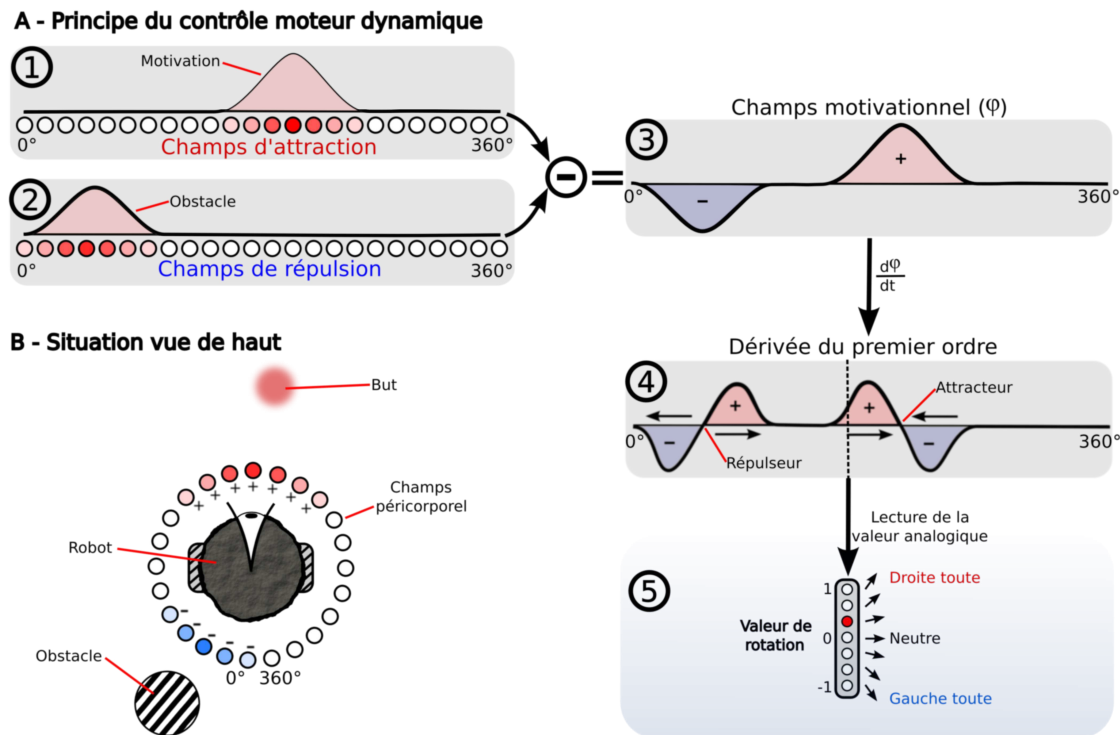


FIGURE 1.8 – Contrôle moteur utilisant les principes des champs de neurones dynamiques. **A** - L'espace pericorporel du robot est représenté par un champ d'attraction (1) et un champ de répulsion (2). Chaque neurone code pour un angle particulier. La soustraction des 2 donne un champ Φ (3) contenant les valences motivationnelles de l'agent (attraction pour les bulles positives, répulsion pour les négatives). Une dérivée du champ (4) forme des points attracteurs et répulseurs. Une activité positive entraîne une rotation à droite tandis qu'une négative entraîne une rotation à gauche. L'agent converge naturellement vers les attracteurs et diverge des répulseurs. Une lecture de valeur analogique (5) permet de transformer le champ 4 en ordres moteurs. **B** - Vue de haut schématique du robot dans son environnement. Chaque neurone du champ pericorporel code pour un angle particulier. Un but est perçu comme une bulle positive (rouge) et un obstacle par une bulle négative (bleue).

lence positive (une source de nourriture, un chemin à suivre, une personne avec qui l'agent veut interagir, etc...). Chacun des neurones de ce champ est le représentant d'une portion angulaire de 1° autour du robot, de sorte que l'ensemble du champ couvre un panorama à 360 degrés. Une bulle d'activité dans ce champ représente donc le cap à suivre vers un but, cap que le robot doit essayer de maintenir.

De la même façon, un second champ similaire (360 neurones) représente les motivations à **valence négative** (un obstacle, un danger, une zone interdite, etc...). Une bulle dans ce champ représente un cap que le robot devra éviter.

Par la soustraction de ces 2 champs, nous obtenons un 3ème champ (que l'on nommera Φ). Ce champ résume l'ensemble des motivations de l'agent, appétitives et aversives. Les activités de Φ peuvent être positives (appétitives) comme négatives (aversives). Une simple dérivée du champ Φ est réalisée au sein d'un 4ème champ. Elle permet de convertir ces potentiels de motivation en potentiels de rotation, exploitable par le robot pour produire des mouvements. Une valeur négative déclenchera une rotation à gauche tandis qu'une valeur positive déclenchera une rotation à droite. La vitesse de rotation est directement proportionnelle à l'amplitude de la motivation.

Ainsi la dérivée d'une bulle positive donne un point attracteur (en terme de rotation), et celle d'une bulle négative donne un répulseur (voir [Cuperlier et al., 2006] pour plus de détails). La dérivée de Φ est définie telle que :

$$\frac{d\Phi}{dt} = f_{att}(\Phi) + f_{rep}(\Phi) \quad (1.11)$$

$$\text{Force attractive : } f_{att}(\Phi) = -\lambda_{att} \cdot \sin(\Phi - \Phi_{att}) \quad (1.12)$$

$$\text{Force répulsive : } f_{rep}(\Phi) = \lambda_{rep} \cdot \sin(\Phi - \Phi_{rep}) \quad (1.13)$$

avec $f_{att}(\Phi)$ (respectivement $f_{rep}(\Phi)$) la force de rotation correspondant à une motivation appétitive (respectivement une motivation aversive). Et λ_{att} (respectivement λ_{rep}) un coefficient d'amplitude associé à l'appétit (respectivement l'aversion). On notera que l'éventualité d'un équilibre parfait n'est pas impossible. Si l'on imagine un obstacle séparant le robot de son but dans un alignement parfait, alors les 2 bulles d'attraction et de répulsion s'annihileraient mutuellement. En pratique, ce cas est peu probable dans un monde réel en raison du bruit résultant de l'imperfection des capteurs et effecteurs (point d'équilibre instable).

L'intérêt d'adopter un codage par champ de neurones dynamique est triple. Il permet de bénéficier des propriétés de filtrage et du compromis compétition/coopération. Le contrôle moteur devient très intuitif et correspond à la dérivée du champ à la position considérée. Enfin, le DNF prodigue une base au sein de laquelle peuvent être fusionnées les actions de stratégies de natures très différentes, pourvu que ces actions adoptent le même codage et le même référentiel. Nous utiliserons par la suite un champ de neurone dynamique de façon similaire pour représenter l'espace pericorporel du robot et ainsi fusionner de manière efficace les orientations proposées par différentes stratégies.

Malgré toutes leurs qualités indéniables, les champs de neurones dynamiques possèdent un inconvénient de taille. Ils possèdent un nombre important de paramètres qui doivent être adaptés pour chaque nouveau problème.

En d'autres termes, les champs de neurones dynamiques sont difficiles à paramétrer et peu enclins à s'adapter au changement, mais ils permettent toutefois d'obtenir une solution efficace pour la fusion/sélection de stratégie, tout en proposant un cadre directement adapté au contrôle moteur.

1.5 Discussion

Les outils et notions présentés dans ce chapitre ont été abordé(e)s de manière volontairement didactique pour rendre la lecture de cette thèse accessible au lecteur non-initié. Nous avons abordé brièvement le fonctionnement du neurone formel, lequel sera l'unité de calcul élémentaire de ses travaux. Nous avons pu voir que de telles unités sont capables d'apprendre un stimulus d'entrée par la modification de leurs poids synaptiques. Même si le fonctionnement physique reste le même, nous avons réalisé une distinction sémantique entre apprentissage par association et apprentissage par catégorisation. Ainsi, nous parlerons de groupe de catégorisation pour désigner une population de neurones entraînés à reconnaître des formes à partir de stimuli présents sur leurs entrées. Nous parlerons également de conditionnement pour désigner

un apprentissage de type conditionnement pavlovien.

Au delà de la notion d'apprentissage, nous avons pu voir le fonctionnement d'un codage en population et son intérêt pour les processus de coopération et compétition dynamiques. Nous exploiterons un tel codage au sein de champs de neurones dynamiques.

Enfin nous avons présenté les 2 plate-formes robotiques mobiles avec lesquels nous réaliserons les différentes expériences. La plate-forme d'intérieur sera la plus utilisée.

Tout ces outils seront utilisés abondamment dans ces travaux et constituent donc une base nécessaire à la compréhension de l'ensemble des chapitres qui vont suivre.

Dans le chapitre suivant, nous tenterons de situer notre approche dans la littérature existante sur les modèles de navigation robotiques.

The ability to navigate in a complex environment is one of the most challenging skills for every animal on this planet. It is crucial for survival, and is even seen as the evolutionary pressure to develop brains. The reason why plants do not have a brain is that they do not have to move.

– Wolpert - Love is a many-moleculed thing.

CHAPITRE 2

Modèles de navigation en robotique

De manière générale, on regroupe sous l'appellation *robots mobiles* l'ensemble des robots à base mobile, par opposition aux robots manipulateurs. En pratique, le terme robot mobile désigne principalement les robots munies de roues. Ceux n'utilisant pas de roue sont généralement désignés par leur type de locomotion (marcheurs, rampants, aériens ou sous-marins).

Historiquement, les robots mobiles autonomes à roues ont été étudiés dès le début des années 50 [Walter, 1953]. Leur faible complexité en fait de bons sujets pour l'étude des systèmes autonomes. Cependant, malgré leur simplicité apparente (mécanisme plan, actionneurs linéaires), ces plate-formes ont soulevé un grand nombre de problèmes difficiles, dont la plupart ne sont toujours pas résolus. Ainsi, alors que les robots manipulateurs se sont aujourd'hui généralisés dans l'industrie, rares sont les applications industrielles qui utilisent des robots pour résoudre des tâches de navigation en totale autonomie. On a vu apparaître, au début des années 2000, quelques plate-formes mobiles autonomes dans l'industrie (chariot guidés) et destinées au grand public (aspirateurs autonomes), mais l'industrialisation de systèmes plus complexes dotés de capacités d'apprentissage bute sur certains problèmes délicats. Contrairement aux robots manipulateurs dont l'espace de travail est restreint et dont les tâches restent connues et répétitives, les robots mobiles sont souvent destinés à évoluer de manière autonome dans des environnements dynamiques, inconnus et peu ou pas structurés (les contrastes et la luminosité peuvent varier fortement, certains objets sont mobiles et enfin la géométrie du monde physique n'est pas connue a priori et peut être complexe).

Savoir se localiser et se déplacer de façon cohérente au sein d'un environnement à priori totalement inconnu fait partie des compétences essentielles pour parvenir à naviguer de manière autonome. Selon le point de vue "*classique*" en robotique, le principal défi provient du fait qu'il soit complexe de construire un modèle interne du monde suffisamment précis pour permettre au système d'opérer en temps réel de façon cohérente. L'un des problèmes majeurs bien connu des roboticiens sous le nom de **SLAM** (Simultaneous Localization and Mapping) [Smith et al., 1987; Leonard and Durrant-Whyte, 1991] est qu'il semble nécessaire de posséder une carte de l'environnement afin de pouvoir se localiser alors même que cette carte doit être construite au fur et à mesure des déplacements. Sans carte, pas d'estimation de position possible. Sans estimation de position, pas de construction de carte possible. Le paradoxe est similaire à celui de l'oeuf et de la poule. Nous aurons l'occasion de voir que la construction d'une carte n'est pas l'unique solution, ni la meilleure, pour résoudre une grande part des problèmes de navigation autonome.

Dans ce chapitre nous aurons l'occasion d'appréhender un état de l'art non-exhaustif des stratégies de navigation existantes. Les premières solutions envisagées pour résoudre ce problème de manière générale nous viennent de la navigation maritime. Nous ferons donc un bref parallèle avec ces méthodes historiques et nous noterons que la plupart d'entre elles restent valables et très utilisées en robotique mobile encore de nos jours. Nous passerons ensuite en revue les méthodes de navigation dites "*classiques*" et leur limitation face à des environnements inconnus et dont les contraintes ne sont pas maîtrisées. Nous verrons comment les animaux sont parvenus à proposer des solutions simples et robustes en abordant le problème sous un autre angle. Le maintien d'invariants sensorimoteurs et la navigation topologique sont des méthodes qui permettent de s'affranchir de manière élégante de problèmes qui semblent difficiles à résoudre vus sous le prisme des méthodes classiques. Nous nous attarderons finalement sur les modèles computationnels existants de la navigation animale. Notre démarche fait partie intégrante de cette approche **bio-inspirée** de la navigation.

2.1 Navigation nautique

Au 15^{ème} siècle, la carte du monde est encore en grande partie muette¹. Les navigateurs se lancent souvent dans l'inconnu tout en souhaitant à la fois reporter d'éventuelles découvertes sur une carte mais également parvenir à revenir à bon port en fin de voyage. Pour cela, ils devaient estimer leur position au cours de l'excursion. A l'époque, la navigation astronomique est encore peu fiable par manque d'instruments précis (une erreur de mesure de seulement 1 degré induit une différence de position d'environ 100km en Europe). Les navigateurs ont alors eut à résoudre un problème de localisation et de cartographie simultanée, problème encore d'actualité en robotique aujourd'hui (SLAM).

Selon Gallistel, la navigation est un processus par lequel on détermine et maintient une trajectoire vers un but [Gallistel and Gelman, 1990]. Selon Levitt et Lawton [Levitt and Lawton, 1990], ce processus se décompose en 3 étapes, ou plutôt 3 questions auxquelles répondre : "Où suis-je ?", "Où se situent les autres lieux vis à vis de moi ?" et "Comment faire pour me rendre dans tel lieu à partir de ma position ?".

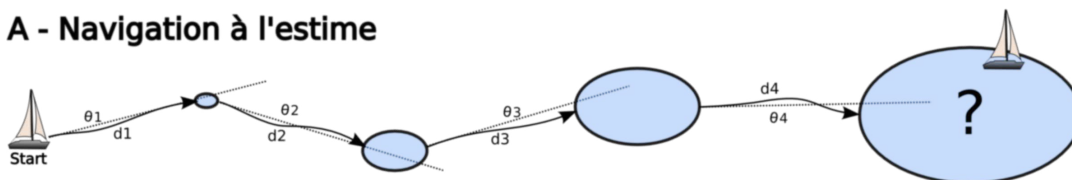
2.1.1 Navigation côtière

Dans le cas le plus simple, le vaisseau navigue en bordure de côte, ce qui lui permet de se localiser assez précisément grâce aux indices visuels environnants (voir fig. 2.1 - B). La base de la navigation "observée" est le point par triangulation ou trilatération. Les marins utilisent les points remarquables **invariants** de l'environnement pour pouvoir trianguler leur position, après quoi ils peuvent estimer le cap à suivre vers un but à l'aide d'une carte de cet environnement. Ces points invariants sont nommés des "**amers**" et l'angle sous lequel ils sont perçus correspond à leur "**azimut**". La méthode d'estimation par triangulation [Haasbroek, 1968] consiste à mesurer les azimuts de 3 amers. On reporte ensuite l'azimut depuis chacune des positions des amers correspondants. L'intersection des 3 droites obtenues définit un triangle, plus ou moins grand en fonction de la précision des relevés. La position du navire est estimée à l'intérieur de ce triangle. La trilatération est une méthode similaire pour laquelle la position est estimée en fonction de la distance aux amers en lieu et place de leur azimut. Dans la cadre de la navigation visuelle

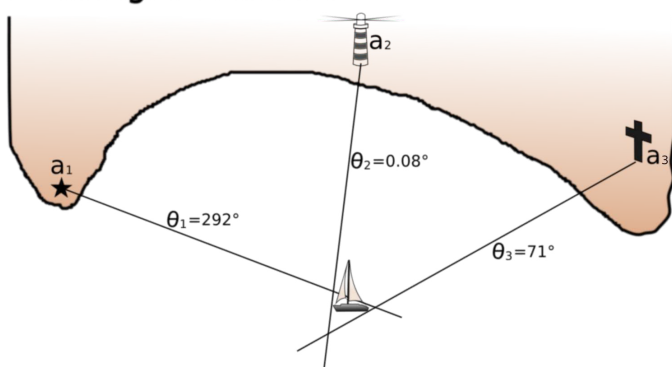
1. <http://expositions.bnf.fr/lamer/bornes/feuilletoirs/conquete/32.htm>

monoculaire, on préférera souvent une mesure par triangulation, puisque la mesure de l'azimut est alors plus précise qu'une estimation de distance aux amers (qui ne peut s'estimer visuellement que par la hauteur apparente des amers lorsque les distances sont importantes).

A - Navigation à l'estime



B - Navigation côtière



C - Navigation astronomique

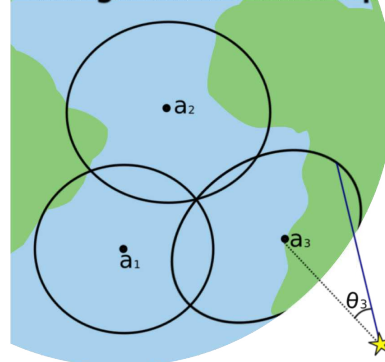


FIGURE 2.1 – Principe de la navigation maritime. **A** - Sans repère visuel, la navigation est réalisée à l'estime. La vitesse et le cap permettent d'estimer sa position, mais cette estimation est soumise au cumul d'erreurs. A chaque nouvelle mesure, les incertitudes s'accumulent, ce qui limite ce type de navigation à des trajets courts. **B** - La richesse des amers visuels aux abords d'une côte permet une localisation précise par triangulation. La mesure de l'azimut (l'angle) perçu de 3 amers permet de déterminer un triangle virtuel au sein duquel se trouve la position réelle. **C** - Au large des côtes, les seuls amers disponibles sont les astres. La connaissance de la position terrestre pour laquelle un astre donné est au zénith permet de se localiser : le relevé de son azimut permet de définir un cercle de positions possibles. La mesure de l'azimut de 3 astres permet de déterminer la position exacte par triangulation (à l'intersection des 3 cercles).

2.1.2 Navigation à l'estime

Lorsque les explorateurs quittèrent la côte pour s'aventurer en pleine mer, la méthode de localisation par triangulation d'amers devenait impossible par manque d'indices visuels. Ils avaient alors recours à la méthode de navigation à l'estime (voir fig. 2.1 - A). Celle-ci consiste à déduire la position du navire à partir de son cap et de la distance parcourue depuis sa dernière position connue. Le trajet est donc estimé par l'intégrale des vecteurs de mouvements élémentaires au cours du temps. Ces vecteurs sont déduits, à chaque point de mesure, par la vitesse et le cap du navire. Le problème inhérent à ce type de méthode récursive est son cumul d'erreurs progressif. Même si la méthode est très précise localement (sur de petites distances), elle intègre les petites erreurs de mesure et devient vite problématique. En effet, en plus d'une dérive propre au bateau et aux courants marins, chaque nouvelle estimation de position introduit une erreur de mesure du cap due à la précision angulaire ainsi qu'une erreur de mesure de la vitesse. Ces limitations rendaient la méthode inefficace lors de voyages sur de longues distances. L'amélioration de la

précision des instruments de mesure à permis de ralentir ce cumul d'erreur mais sans repères externes fixes l'utilisation de la navigation à l'estime restait limitée.

2.1.3 Navigation astronomique

Pour parvenir à se localiser en pleine mer sans accumulation d'erreur, les explorateurs ont dû prendre en compte les seules informations visuelles disponibles loin des côtes : les étoiles (voir fig. 2.1 - C). Avec l'invention d'outils de mesure beaucoup plus précis, la navigation astronomique est devenue exploitable (deuxième moitié du 19ème siècle). Celle-ci consiste à déterminer sa position à l'aide de l'observation des astres et la mesure de leur azimut (c'est-à-dire l'angle entre l'astre et l'horizon) [Jean, 1990]. La connaissance de la position terrestre pour laquelle un astre donné est au zénith permet de se localiser : le relevé de son azimut permet de définir un cercle représentant l'ensemble des positions terrestres pour lesquels l'astre en question est perçu sous cet azimut. La mesure de l'azimut de 3 astres distincts permet de déterminer la position exacte par triangulation. L'intersection des 3 cercles se trouve être la seule position garantissant les 3 observations. Cette méthode est appelée méthode par interception².

Finalement la méthode de navigation la plus fiable consiste à cumuler navigation à l'estime et navigation observée. La navigation à l'estime fournit une localisation plus précise mais est soumise au cumul d'erreurs sur de longue distance. La position devait donc subir une **recalibration** périodique par le biais d'une estimation de position sur la base d'indices astronomiques afin de limiter cette dérive. Nous allons voir que ces principes sont toujours utilisés en robotique mobile aujourd'hui.

2.2 Navigation en robotique classique

Selon le point de vue de l'approche classique de la navigation robotique, la construction d'une carte précise de l'environnement est nécessaire pour qu'un agent parvienne à planifier ses actions et puisse naviguer de manière autonome dans cet environnement. Selon Newman, cette capacité est considérée comme primordiale pour la durée de vie d'agents mobiles autonomes [Leonard and Newman, 2003] La cartographie d'un environnement inconnu est généralement considérée comme l'un des problèmes les plus importants pour atteindre une réelle autonomie en robotique mobile [Thrun, 2002b; Leonard and Newman, 2003].

Les premiers modèles internes de l'environnement ont été proposés par [Nilsson, 1984] puis [Moravec and Elfes, 1985] dans les années 80. Le modèle est basé sur une grille représentant l'espace libre et occupé par les obstacles (occupancy grid). Une méthode alternative basée sur une représentation métrique à ensuite été proposée par Chatila et Laumond [Chatila and Laumond, 1985]. D'autres approches proposent une cartographie topologique (décrivant les connexions entre lieux) [Kuipers and Byun, 1991; Mataric, 1992]. A partir des années 90, le domaine est dominé par l'approche probabiliste, dont les travaux de Smith, Self, et Cheeseman [Smith et al., 1987] introduisent un cadre statistique pour répondre au problème du SLAM (c'est à dire, parvenir à cartographier l'environnement tout en s'y localisant dans le même temps). Ces approches probabilistes sont principalement représentées par 2 grandes familles : l'une d'elle utilise la méthode du *filtrage de Kalman* [Kalman, 1960], tandis que l'autre se base sur l'algorithme *Expectation Maximization* [Dempster et al., 1977] pour réaliser ces estimations.

2. http://en.wikipedia.org/wiki/Celestial_navigation

Le filtre de Kalman permet d'estimer les états d'un système linéaire dynamique à partir d'une série de mesures incomplètes ou bruitées. Il est utilisé dans une large gamme de domaines technologiques (radar, vision électronique, communication ...). Ce filtre fait l'hypothèse que le bruit est purement gaussien, ce qui est une approximation insuffisante dans de nombreux cas. De plus, il ne permet de traiter que des problèmes impliquant des dépendances linéaires entre les différentes variables [Thrun, 2002b]. Une variante, le filtrage de Kalman étendu permet de traiter les problèmes non-linéaires [Julier and Uhlmann, 2004].

Le filtrage particulaire est une méthode permettant de ne pas se limiter à de telles hypothèses de bruit gaussien et de linéarité, mais cette technique présente un coût computationnel important [Gordon et al., 1993]. Cette technique est utilisée pour la localisation de robot mobile par Thrun et al. [Thrun, 2002a]. De nombreuses variantes de ces méthodes existent, mais leur description et comparaison est en dehors du champ de cet état de l'art.

Ces méthodes probabilistes tentent de résoudre 5 problèmes majeurs liés à la cartographie autonome [Thrun, 2002b] : (1) les bruits de mesures ne sont pas statistiquement indépendants. La nature de ces bruits dépend fortement du contexte. Par exemple, l'utilisation d'une caméra panoramique non calibrée crée une erreur de localisation des amers dépendante de l'orientation. S'affranchir d'un bruit non-centré (biaisé) s'avère extrêmement compliqué. De même, pour une navigation à l'estime, le cumul d'erreurs affecte la manière dont les futures mesures seront interprétées (voir fig. 2.2 - a). (2) La dimensionnalité de la représentation augmente considérablement avec la taille de l'environnement et la précision du modèle, ce qui implique une quantité de mémoire importante. (3) Le problème le plus dur à résoudre est celui de la correspondance qui survient lors de la clôture d'une boucle au sein d'une trajectoire (*correspondence* ou *loop closing problem*) [Williams et al., 2009]. La difficulté réside dans le fait de déterminer si plusieurs mesures, enregistrées à des moments différents, correspondent à la même entité physique dans l'environnement. Ce phénomène provient de la nature ambiguë de la perception, phénomène bien connu en psychologie et expliqué par la théorie de la "*Gestalt*" [Smith, 1988b]. (4) L'aspect dynamique du monde réel peut rendre inopérant le système si des éléments essentiels à la localisation viennent à disparaître. Il est alors nécessaire d'effectuer un réapprentissage pour s'adapter à ces changements. Enfin (5), le système doit pouvoir planifier son chemin au sein d'un environnement partiellement connu. Les solutions doivent bien souvent faire face à un compromis entre réactivité et fiabilité. L'utilisation de méthodes statistiques [Thrun, 2002b] permet de s'affranchir du problème de la navigation en environnement inconnu mais implique de nombreux passages dans chaque lieu, ce qui reste incompatible avec le besoin d'une grande réactivité au stade initial.

En restant dans le cadre de l'approche classique de la robotique, les méthodes basées sur la construction de modèles internes du monde **sans a priori** sont finalement celles qui donnent les meilleurs résultats. Assez récemment une telle méthode a permis des performances impressionnantes de navigation en intérieur sur le long terme (plusieurs semaines) avec changements environnementaux [Andreasson et al., 2005]. Cependant, le passage à l'échelle à des environnements plus larges et moins structurés (extérieurs) pose encore de nombreux problèmes tels que l'explosion de la taille mémoire du modèle interne, l'odométrie non fiables sur terrains accidentés ou encore le recollement de cartes.

D'autres méthodes plus ad-hocs donnent de bons résultats en réponse à des problèmes spécifiques. Les travaux de Leonard et Newman [Leonard and Newman, 2003] montrent, par exemple, qu'il est possible de développer un algorithme dont le temps de convergence est constant

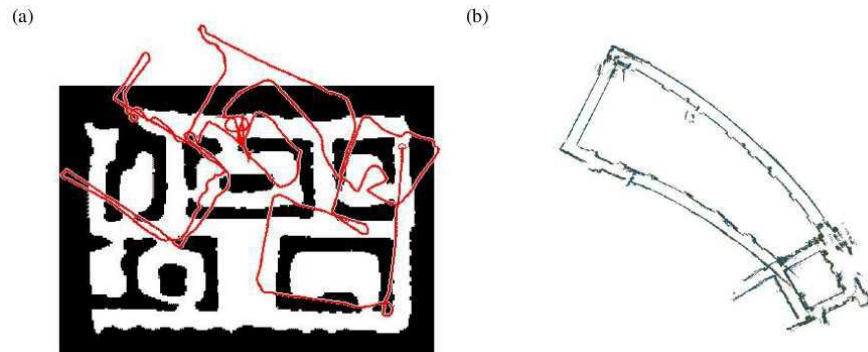


FIGURE 2.2 – Figure extraite de [Thrun, 2002b]. **a** - Exemple de trajectoire obtenue à l'aide d'information odométrique uniquement. La figure montre l'impact du cumul d'erreur sur la trajectoire. De petites erreurs peuvent avoir un effet important sur de grandes distances. **b** - Exemple de cartographie d'un environnement intérieur cyclique. La figure montre le problème de correspondance qui survient lorsque le robot achève une boucle au cours de sa trajectoire. Parvenir à établir cette correspondance est un des plus importants défis dans le domaine du SLAM.

dans le cas où le robot est autorisé à faire des visites répétées de toutes les régions de l'environnement. Steux et Hamzaoui [Steux and Hamzaoui, 2010] sont parvenus à développer un algorithme léger et très simplifié de SLAM dont le code ne dépasse pas 200 lignes. Cet algorithme permet au robot d'atteindre une vitesse de 2.5m/s en embarquant l'ensemble des calculs sur un processeur ARM9. Nguyen et al [Nguyen et al., 2006] propose un modèle également très léger basé sur l'hypothèse d'un environnement formé de plans orthogonaux (ce qui est souvent le cas en intérieur) permettant ainsi de simplifier considérablement la complexité des calculs.

Bien que très utilisées dans certains domaines précis, ces méthodes requièrent la plupart du temps un scanner laser précis et donc très coûteux. De plus, elles sont fortement dépendantes du modèle a priori du monde. Des travaux récents visent justement à s'adapter aux déplacements et à l'occlusion d'amers visuels et à se remettre à jour dynamiquement [Lee et al., 2006]. D'autres méthodes donnent de bons résultats dans des environnements fortement dynamiques comme l'algorithme développé par Foka et Trahanias [Foka and Trahanias, 2010]. L'algorithme permet à un robot mobile de se localiser et naviguer dans un milieu encombré (une foule d'êtres humains par exemple). Des approches ensemblistes sont également utilisées pour la robotique sous-marine [Jaulin, 2009, 2011] pour s'affranchir des conditions données par l'environnement (Les fonds marins sableux sont des lieux très pauvres en amers et présentant de fortes redondances).

Malgré de nombreux progrès réalisés dans ce domaine, le SLAM présente tout de même des défis importants. Aujourd'hui, il existe de nombreuses méthodes pour cartographier et se localiser simultanément dans un environnement si celui-ci est statique, structuré et de taille limitée. La navigation au sein d'environnements complexes, dynamiques, non-structurés et de grandes tailles restent encore un problème de recherche largement ouvert [Thrun, 2002b].

2.3 Navigation animale

La plupart des animaux possèdent des capacités à trouver leur chemin sans l'aide d'aucune carte ou d'instruments particuliers. Les sternes arctiques (espèce d'oiseaux), les papillons monarques ou encore les saumons migrent régulièrement sur des milliers de kilomètres et parvi-

ennent à trouver leur site de reproduction [Dingle Hugh and Drake, 2007]. Selon Wolpert, la capacité à naviguer dans un environnement complexe est l'une des compétences les plus difficiles à réaliser pour les animaux : cette compétence est cruciale pour leur survie et à certainement été un moteur pour le développement du cerveau au cours de l'évolution [Wolpert, 2002]. Cette théorie ne vaut cependant que pour certains animaux, comme les mammifères, et pour des tâches de navigation sophistiquées, comme la reconnaissance de lieux et la planification de trajectoire. Nous allons voir que la plupart des insectes sont tout à fait capables d'exhiber certains comportements de navigation simples et robustes sans besoin d'un système nerveux complexe [Srinivasan and Zhang, 2003; Wystrach et al., 2013].

De nombreux travaux en éthologie se sont intéressés à l'étude des stratégies utilisées dans la nature pour pouvoir naviguer. Von Frisch a montré que les abeilles peuvent naviguer à l'aide du soleil, de la polarisation de la lumière ou du champ magnétique terrestre, bien qu'elles préfèrent s'en référer au soleil lorsque celui-ci est disponible [Von Frisch, 1955]. Keeton a montré que les pigeons voyageurs pourraient même faire usage d'un éventail d'indices aidants à la navigation, comprenant le soleil, le champ magnétique terrestre, l'olfaction et la vision [Keeton, 1971]. Lockley a démontré qu'une espèce de petit oiseau marin, le Puffin des Anglais, peut s'orienter et revenir à son nid très rapidement à condition que le soleil ou les étoiles soient visibles [D. Lack, 1938].

La caractéristique importante à noter ici est qu'une partie des solutions utilisées chez les animaux ne semblent pas recourir à une carte de l'environnement ni à l'utilisation de connaissances a priori. Bien souvent, le maintien d'**invariants** dans les perceptions suffit à faire émerger un comportement cohérent et robuste. Les animaux semblent prendre des décisions en se basant sur une mesure de corrélation entre l'information actuelle et l'information mémorisée. Leur développement semble basé sur des réflexes et conditionnements relativement simples [Piaget, 1936]. De plus, il semble que l'information extraite et mémorisée se limite à l'information utile pour la tâche. Par exemple, dans le cadre d'une tâche de navigation vers un but, la reconnaissance d'un objet est codée de telle manière que celui-ci puisse être utile pour rejoindre le but [Gaussier and Zrehen, 1995]. Ce point de vue, adopté par de nombreuses espèces, permet à l'inverse des techniques d'intelligence artificielle classiques, de résoudre des tâches complexes avec la seule aide d'un cerveau minimaliste.

Les insectes sont des exemples intéressants pour illustrer ce principe. Certaines espèces de fourmis n'utilisent que l'information visuelle directement accessible sur un lieu important afin d'apprendre à retrouver ce lieu [Wehner and Rüber, 1979]. Une expérience permet de montrer ce phénomène : un bâton décoré de 2 anneaux noirs est placé à la sortie du nid (voir fig. 2.3 - A et B). La fourmi apprend la position de son nid en mémorisant l'angle sous lequel elle perçoit le motif du bâton. Lorsque ce dernier est remplacé par un bâton 2 fois plus haut, la fourmi se met à chercher son nid à la position pour laquelle les motifs du bâton sont perçus sous le même angle que ceux appris. De la même manière, les guêpes parviennent à retrouver leur nid en apprenant sa position relative aux amers environnants. Tinbergen a réalisé une expérience qui met en évidence cette méthode [Tinbergen, 1969]. La guêpe apprend à reconnaître la position de son nid vis à vis de pommes de pins disposées autour de celui-ci. Elle tourne systématiquement autour de ces pommes de pins pendant plusieurs secondes avant de s'éloigner. Lorsqu'elle revient, les pommes de pins ont été volontairement déplacées de quelques mètres du nid. La guêpe tente alors de retrouver son nid au centre des pommes de pins (voir fig. 2.3 - C et D).

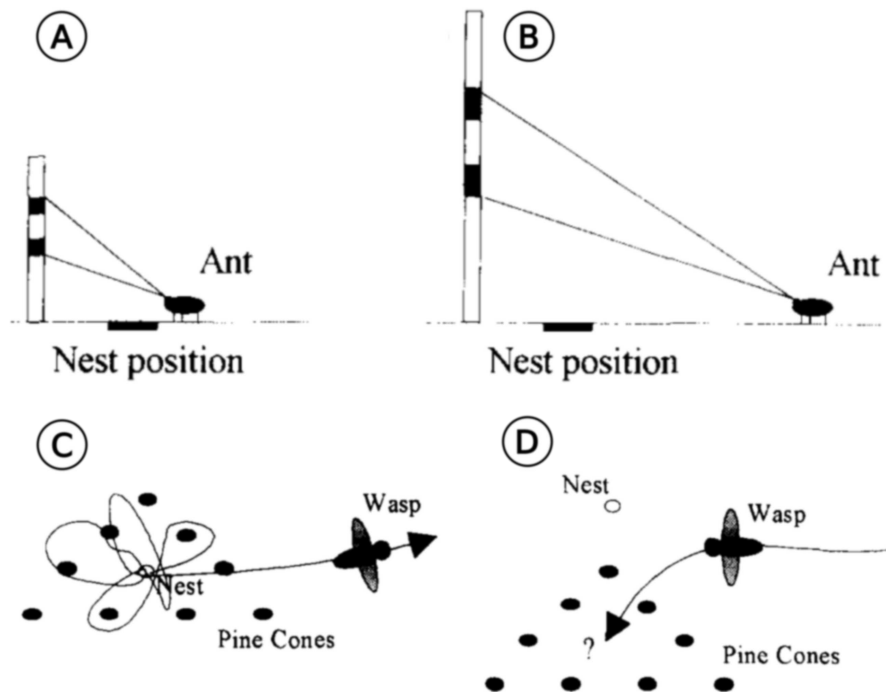


FIGURE 2.3 – Figure extraite de [Gaussier and Zrehen, 1995]. **A** - La fourmi apprend la position de son nid en mémorisant l'aspect visuel d'un bâton. **B** - Si le bâton est remplacé par un objet similaire 2 fois plus grand, la fourmi cherche son nid plus loin et ne parvient pas à le retrouver [Wehner and Rüber, 1979]. **C** - Avant de s'éloigner de son nid, la guêpe tourne autour de pommes de pins, disposées tout autour du nid, afin d'apprendre la position de celui-ci. **D** - Si la position des pommes de pins est déplacée (translation) alors la guêpe ne parvient pas à retrouver son nid [Tinbergen, 1969].

2.3.1 Intégration de chemin

De nombreux animaux ont la capacité d'évaluer en permanence la distance les séparant de leur nid ainsi que l'orientation à prendre pour le rejoindre, comme si un mécanisme de mémorisation du chemin parcouru leur permettait l'accès à tout moment au vecteur de retour au nid (homing vector) où bien à une position ultérieure [Etienne and Jeffery, 2004]. Pour certains animaux, ce vecteur permet un retour direct au nid même après plusieurs heures d'exploration de l'environnement. En 1873, Darwin citait un explorateur de la Sibérie du Nord, Von Wrangell, qui avait observé que les natifs de cette contrée étaient capables de garder un cap en direction d'un point particulier sur de longues distances, malgré des changements de direction incessants et la totale absence de repères célestes ou terrestres [Darwin, 1873]. Le point crucial, relevé par Darwin, est qu'aucun des repères directionnels lointains, comme les repères célestes, ou les autres indices purement directionnels comme le champ géomagnétique, ne peut diriger le déplacement vers un point particulier de l'espace. Ces indices fournissent, au mieux, la direction d'un point unique de l'espace. La détermination d'une position ne peut résulter que d'un processus d'intégration de chemin : une navigation à l'estime (dead reckoning).

Là encore, la fourmi est un bon exemple pour illustrer un tel processus. Cette dernière quitte généralement le nid à la recherche de nourriture en ayant principalement recours à une stratégie d'exploration aléatoire [Müller and Wehner, 1988; Wittlinger et al., 2006; Grah et al., 2005].

Lorsqu'elle tombe sur une source de nourriture, il a été observé qu'au lieu de parcourir la trajectoire en sens inverse, la fourmi est capable de rejoindre son nid en ligne droite en empruntant le chemin le plus court. Les fourmis forestières pourraient utiliser des indices visuels ou olfactifs pour retrouver le chemin du nid. Les fourmis du désert, quant à elles, n'ont pas accès à de tels indices visuels et les vents puissants empêchent l'utilisation d'indices olfactifs. Ces fourmis ont recours à un mécanisme d'intégration de chemin basé sur des informations idiothétiques : elles comptent leurs pas (voir fig. 2.4). La fourmi du désert "cataglyphis fortis" est ainsi capable de retourner à son nid après une exploration sur plusieurs centaines de mètres [Wehner and Srinivasan, 1981].



FIGURE 2.4 – Mécanisme d'intégration de chemin chez la fourmi. Figures extraites de [Müller and Wehner, 1988; Wittlinger et al., 2006; Grah et al., 2005]. **A** - Trajectoire d'une fourmi s'éloignant de son nid jusqu'à tomber sur une source de nourriture (F). Malgré une trajectoire complexe et hasardeuse, la fourmi est ensuite capable de retourner à son nid en ligne droite. **B** - Lorsque la taille des pattes de la fourmi est modifiée (prolongement ou amputation) après qu'elle ait trouvé la source de nourriture, cette dernière sur-estime ou sous-estime la distance au nid. **C** - Illustration d'une fourmi dont les pattes ont été allongées artificiellement.

Une fois la source de nourriture découverte par la fourmi, les chercheurs procèdent à une modification de la taille de ses pattes : un allongement artificiel par le biais d'échasses, ou bien un raccourcissement par amputation. Lors du retour au nid, les fourmis amputées sous-estiment la distance au nid, tandis que celles dotées d'échasses la sur-estiment.

Les insectes volants comme l'abeille sont également capables d'intégrer leur chemin en utilisant principalement le flot optique, l'azimut du soleil ou le champ magnétique du ciel [Srinivasan and Zhang, 2003; Wehner, 1994]. Même si ce mécanisme a été principalement étudié chez les rats et les insectes, il semble aussi faire partie intégrante de la navigation spatiale chez l'humain [Maguire et al., 1998]. Les chiens, les hamsters et les humains utilisent principalement leur proprioception [Seguinot et al., 1998; Loomis et al., 1993]. Le choix de la modalité sensorielle varie d'une espèce à l'autre et dépend également de la confiance en cette stratégie dans une situation donnée [Jeffery, 1998] (voir fig. 2.5).

2.3.2 Région hippocampique et cellules de lieu

L'hippocampe est une structure majeure du cerveau des mammifères appartenant au système limbique (voir fig. 2.6). Il joue un rôle très important dans la navigation spatiale, la mémorisation à court terme et la consolidation de la mémoire à long terme. Chez l'Homme et le primate, il se situe dans le lobe temporal médian, sous la surface du cortex avec lequel il est en étroite collaboration. Il se compose de trois sous-structures : le subiculum, la corne d'Ammon (composée des

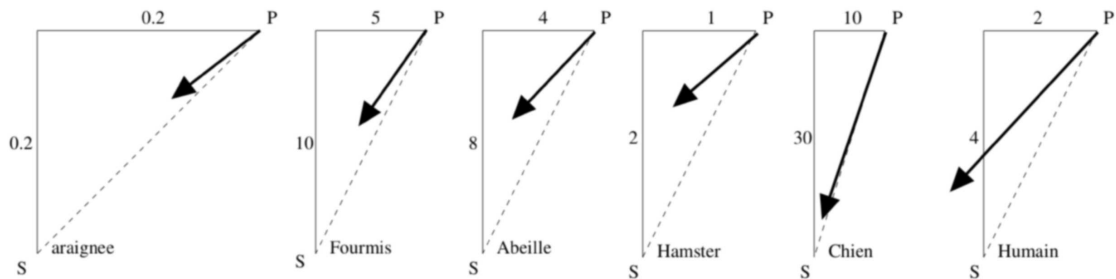


FIGURE 2.5 – Figure extraite et modifiée de [Etienne and Jeffery, 2004]. Comparaison inter-espèce de l'erreur d'intégration de chemin au cours d'une expérience de privation d'information allothétique. Les sujets effectuent un parcours en forme de "L", de (S) à (P) en étant privés de stimuli externes. Une fois en (P), ils tentent de retourner au point de départ par le chemin qu'ils estiment le plus court. Araignées, hamsters et humains sont privés des sens visuel, auditif, olfactif et tactil. Fourmis et abeilles sont privées de références comme le soleil. Les chiens sont privés de vision et d'audition à l'aller mais pas lors du trajet retour. Le trait en pointillés désigne le vecteur de retour réel, tandis que la flèche représente le vecteur emprunté par le sujet.

aires CA1, CA2 et CA3) et le gyrus denté. Le système limbique, quant à lui, est un groupement de structures contrôlant les émotions primaires comme la colère, la peur ou le plaisir, la modulation du comportement et la formation de la mémoire [Broca, 1871].

L'hippocampe ferait office de mémoire à court terme et participerait à la consolidation de souvenirs récents dans le cortex, siège de la mémoire à long terme. Au cours de cette consolidation, l'hippocampe se désengage progressivement de la mémorisation et récupération des souvenirs. Le caractère progressif des amnésies rétrogrades a notamment été observé par Scoville et Milner [Scoville and Milner, 2000].

Chez le rat, la région hippocampique a été identifiée comme responsable de la mémoire spatiale [O'Keefe and Dostrovsky, 1971].

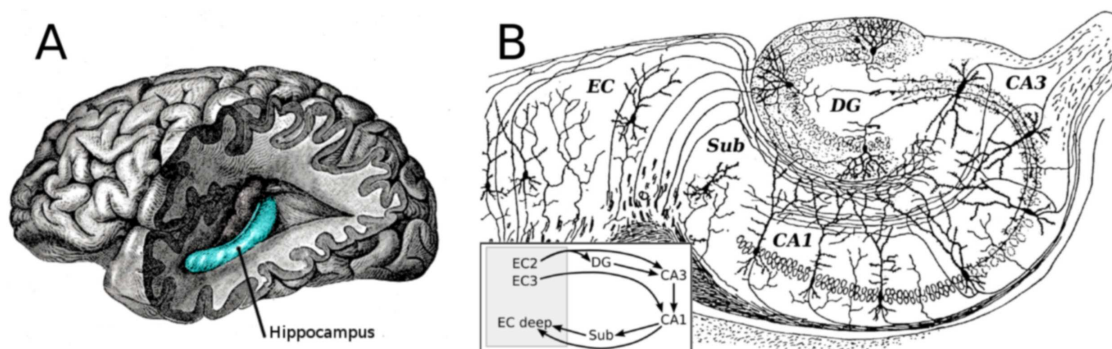


FIGURE 2.6 – **A** - L'hippocampe est situé dans le lobe temporal médian. C'est une structure paire, présente de manière symétrique dans chaque hémisphère. **B** - Représentation d'une coupe transversale de la région hippocampique du rat ainsi que des principales structures et connexions excitatrices les liants. On retrouve notamment le gyrus denté (DG) et la corne d'Ammon (CA1-3) qui forme l'hippocampe proprement dit. On trouve également la structure para-hippocampique constituée notamment du cortex entorhinal (EC) et du subiculum (Sub).

De nombreuses cellules situées dans les régions CA1 et CA3 de l'hippocampe présentent des activités, en terme de fréquence de décharge, correspondant à des positions spatiales précises de l'animal dans l'environnement [O'Keefe and Dostrovsky, 1971]. Ces neurones sont communé-

ment appelés "cellules de lieux" puisqu'ils répondent pour des endroits précis de l'espace. Leur activité est caractérisée par une réponse maximale en un point de l'environnement. Cette projection de la réponse neuronale sur l'environnement est appelée "champ de lieu" de la cellule [Rédei, 2008]. Le protocole standard d'enregistrement de ces cellules consiste à implanter des électrodes dans le cerveau d'un rat puis de le placer dans un enclos fermé. La position du rat et la fréquence moyenne de décharge des cellules de lieu sont enregistrées simultanément, après quoi il est possible de tracer l'activité d'une cellule donnée superposée à la trajectoire obtenue. Les champs de lieu se recouvrent entre cellules codant pour des lieux proches. La taille des champs semble principalement dépendre de la taille de l'enclos. Les activités des cellules apparaissent quelques minutes après la première entrée du rat dans l'enclos et peuvent parfois mettre plusieurs semaines pour converger vers une représentation stable [Wilson and McNaughton, 1993]. Après que les champs aient convergés vers une activité invariante, ils restent stables durant des semaines, voire des mois [Thompson and Best, 1990].

On sait également que l'hippocampe possède, avec le septum, un rôle important pour la détection de la nouveauté et la neuromodulation pour permettre l'apprentissage de nouveaux motifs [Hasselmo and Fehlau, 2001]. L'acétylcholine sécrétée dans le septum permettrait d'inhiber l'activation de motifs anciens afin de faciliter l'acquisition des nouveaux [Hasselmo and Schnell, 1994].

Dans la suite de ces travaux nous adopterons la représentation de la fig. 2.7-B pour parler de la réponse (en rouge) d'une cellule unique superposée à la trajectoire (en noir) de l'agent. D'autres part, nous préférons la représentation de la fig. 2.7-C pour présenter les zones pour lesquels différentes cellules impliquées sont gagnantes (une couleur différente pour chaque cellule).

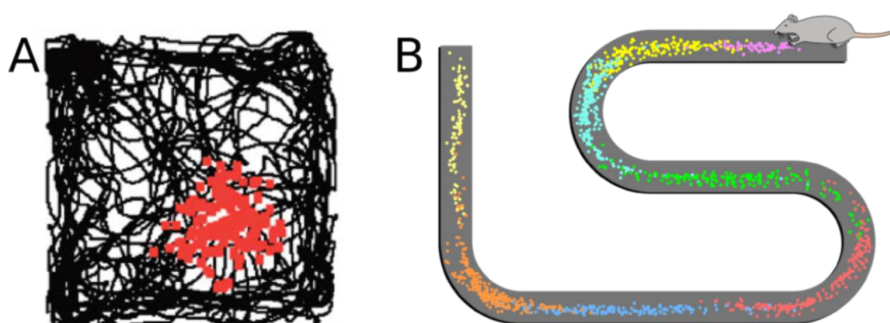


FIGURE 2.7 – **A** - Figure extraite de [Bird and Burgess, 2008], champ de lieu d'une cellule hippocampique dans un environnement bi-dimensionnel. Des électrodes, implantée dans le cerveau d'un rat, permettent d'enregistrer l'activité des cellules de lieu pendant que celui-ci explore un enclos de 1m^2 . L'activité (en rouge) est superposée à la trajectoire (en noir). **B** - Figure issue d'internet sous licence libre. Champs de lieu de plusieurs cellules hippocampiques dans un environnement mono-dimensionnel. Chaque couleur correspond à l'activité d'une cellule de lieu différente. On observe des recouvrements entre champs de cellules proches.

Bien que les signaux qui permettent d'activer ces cellules soient encore difficiles à caractériser, il apparaît que leur activité est fortement liée aux indices visuels présents dans le milieu. La vision joue un rôle primordial dans l'activation des cellules de lieu puisqu'une rotation de ces points d'intérêt en bordure de l'environnement entraîne une rotation des champs de lieu associés. De même, les champs sont homothétiques de la taille de l'environnement [Muller and Kubie, 1987]. Cependant, d'autres modalités peuvent venir compléter le code spatial. De nom-

breuses expériences montrent, par exemple, le rôle du système vestibulaire et des mécanismes d'intégration de chemin [Foster et al., 1989; Kelemen and Fenton, 2010], des informations olfactives [Zhang and Manahan-Vaughan, 2013], auditives [O'Keefe and Nadel, 1978] et tactiles [Save et al., 1998] dans la construction de ce code. Des études montrent également que les cellules de lieu répondent lorsque l'animal est plongé dans le noir complet et ne peut plus utiliser de repères visuels [Quirk et al., 1990] bien que la précision de la réponse se dégrade au fil du temps. Ces travaux montrent l'importance de certaines informations idiothétiques dans l'activation des cellules de lieu.

Certaines études suggèrent que cette intégration d'information multimodale aurait lieu en amont de l'hippocampe, dans le cortex entorhinal (EC), qui constitue la principale entrée de l'hippocampe [Gothard et al., 2001]. Une étude montre qu'en cas de conflits entre modalités idiothétique et visuelle, les champs de lieu peuvent devenir instables [Knierim et al., 1998]. Des activités de cellules de lieu ont également été observées dans d'autres régions cérébrales proches de l'hippocampe, notamment dans le cortex entorhinal. Ces cellules de lieu ont la particularité d'exhiber des champs beaucoup plus diffus [Quirk et al., 1992; Sharp, 1999b]. Le gyrus denté (DG) constitue une voie isolée de communication entre EC et le reste de l'hippocampe. D'autres structures comme le subiculum, le cortex retrosplénial, le thalamus, le striatum ou encore le cortex entorhinal jouent également un rôle très important dans la localisation de l'animal puisqu'elles hébergent des cellules de direction de la tête (head direction cells). Ces cellules codent chacune pour une certaine orientation de la tête, agissant de ce fait comme une boussole relative [Sargolini et al., 2006].

Récemment, d'autres cellules particulières ont été découvertes lors d'enregistrement dans le cortex entorhinal dorso-caudal médian (dMEC). Ces cellules, nommées "cellules de grille" (grid cells), ont la particularité de présenter des champs de lieu étroits et multiples. Leur réponse est périodique dans l'environnement, formant ainsi un pavage régulier de l'espace, à l'instar d'une grille [Hafting et al., 2005]. Les cellules de grille pourraient être à l'origine d'un système de localisation spatial basé majoritairement sur des informations idiothétiques. Nous reviendrons plus en détail sur leur fonctionnement dans le chapitre 4.

La découverte des cellules de lieu a stimulé la compréhension du fonctionnement de la navigation spatiale chez l'animal. Ces cellules constituent alors la première preuve de l'implication de l'hippocampe dans des tâches de navigation et a réactualisé l'idée de l'utilisation de "cartes cognitives" [O'Keefe and Nadel, 1978]. La notion de carte cognitive correspondrait à une représentation interne de l'environnement permettant à l'agent d'inférer son chemin parmi plusieurs. Elle se base sur l'apprentissage de la topologie existante entre différents états. Ainsi l'animal pourrait remettre en cause la planification de son chemin lors de l'apparition d'un raccourci ou encore sélectionner le chemin optimal vers un but particulier. Le concept de carte cognitive avait déjà été proposée par Tolman 40 ans plus tôt [Tolman and Honzik, 1930].

La navigation vers un but unique n'implique pas nécessairement la construction d'une carte topologique de l'environnement. Dans ce cas, il peut suffire d'utiliser une stratégie purement sensori-motrice en associant simplement une action (une direction à suivre) à un lieu. Dans la suite de mes travaux, nous n'utiliserons pas de carte cognitive mais seulement des associations entre lieux et actions à réaliser. Ce choix nous permettra, dans un premier temps, de pouvoir présenter des solutions dans le cas, plus simple, de la navigation sensori-motrice.

2.4 Navigation robotique bio-inspirée

Basée sur les nombreuses études des solutions utilisées chez l'animal pour naviguer, une approche concurrente de la navigation robotique classique s'est développée en proposant des modèles inspirés de la biologie. Cette approche a permis l'apparition de robots réactifs, rapides et capables de comportements simples et robustes. A l'inverse de l'intelligence artificielle classique, les solutions obtenues sont souvent sous-optimales mais adaptatives et se prêtent particulièrement bien aux environnements inconnus. Certaines approches bio-inspirées se veulent également biologiquement plausible et, de ce fait, permettent un enrichissement mutuel des disciplines de la robotique et des sciences du vivant. Les expériences robotiques permettent de confirmer ou d'infirmer certaines hypothèses sur les modèles.

A partir des classifications proposés par les ethologues [Gould, 1986], différents modèles d'architectures de contrôle pour la navigation ont été proposés [Trullier and Meyer, 1997]. [Mallot and Franz, 2000] proposent de faire une distinction hiérarchique entre stratégies locales de retour au nid (fig. 2.8) et stratégies de recherche de chemin (fig. 2.9). Un classement hiérarchique existe également au sein de chacune des 2 catégories. Parmi les stratégies locales : l'agent a recours à une exploration aléatoire en l'absence d'indice sensoriel. Il peut suivre un sentier si celui-ci existe (dépôt de phéromones ou traces visuelles par exemple). Il peut viser le nid par remontée de gradient si un indice sensoriel saillant y est présent. Enfin, l'agent peut rejoindre le nid par l'apprentissage de relations spatiales entre celui-ci et des amers proches, sans codage de la position du nid lui-même.

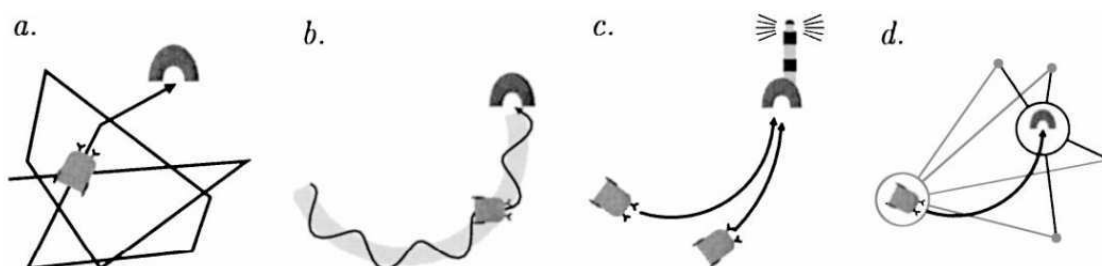


FIGURE 2.8 – Hierarchie des stratégies locales de retour au nid (homing) selon Franz et Mallot. Figure extraite de [Mallot and Franz, 2000]. **a** - Exploration aléatoire. **b** - Suivi de piste ou de sentier. **c** - Visée du nid (nécessite un indice sensoriel saillant sur le nid). **d** - Guidage vers le nid par le biais de relations spatiales entre amers environnants.

Parmi les stratégies de recherche de chemin : Si l'environnement est relativement statique, un chemin peut être appris par une séquence simple d'associations entre états (lieux) et actions (direction à suivre). Dans le cas d'apparition d'obstacles, des chemins multiples peuvent être appris par la construction d'une topologie entre lieux proches. L'agent connaît ainsi une autre route en cas d'échec sur la première (navigation topologique). Enfin, l'exploration des alentours permet de découvrir des raccourcis (navigation métrique).

2.4.1 Modèles inspirés des insectes

L'une des premières constatations de l'approche bio-inspirée a été de dire que, dans de nombreux cas, la construction d'une carte précise du monde n'est pas nécessaire pour résoudre la tâche. Cette constatation est particulièrement vraie dans le cas de l'insecte, dont les ressources

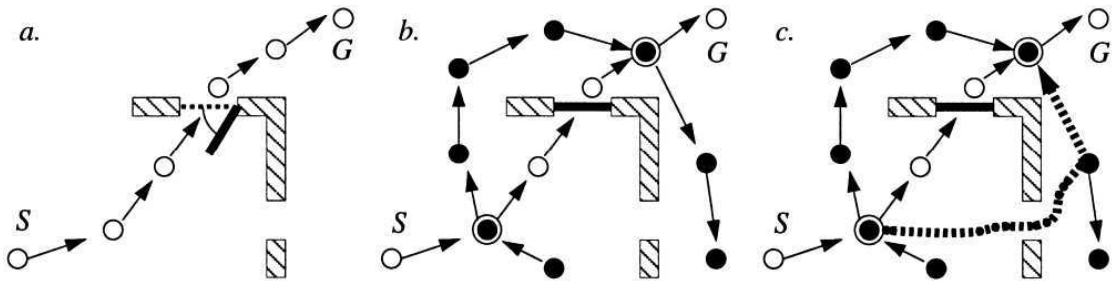


FIGURE 2.9 – Hierarchie des stratégies de planification de chemin selon Franz et Mallot. Figure extraite de [Mallot and Franz, 2000]. **a** - Chemin simple défini par une séquence d'associations état/action. La stratégie ne s'adapte pas à l'apparition de nouveaux obstacles (e.g porte fermée). **b** - Des chemins multiples sont envisageables après la construction d'une topologie entre états. L'agent peut choisir une autre route en cas d'échec. **c** - Explorer les alentours permet à l'agent de découvrir des raccourcis (pointillés) aux travers de lieux inexplorés.

cérébrales sont limitées. Dans de nombreux cas, l'animal n'a pas besoin de maintenir un modèle interne du monde, le monde lui-même est son meilleur modèle [Brooks, 1991].

Probablement, la première solution bio-inspirée proposée pour la navigation d'un robot mobile remonte aux années 50 avec le modèle de la "tortue" de Walter [Walter, 1953]. Le robot, doté de roues, de capteurs photosensibles et de lampes à vide, était capable de remonter un gradient de lumière, de déclencher un arrêt sur obstacles, de se recharger en totale autonomie et même de mimer un apprentissage par conditionnement pavlovien (association entre vision et son) jusqu'alors réservé aux systèmes vivants.

Plus tard, Braitenberg améliora ce principe en montrant comment des véhicules à l'architecture minimaliste (2 neurones) pouvaient exhiber des comportements complexes [Braitenberg, 1984]. Inspirés de l'abeille, les véhicules étaient capable d'évitement d'obstacles et de phototaxie (suivi de lumière). Un panel de comportements complexes émergeait de la dynamique de la boucle entre agent et environnement, et ce, en ne modifiant que le poids des synapses (comportement peureux, amoureux, agressif, etc...).

Les travaux de Brooks au début des années 90 ont permis l'apparition de contrôleurs réactifs [Brooks, 1991]. Son architecture de "Subsumption" a permis le contrôle et l'arbitrage en temps réel d'un ensemble de comportements en parallèle, ce qui restait jusque là impossible avec des contrôleurs basés sur un modèle du monde.

Franceschini puis Floreano ont travaillé sur des modèles, inspirés de la mouche, capables de maintenir le flot optique constant (à partir de la théorie écologique de Gibson [Gibson, 1978]). Ces modèles permettent à un drone volant d'atterrir en toute sécurité sans besoin d'une estimation complexe de la distance ou de la géométrie du sol. La vitesse du drone est proportionnelle à sa distance aux obstacles [Ruffier and Franceschini, 2005; Floreano et al., 2009].

Cartwright et Collet proposent un modèle inspiré des abeilles pour la mémorisation du but [Cartwright and Collett, 1983]. Le modèle réalise une détection panoramique d'amers (snapshot) et permet l'estimation de la direction au but en sommant les contributions des déplacements angulaires de chaque amer. Ce modèle précède les premiers modèles de cellules de lieu. L'association d'amers est difficile et nécessite de connaître l'orientation de l'agent. Une variante améliorée et basée sur la couleur a été proposée par Gourichon et al. [Gourichon et al., 2002]. Gaussier et Zrehen propose un modèle minimaliste, inspiré des insectes, supposant que ces derniers peuvent reconnaître individuellement les amers pour accéder à leur azimut. Ce modèle permet un

retour au nid par l'apprentissage des relations spatiales entre les amers environnants et le lieu but [Gaussier and Zrehen, 1995].

2.4.2 Modèles inspirés des mammifères

L'implémentation d'algorithmes copiant le comportement des insectes a donné de très bons résultats mais le passage à des mécanismes cognitifs plus complexes tels que ceux utilisés par les mammifères reste encore un challenge dans les systèmes actuels. Le début des années 1990 a vu naître les premiers modèles computationnels de cellules de lieu hippocampiques. En 1994, Burgess propose un premier modèle basé sur la hauteur apparente de 2 amers visuels (les murs de l'enceinte) [Burgess et al., 1994]. Dans le même temps, Gaussier et Zrehen aboutissent à l'architecture **PerAc** qui servira de base à nos travaux (l'architecture est présentée en détails dans la section 2.5) [Gaussier and Zrehen, 1995].

Quelques années plus tard, Redish et Touretzky proposent un modèle théorique des processus de navigation à l'oeuvre dans la région hippocampique [Redish and Touretzky, 1997]. Dans ce modèle, le code spatial des cellules de lieu est créé à partir d'informations visuelles (vues locales) et d'informations idiothétiques par le biais du subiculum (intégration de chemin). La sélection de l'action serait réalisée en partie par les ganglions de la base (nucleus accumbens) à partir d'informations spatiales fournies par l'hippocampe.

Arleo et al. propose ensuite un modèle de l'hippocampe basé sur des cellules de lieu et de direction de la tête combinant des informations allothétiques (visuelles) et idiothétiques (proprioceptives) [Arleo and Gerstner, 2000; Arleo and Rondi-Reig, 2007]. Le système visuel projette dans le cortex entorhinal superficiel, tandis que le système idiothétique (intégration de chemin) projette dans EC profond. L'apprentissage non supervisé de cellules de lieu est réalisé dans les couches superficielles de EC par une règle de Hebb. Le processus d'intégration de chemin est modélisé par des cellules dont l'activité est une fonction gaussienne de la position et de la direction du robot et fait abstraction de la plausibilité biologique. Finalement le recrutement de cellules de lieu multimodales est réalisé de manière autonome à partir d'un seuil de reconnaissance. L'environnement des expériences est réel mais contrôlé et très simplifié : les murs sont striés afin de faciliter la détection de fréquences (amers artificiels) pour la reconnaissance des cellules de vue. Le pavage automatique génère quelques 800 cellules de lieu pour un petit environnement (80x80cm). Les cellules de lieu obtenues servent de base pour un apprentissage par renforcement de l'environnement, mais cette méthode requiert un temps trop important pour des environnements de grandes tailles.

D'autres travaux se basent également sur des modèles de cellules de lieu pour étudier des solutions intéressantes de sélection de l'action mais se contentent d'une modélisation en simulation [Foster et al., 2000; Dollé et al., 2008]. De plus, l'activité des cellules est une fonction gaussienne de la position connue du robot.

Assez récemment, Milford et Wyeth ont proposé un modèle bio-inspiré, nommé "RatSLAM", pour la cartographie d'un environnement de très grande taille [Milford and Wyeth, 2008]. Le modèle s'inspire de la structure hippocampique du rat et permet à un véhicule mobile de construire en temps réel une carte d'une taille conséquente à partir d'une simple webcam (un quartier urbain complet pour un trajet de 66 km et la catégorisation de 12000 lieux). Une recalibration de la position est effectuée lorsque le système reconnaît visuellement un lieu et un mécanisme de maintien de cohérence permet de corriger la carte à partir de l'odométrie et d'une compéti-

tion entre de nombreuses poses caméra. Néanmoins, le système est testé en passif d'une part et se veut également "biologiquement inspiré" sans pour autant présenter une grande plausibilité biologique.

La navigation au travers de grands environnements intéresse également Cummins et Newman [Newman and Cummins, 2008] qui utilise une méthode dite d'"apparence seule". Le système assigne à chaque nouveau panorama un lieu rencontré précédemment, ou bien un nouveau lieu si la différence constatée est trop importante. Cette méthode ne nécessite aucune estimation de la position métrique du robot et a permis de naviguer sur plus de 1000 kilomètres en milieu urbain.

2.5 Per-Ac : un modèle de couplage sensori-moteur pour la perception

Dans de précédents travaux datant du début des années 90, Gaussier et Zrehen ont développé un modèle générique de couplage permettant à un robot d'apprendre des conditionnements entre informations sensorielles (la reconnaissance d'un lieu spatial, d'une position particulière d'un membre, d'une expression faciale, etc...) et action (une direction à suivre, une configuration motrice) [Gaussier and Zrehen, 1995] (voir fig. 2.10). L'élégance du modèle est de proposer une architecture générique de bas niveau et qui ne requiert aucune représentation symbolique du but. Pour la navigation spatiale, l'équipe Neurocybernetique du laboratoire ETIS a proposé un modèle de l'hippocampe et des boucles corticales permettant de contrôler un robot mobile lors de tâche de navigation visuelle [Gaussier et al., 2002; Giovannangeli and Gaussier, 2010; Banquet et al., 2005] et proprioceptive [Gaussier et al., 2007].

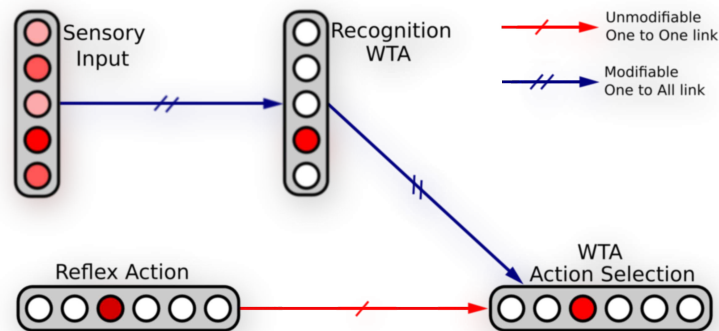


FIGURE 2.10 – Per-Ac : Une architecture générique de couplage entre Sensation et Action. Une voie inconditionnelle (en bas) déclenche des actions de manière réflexe. Une autre voie non-réflexe (en haut) apprend à reconnaître une situation sensorielle par le biais d'une compétition (WTA), puis à associer cette situation à l'action déclenchée par la voie réflexe. Une fois l'apprentissage terminé, la reconnaissance de la situation déclenchera l'action correspondante, même lorsque la voie réflexe est inactive.

2.5.1 Modèle de cellules de lieu visuelles

Une image panoramique est analysée séquentiellement pour en extraire un ensemble de points d'intérêt qu'une cellule de lieu sera amenée à reconnaître.

La caméra monoculaire embarquée, montée sur servomoteur pan-tilt, permet la capture de 15 images successives, correspondant à un panorama à 360 degrés³. Le système réalise ensuite un traitement d'images qui consiste en (1) le calcul de l'image du gradient (faisant apparaître les contours des objets à fort contraste dans la scène), puis (2) une convolution avec un filtre DoG (différence de gaussienne) qui permet de faire ressortir les points de forte courbure des contours. Ces points de courbure sont intéressants puisqu'ils sont robustes et invariants aux changements d'échelle. Ils correspondent la plupart du temps à des coins d'objets ou des éléments saillants anguleux (e.g. les 4 coins d'un tableaux sombre sur fond clair constituent des points d'intérêt robustes). Ces traitements permettent, à partir d'images brutes, de mettre en évidence une constellation de points d'intérêt dans le panorama. Afin de reconnaître ces points particuliers, le système apprend de petites imagerie circulaires centrées sur chacun d'eux. Une transformation log-polaire est réalisée sur chaque imagerie afin de rendre la reconnaissance plus robuste aux petites rotations et aux changements d'échelle. Chaque imagerie (32*32 pixels) est apprise par un neurone codant pour un amer (*landmark*). Autrement dit, une "**cellule d'amer**" apprend à reconnaître une imagerie particulière. L'activité d'une cellule d'amer dépend de la différence entre l'imagerie perçue et l'imagerie apprise par cette même cellule.

De la même façon, une "**cellule de lieu**" (place cell) apprend à reconnaître une constellation d'amers associés à leurs azimuts (voir fig. 2.11). L'activité d'une cellule de lieu dépend de la différence entre la constellation amers/azimuts perçue et celle apprise par la cellule. Selon ce modèle, un lieu dans l'environnement est donc représenté par un agencement particulier d'images dans le panorama. Chaque cellule est capable de reconnaître un lieu particulier de manière robuste. Cette robustesse provient en partie du grand nombre d'images extraites (75 par panorama), ce qui permet de reconnaître un lieu même lorsque des objets ont été déplacés ou bien qu'une partie du panorama est occultée. La robustesse provient également de l'utilisation d'un mécanisme de compétition (WTA) qui sélectionne la cellule de lieu la plus active au temps t , c'est à dire le lieu le mieux reconnu. De cette manière, la reconnaissance d'un lieu ne dépend pas d'un niveau absolu de reconnaissance, mais bien d'un niveau relatif entre l'activation des différents lieux. Dans ce modèle, les images locales extraites correspondraient à l'information "What" (Quoi) codée dans le cortex perirhinal ou dans d'autres aires du chemin visuel ventral du cortex temporal du rat [Kolb and Tees., 1990]. Les angles sous lesquels sont vu les amers correspondraient à l'information "Where" (Où), fournie par le cortex pariétal par le biais de la région parahippocampique. La fusion des informations "What" et "Where" serait réalisée dans la couche superficielle du cortex entorhinal ou dans le cortex postrhinal [Suzuki et al., 1997; Burwell and Hafeman, 2003]. Dans ce modèle, cette fusion est réalisée par un réseau nommé PrPh (Perirhinal & Parahippocampique). Un neurone du groupe PrPh ne s'active que lorsqu'un amer est reconnu sous un azimut particulier appris pour cet amer. L'azimut absolu est obtenu par la somme entre la position relative de l'amer dans l'image, l'angle relatif de la caméra (tête) et la direction absolue du robot. Le système de cellules de direction de la tête fournit des informations sur l'orientation du robot. Ce système est alimenté par des données provenant d'une boussole magnétique, mais peut se baser uniquement sur une boussole visuelle [Leprêtre et al., 2000; Giovannangeli and Gaussier, 2007] et être complété par des informations proprioceptives.

Un système de recrutement permet d'encoder de nouveaux amers si ceux-ci n'avaient jamais été appris. L'activité de cette population de neurones permet donc d'identifier un amer perçu.

3. Le modèle n'a pas besoin de réaliser un panorama complet et peut se contenter de mettre à jour une mémoire dynamique. Il fonctionne également si le robot se déplace pendant la prise de panorama

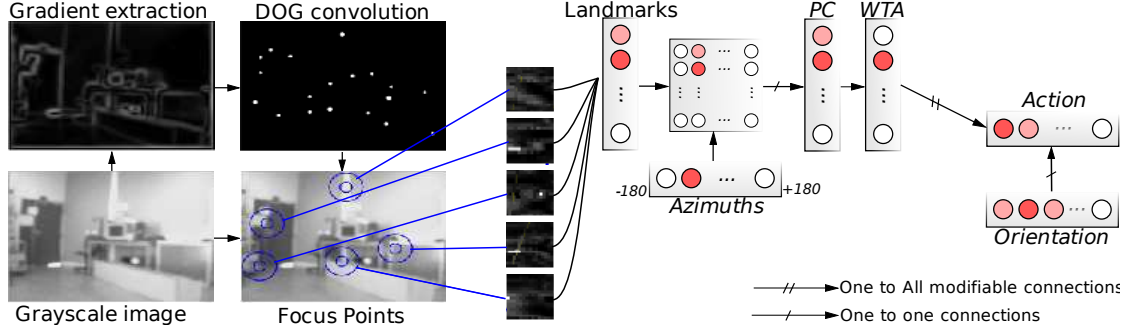


FIGURE 2.11 – Modèle de cellules de lieu : architecture Per-Ac pour la navigation visuelle. A chaque instant, la caméra du robot prend un panorama visuel de l’environnement à 360 degrés. L’image du gradient de ce panorama est convoluée avec un filtre par différence de gaussienne (DoG) pour en extraire une constellation de points saillants. Le système apprend ensuite des imagettes locales centrées sur les points plus saillants. Ces imagettes sont apprises par des cellules d’amer (landmarks). Un réseau PrPh fusionne les amers et leur azimuts respectifs. Chaque cellule de lieu (PC) apprend ensuite à reconnaître une constellation amers-azimuts particulière. Un mécanisme de compétition (WTA ou softWTA) sélectionne la (ou les) cellule(s) de lieu la (les) plus active(s) au temps t . Enfin, une association est apprise entre la cellule de lieu gagnante et l’orientation actuelle du robot, après quoi le robot s’orientera dans la direction apprise lorsque la cellule correspondante est de nouveau gagnante.

L’équation pour l’apprentissage des amers en fonction des imagettes locales est la suivante :

$$\frac{dW_{ij}^{VA}(t)}{dt} = R_i^A(t) \cdot X_j^V(t) \quad (2.1)$$

Où W_{ij}^{VA} représente le poids synaptique entre un neurone de vue locale j et un neurone d’amers i . $R_i^A(t)$ est un signal de type échelon binaire permettant le recrutement d’un neurone lorsque l’activité du gagnant de la population de neurones passe en dessous du seuil de vigilance ν^A . X_j^V est l’activité d’une cellule ayant appris à reconnaître un amer (vue locale), après transformation log-polaire (un neurone par pixel).

L’équation pour le calcul de l’activité X_i^A d’un neurone i recruté correspondant à un amer visuel est la suivante :

$$X_i^A(t) = 1 - \frac{\sum_j |W_{ij}^{VA}(t) - X_j^V(t)| \cdot H_\epsilon(W_{ij}^{VA}(t))}{\sum_j H_\epsilon(W_{ij}^{VA}(t))} \quad (2.2)$$

avec H_ϵ la fonction de Heavyside telle que $H_\epsilon(x) = \begin{cases} 0 & \text{si } x < \epsilon \\ 1 & \text{si } x \geq \epsilon \end{cases}$

L’activité $X_i^A(t)$ d’un neurone non recruté est aléatoire (sa valeur est comprise entre 0 et B , B représentant le niveau de bruit neuronal maximal). La fonction H_ϵ permet de détecter les poids ayant appris (ϵ est une valeur positive, proche de 0).

L’apprentissage du groupe PrPh est contrôlé par les équations suivantes :

$$\frac{dW_{ij}^{\phi P}(t)}{dt} = R_i^P(t) \cdot H_\gamma(X_j^\phi(t)) \quad (2.3)$$

$$\frac{dW_{ij}^{AP}(t)}{dt} = R_i^P(t) \cdot H_\gamma(X_j^A(t)) \quad (2.4)$$

$R_i^P(t)$ est un signal de recrutement de type échelon binaire, déclenché lorsque l’activité du gagnant de la population de neurones passe sous un seuil de vigilance ν^P . La fonction H_γ

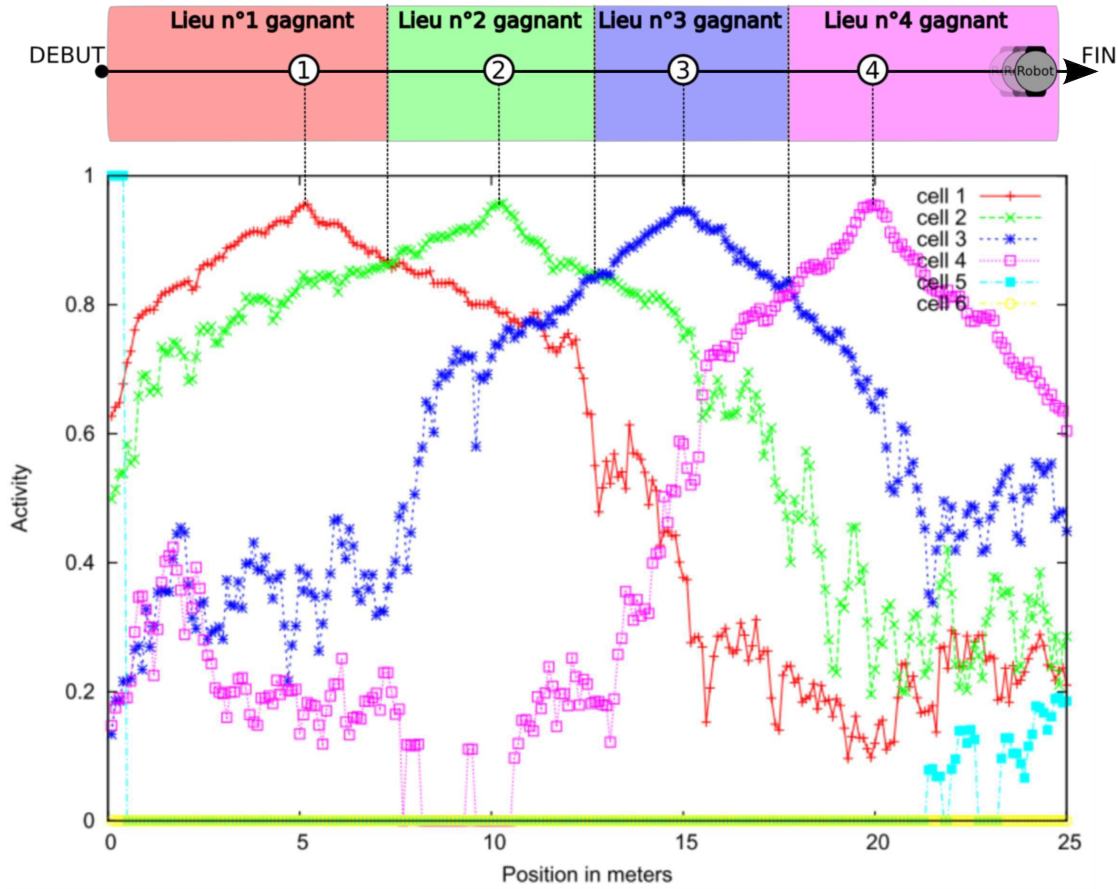


FIGURE 2.12 – Activité de 4 cellules de lieu enregistrées sur une trajectoire rectiligne de 25m en environnement extérieur. Les 4 maxima correspondent aux endroits précis où les lieux ont été appris. Les points de croisement correspondent aux frontières entre lieux gagnants. Le modèle montre de bonnes capacités de généralisation puisque le champ de lieu d’une cellule (champ pour lequel l’activité de la cellule est au dessus du bruit) est plus large que la zone pour laquelle cette cellule est gagnante. La cellule n°5 correspond à un lieu appris en dehors de l’enregistrement présenté. La cellule n°6 n’a pas été recrutée et son activité est donc nulle.

détecte les neurones fortement activés en entrée du PrPh (γ étant proche de 1).

L’équation pour le calcul de l’activité X_i^P du neurone i du groupe PrPh est la suivante :

$$X_i^P(t) = \max(\lambda_i(t) \cdot X_i^P(t - dt), f(\max_j W_{ij}^{\phi P} \cdot X_j^\phi \cdot \max_k W_{ik}^{AP} \cdot X_k^A - \theta)) \quad (2.5)$$

avec f une fonction de seuil telle que $f(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$

θ est le seuil d’activation des neurones du PrPh. X^A représente l’activité de l’amer reconnu et X^ϕ l’azimut sous lequel il est perçu. Un seul amer est traité à chaque itération ce qui implique que le PrPh fonctionne comme une mémoire à court terme, dont le facteur d’oubli est $\lambda_i(t)$. Il permet de conserver en mémoire les amers perçus pendant les prises de vue d’un panorama. La mémoire d’un amer est remise à jour par une meilleure reconnaissance et également après un temps défini arbitrairement (après un ou plusieurs panoramas) [Giovannangeli, 2007].

L'activité des cellules de lieu, même en condition extérieure, montre un pic d'activité sur le lieu appris et généralise assez bien sur de grandes distances (2 à 3m en intérieur et 20 à 30m en extérieur) (voir fig. 2.12). Ces profils d'activité large ne correspondent pas aux enregistrements hippocampiques (CA1 et CA3) mais plutôt aux profils de cellules enregistrées dans le subiculum ou le cortex entorhinal [Quirk et al., 1992; Sharp, 1999b].

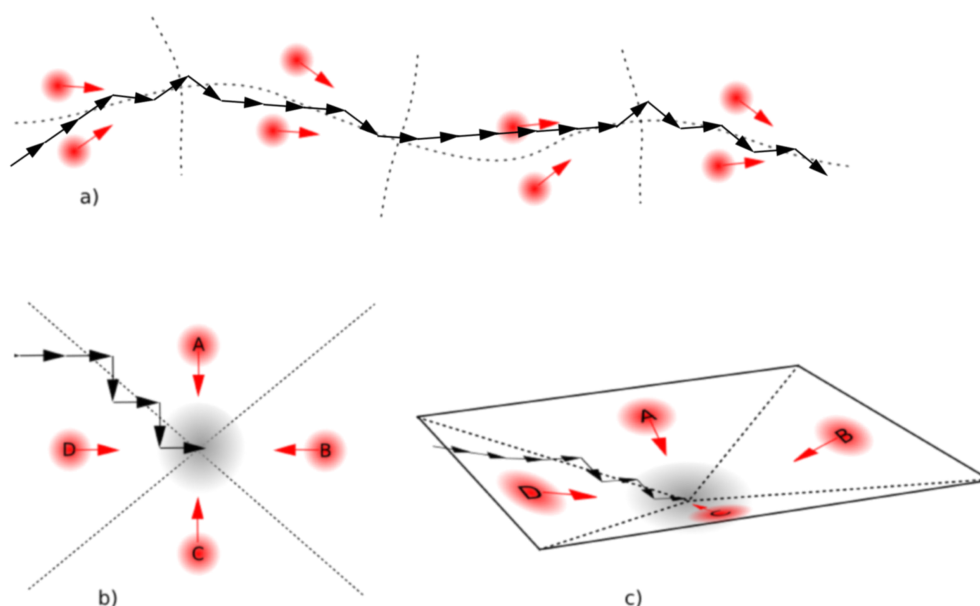


FIGURE 2.13 – Principe de l'apprentissage de comportements simples à base de quelques associations Lieu/Action. Une action est associée à la cellule de lieu gagnante. Chaque lieu est ainsi associé à une direction à suivre (flèche rouge). Chaque fois que le robot se trouve dans un lieu, il se tourne automatiquement dans la direction apprise dans ce même lieu. Ce mécanisme permet au robot d'apprendre des comportements de type ronde (a), ou encore de type retour au nid (homing)(b) à partir de quelques associations Lieu/Action seulement. L'idée maîtresse ici est que la trajectoire n'est pas prescrite mais émerge de la dynamique sensorimotrice en créant des bassins d'attractions (c). A titre d'exemple, le comportement appris en (b) permet au robot de converger vers un lieu but au centre. Pourtant ce lieu n'est pas codé dans la mémoire du robot, mais émerge naturellement comme le fond du bassin d'attraction généré par la dynamique de reconnaissance des lieux A, B, C et D. Dans un monde réel bruité, ce type d'apprentissage est plus robuste qu'une simple remontée de gradient et permet au robot de converger vers le but en "tombant" dans un bassin d'attraction.

Chaque cellule de lieu peut ensuite être associée à une action particulière (une orientation à suivre dans notre cas) et peut servir de base à des stratégies sensori-motrices simples. Dans le cadre de l'apprentissage d'une tâche de navigation supervisée, l'humain peut donner la direction à suivre comme signal désiré. Le robot apprendra alors à associer l'état courant (la cellule de lieu gagnante) à la direction désirée. On parlera d'**association Lieu/Action**. A partir de 3 cellules (ou plus) dont les actions associées pointent vers un but virtuel, il est possible de construire un bassin d'attraction autour de ce but (fig. 2.13-b). La dynamique de reconnaissance des lieux permet au robot de converger naturellement vers le but, sans que celui-ci ne soit explicitement codé [Gaussier and Zrehen, 1995; Gaussier et al., 2000]. Cette stratégie permet également l'apprentissage de ronde (voir fig.2.13-a) [Giovannangeli and Gaussier, 2010]. Ce sont ces rondes sensori-motrices que l'on étudiera abondamment dans la suite de ces travaux. Cette stratégie d'association Lieu/Action se limite cependant à un environnement simple dans

lequel un choix entre plusieurs chemins possibles n'est pas nécessaire. Si plusieurs chemins sont possibles, ou que de nouveaux obstacles induisent des minima locaux, on peut avoir recours à une stratégie de planification basée sur la construction d'une carte cognitive [Gaussier et al., 1997]. Atteindre plusieurs buts reste cependant possible avec une stratégie sensorimotrice simple si un bassin d'attraction différent est créé pour chaque but [Gaussier et al., 2000].

2.5.2 Modèle d'intégration de chemin

Outre ce modèle de cellules de lieu permettant une stratégie de navigation visuelle, l'équipe a également développé un modèle d'intégration de chemin à partir d'informations idiothétiques [Gaussier et al., 2007] basé sur les travaux de [Mittelstaedt, 2000]. Une telle stratégie se révèle essentielle dans certaines conditions de privation sensorielle (lorsque l'agent est plongé dans le noir par exemple) et peut être mise en place sur une plate-forme robotique possédant des capteurs odométriques. La stratégie se base sur 2 informations codant pour un mouvement élémentaire du système : la vitesse linéaire ainsi que la vitesse de rotation instantanée (voir fig. 2.14). Les mouvements élémentaires du robot définissent à chaque itération des vecteurs unitaires dans un champ d'entrée V_i . Un champ de sortie D_i somme simplement l'activité de V_i au cours du temps. Dans ce modèle, les champs V_i et D_i sont des champs neuronaux simplifiés (sans interactions latérales) et supposés de même taille (121 neurones). Chaque neurone code pour une portion angulaire du panorama total à 360° (codage en population). Un nombre impair de neurones est préférable pour des raisons de symétrie, et le nombre de neurones utilisés est un multiple de 360. Par la suite, nous parlerons de **champ d'intégration de chemin** pour désigner le champ de sortie D_i .

Un masque sinusoïdal non-négatif⁴ est convolué autour de la direction de chaque mouvement élémentaire dans le champ de neurone d'entrée V_i :

$$V_i(t) = 1 + \cos(\Theta(t) - \theta_i) \quad (2.6)$$

Avec $\theta_i = -2\pi \frac{i}{N}$

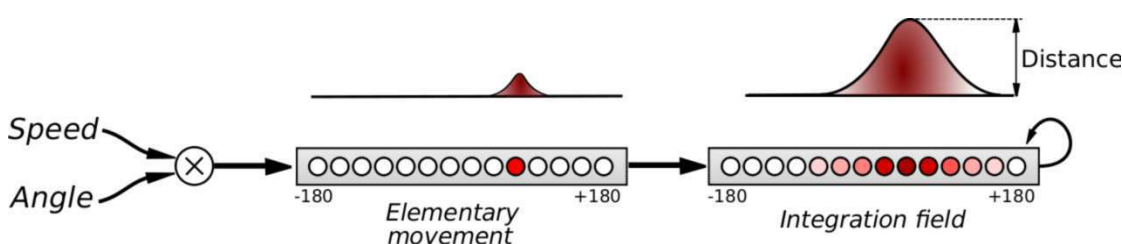


FIGURE 2.14 – Modèle neuronal utilisé dans notre architecture pour le calcul de l'intégration de chemin. Un champ neuronal avec liens récurrents représente une mémoire du chemin parcouru. Le champ intègre les mouvements élémentaires du robot à chaque itération. L'index du neurone le plus actif représente l'orientation du vecteur de mouvement global et son amplitude, la distance globale parcourue. Les mouvements élémentaires sont représentés au sein d'un vecteur de neurones alimenté par l'orientation absolue du robot et sa vitesse linéaire instantanée.

Le champ de sortie D_i réalise simplement une intégration temporelle de l'activité du champ d'entrée V_i . L'amplitude d'un mouvement élémentaire est modulée par la vitesse linéaire du

4. Une fonction gaussienne ou triangulaire peuvent en pratique faire l'affaire mais la démonstration implique l'usage du cosinus

robot. L'orientation du robot au temps t sera noté $\Theta(t)$. L'activité du champ d'intégration de chemin est défini telle que :

$$D_i(t) = [D_i(t - dt) + \alpha V_i(\Theta(t)) - r(t)]^+ \quad (2.7)$$

Avec $r(t)$ un signal de remise à zéro de type échelon binaire, utilisé pour la recalibration du champ d'intégration de chemin. $[x]^+ = x$ si $x > 0$ et 0 sinon. α représente le gain du système permettant d'empêcher la saturation du champ ($0 < \alpha < 1$).

Sans connections latérales, chaque neurone D_i réalise simplement la somme de toutes ses activités passées depuis la dernière remise à zéro. Le champ D encode le chemin total parcouru depuis le point de départ et conserve une forme sinusoïdale (voir fig. 2.15). L'amplitude du neurone gagnant reflète la distance parcouru en ligne droite (à vol d'oiseaux) et son abscisse représente la direction du mouvement global (voir [Gaussier et al., 2007] pour une description complète).

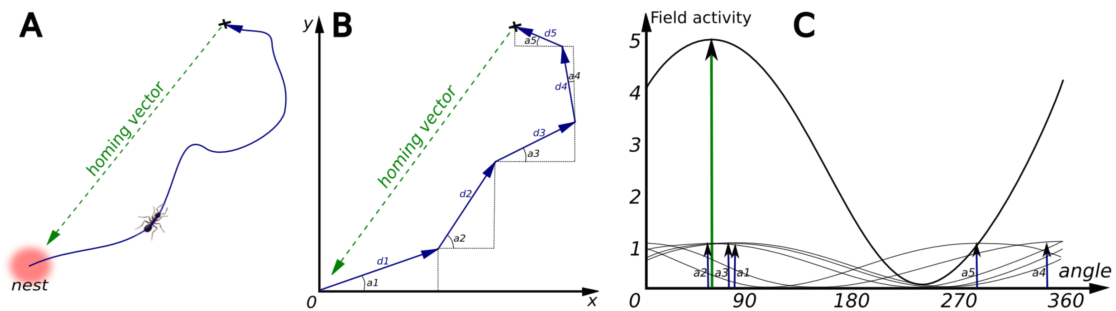


FIGURE 2.15 – Principe de l'intégration de chemin. **A** - Les animaux ont cette capacité d'évaluer la distance les séparant de leur nid ainsi que l'orientation à prendre pour le rejoindre (homing vector), et ce simplement en intégrant leurs mouvements au fil de leur déplacement. **B** - De la même manière, un robot peut estimer en temps-réel son vecteur-retour en cumulant angle et norme de chacun de ses vecteurs mouvements élémentaires. **C** - Dans notre modèle, ces vecteurs sont représentés par des sinusoïdes dont l'amplitude correspond à la distance et la phase à l'orientation. On montre que la résultante de la somme de ces sinusoïdes n'est autre qu'une sinusoïde correspondant au vecteur-mouvement global.

Cette stratégie se trouve être plus précise localement en comparaison d'une stratégie basée sur des informations allothétiques (externes) telle que la navigation visuelle. Néanmoins l'intégration de chemin possède l'inconvénient d'introduire une erreur cumulative d'autant plus grande que l'on navigue depuis longtemps. Par la suite, il sera nécessaire de recourir à un mécanisme de recalibration pour limiter cette dérive. Nous reviendrons sur ce mécanisme de recalibration dans le chapitre 4.

Dans le chapitre suivant, nous allons nous intéresser à l'analyse du modèle de navigation sensori-moteur PerAc en adoptant un point de vue probabiliste. Nous verrons les liens existants entre ce modèle réactif et d'autres modèles statistiques de reconnaissance de lieu, basés sur une approche bayésienne.

Un parallèle entre formalisme neuronal et bayésien

La navigation robotique autonome en environnement réel repose sur des mesures bruitées. C'est la raison pour laquelle la plupart des travaux robotique de localisation, de cartographie ou encore de planification sont basés sur des algorithmes probabilistes. La nature statistique de ces méthodes permet de prendre en compte les incertitudes présentes sur les données mesurées au détriment d'une faible réactivité. Pourtant, la nature nous montre que la plupart des animaux sont tout à fait capables d'extraire l'information utile à partir de signaux bruités, tout en conservant cet aspect réactif. Les approches neuromimétiques peuvent donc nous fournir des solutions robustes et innovantes pour aborder ce genre de problème. Par ailleurs, de récents travaux ont démontré que le cerveau était capable de réaliser des calculs très similaires aux processus bayésiens [Deneve S, 2004]. De même qu'il est connu depuis longtemps que l'apprentissage d'un conditionnement classique est équivalent au calcul d'une probabilité conditionnelle [Rescorla, 1988]. Les approches bayésiennes et neuromimétiques tendent donc à se rapprocher.

Nous proposons, dans ce chapitre, de comparer ces 2 approches dans le cadre de l'auto-évaluation et de la navigation sur de grandes distances. En prenant inspiration de travaux en psychologie et neurobiologie, nous avons d'ores et déjà proposé des solutions pour permettre à un robot de naviguer dans de petits environnements (typiquement 1 salle) et pour de courtes durées (quelques heures) : se référer à la section 2.5 du chapitre précédent. Cependant, plusieurs questions restent ouvertes si l'on désire réaliser le même type de tâches dans des environnements fortement dynamiques et sur de longues périodes temporelles. Les systèmes de navigation robotique autonomes font face à des problèmes d'imprécisions des mesures liées au capteurs et d'incertitudes liées à la contrainte d'un environnement dynamique. L'exploration et la navigation sur le long terme sont des tâches assez ardues. Le changement d'échelle dans la taille de l'environnement implique un apprentissage et une gestion d'un volume de données conséquent qui ouvre certaines interrogations. Comment outrepasser le problème du rapport signal à bruit qui tend à décroître avec le volume d'information ? Comment lever l'ambiguïté présente sur les informations visuelles et réduire l'incertitude de la localisation ? Les systèmes robotiques visant un comportement autonome doivent faire face à 2 principaux défis. Premièrement, le système doit être capable d'extraire et apprendre de manière active l'information pertinente en ignorant celle inutile à la tâche. Deuxièmement, le système doit être capable d'évaluer la qualité de son comportement et l'adapter en fonction de la situation.

Dans ce chapitre nous nous intéressons au premier point en proposant une réflexion et des solutions au problème de l'extraction de l'information pertinente pour la tâche. On notera qu'une telle capacité à sélectionner l'information pertinente implique en elle-même une certaine capacité d'évaluation.

Nous commencerons par décrire les limitations du modèle d'apprentissage figé précédent. Adopter une approche probabiliste permet de réduire l'ambiguïté inhérente à toute perception visuelle en donnant au système une capacité à inférer sa position par l'apprentissage de la pertinence des amers dans un lieu donné. Nous ferons un rapide tour des méthodes existantes dans la littérature traitant ce type de problème avant d'analyser notre propre modèle de cellules de lieu hippocampiques sous le prisme de l'approche probabiliste. Nous n'essaierons pas de comprendre le fonctionnement de ce réseau neuronal en termes probabilistes, mais plutôt de proposer deux modèles probabilistes équivalents au modèle neuronal. Ces 2 modèles probabilistes correspondent à 2 façons différentes d'aborder le problème. Ils utiliseront simplement les mêmes entrées et viseront à fournir des sorties équivalentes en empruntant finalement au modèle neuronal que les notions que ce dernier manipule.

A partir de cette comparaison entre approche neuronale et bayésienne, nous montrerons comment un robot est capable d'apprendre à inférer la reconnaissance d'un amer à partir de la perception de son azimut. Nous montrerons, de la même manière, comment construire des cellules de lieu "probabilistes" en apprenant à inférer la reconnaissance d'un lieu à partir de la perception d'une constellation d'amers-azimuts donnée.

3.1 Limitations du modèle actuel

Comme nous l'avons vu précédemment, notre architecture de reconnaissance de lieu a donné de bons résultats lors de tâches de navigation dans de petits environnements et sur de courtes durées. Les capacités de généralisation du système tenaient la première place, et l'usage d'une simple compétition permettait de bonnes performances comportementales lors de tâches de navigation typiquement limitées à une pièce et sur une période temporelle restreinte à quelques heures. On peut se poser la question de la capacité de l'architecture à traiter des problèmes de plus grande taille (plusieurs salles, environnements extérieurs) et à maintenir des performances identiques à long terme (plusieurs jours, semaines, mois). Le système fait alors face à plusieurs problèmes majeurs :

1. La nature ambiguë des informations visuelles amène à des interprétations multiples des vues locales extraites. Lors d'une navigation dans un environnement complexe, une cellule de lieu peut ainsi présenter des champs multiples, c'est à dire répondre à plusieurs endroits dans l'environnement (voir fig. 3.1). La capacité d'un simple mécanisme de compétition à proposer une interprétation cohérente diminue avec l'augmentation du nombre de vues apprises (le rapport signal à bruit tend vers 0). La distance entre les prototypes appris diminue avec le nombre d'amers et la probabilité de fausses alarmes augmente fatalement [Giovannangeli, 2007]. Ce problème est bien connu en robotique mobile sous le nom d'*aliasing perceptif* (perceptual aliasing) lorsque différentes portions de l'environnement se ressemblent fortement [Marsland et al., 2001]. De précédents travaux ont montré qu'une information de contexte apprise à la volée pouvait constituer une première solution à ce problème en permettant à la fois de limiter la compétition (augmenter le rapport signal à bruit) mais aussi de limiter la puissance de calcul nécessaire en ne calculant

- globalement que l'activité des neurones associés au bon contexte [Cuperlier et al., 2008].
2. Dans l'architecture actuelle, tous les amers sont traités de manière identique. Le système est incapable de procéder à un tri. Pourtant les amers peuvent avoir des caractéristiques différentes (proximaux ou distaux, mobiles ou immobiles) et être plus ou moins pertinents selon la tâche envisagée. Il est donc nécessaire d'introduire une capacité à prédire la position ou le déplacement de ces amers afin de les caractériser. Ceux dont la position angulaire reste invariante (sur une certaine distance) devraient avoir plus de poids dans la reconnaissance que les amers mobiles, intermittents, redondants ou ambigus. Par ailleurs, le temps de calcul évolue de façon linéaire (complexité en $O(N)$) du nombre de lieux appris [Giovannangeli, 2007]. L'ajout d'un mécanisme attentionnel de sélection des amers pertinents permettrait non seulement de réduire les ambiguïtés mais également d'obtenir une complexité asymptotique (bornée) ou logarithmique.
 3. L'environnement réel est souvent fortement dynamique. Des amers stables sont amenés à disparaître, tandis que d'autres se révèlent être mobiles et ne devraient donc pas être pris en compte. L'apprentissage actuel "en un coup" ne permet pas de rendre compte de cet aspect dynamique. Même si l'on considère le cas où de tels changements environnementaux ne se produisent que rarement, le cumul de petites modifications et les variations de luminosité posent le problème du maintien de l'activité des cellules de lieu dans le temps. En effet, les tests actuels se limitent à des expériences de quelques heures et le système est incapable de s'adapter à des reconfigurations lentes de l'environnement ou aux changements saisonniers. L'apprentissage ne peut donc être statique, il doit lui aussi être dynamique et adaptatif. Ce problème est connu sous le nom de *dérive des concepts* (concept drift) [Tsymbal, 2004; Widmer and Kubat, 1996]. Les propriétés statistiques de la variable que le modèle essaie de prédire évoluent au cours du temps d'une manière a priori imprévisible. Ceci pose problème puisque les prédictions deviennent moins exactes au fur et à mesure que le temps passe. Pour permettre au système de traiter ce problème de dérive, celui-ci doit être capable de mettre à jour son modèle régulièrement en prenant en compte les nouvelles observations. Parmi les différentes approches proposées par la communauté du "*machine learning*" [Miroslav Kubat, 2004; Tsymbal, 2004; Witten and Frank, 2005], l'apprentissage incrémental est considéré comme la méthode la plus adaptée à la résolution de ce problème.

Jusqu'à présent, la question de savoir si un amer est pertinent ou non pour la navigation a été largement ignorée ou remplacée par la préselection de modèle a priori servant à définir ce qu'est un "bon" amer [Marsland et al., 2001]. Le risque est double : le superviseur humain peut sélectionner des amers difficilement reconnaissables par la machine, ou bien, au contraire, négliger des amers qu'il juge sans intérêt mais qui pourtant seraient pertinents pour le robot.

3.2 Approche bayésienne et notion d'inférence

L'inférence statistique consiste à déduire les caractéristiques statistiques inconnues d'une population à partir d'un échantillon de cette population. Les caractéristiques de l'échantillon, une fois connues, reflètent avec une certaine marge d'erreur celles de la population. Cette méthode de réflexion est souvent attribuée à l'homme et serait fonctionnelle dès la naissance. Pourtant

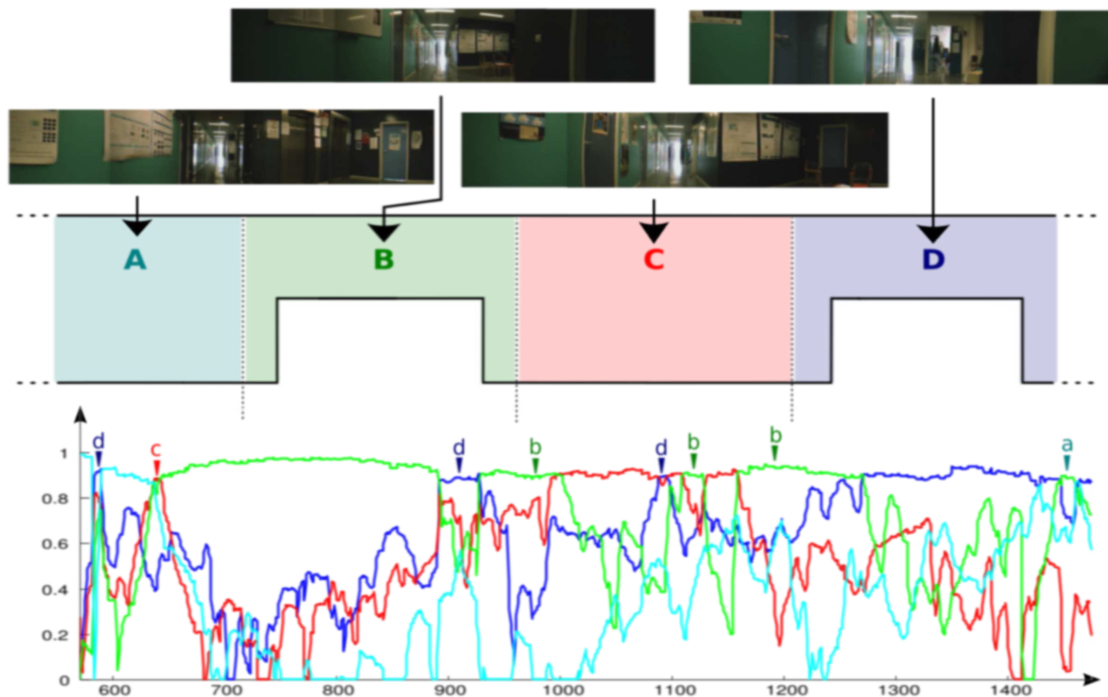


FIGURE 3.1 – Expérience mettant en évidence la présence de nombreuses ambiguïtés visuelles au sein d'un couloir. Le robot a appris 4 lieux régulièrement espacés (A,B,C et D) dans le couloir. Le niveau de reconnaissance de ces lieux est ensuite enregistré lors d'une navigation de A vers D à vitesse constante. Chaque panorama à 180° correspond à la vue subjective du robot au moment de l'apprentissage du lieu correspondant. L'environnement est ici particulièrement ambigu avec des lieux qui se ressemblent fortement. Le tracé des niveaux d'activité montre de nombreux faux positifs.

la vision dominante du XXème siècle est celle de la confusion du bébé : selon W. James, le bébé est un organisme soumis à des entrées sensorielles et incapable pendant une très longue période de les organiser convenablement [James, 1950]. Ce point de vue est partagé par la pédiatrie de l'époque qui considère alors le très jeune enfant comme dépourvu de capacité de structurer ses sensations. Cette vision a également guidé l'oeuvre de Piaget [Piaget, 1967] qui considérait l'enfant de moins de 4 ans comme "précausal", c'est à dire incapable de remonter aux causes de ce qu'il percevait. Durant la première année de vie, selon lui, l'enfant ne vivrait que dans le présent, incapable de saisir la permanence des objets, incapable de se représenter un objet disparu et encore moins de créer des modèles mentaux abstraits de ses connaissances.

Plus récemment (2004), de nouvelles visions réfutent la théorie piagetienne en proposant l'idée d'une rationalité précoce chez l'enfant. Les enfants semblent posséder des théories intuitives du monde physique, biologique, psychologique et de sa structure causale [Schulz and Gopnik, 2004]. Une première théorie propose que le bébé dispose dès la naissance d'un noyau de connaissances fondamentales et innées (*core knowledge*). Ces connaissances comprendraient les concepts d'objet [Kellman and Spelke, 1983], de nombre [Mehler and Bever, 1967], d'espace [Landau et al., 1981] et même de ce que les auteurs appellent une "*grammaire universelle*" [Chomsky, 1961; Gleitman et al., 1995; Komarova et al., 2001].

Selon une seconde théorie, quand l'enfant joue, il réaliserait en réalité des expériences permettant de tester des modèles comme le ferait un scientifique. Le bébé serait capable de ma-

nipuler des probabilités, de sélectionner les variables pertinentes, d'éliminer les variables non-pertinentes, de repérer des ambiguïtés ou interprétations multiples et enfin de lever ces ambiguïtés par le jeu [Gopnik, 2001; Gopnik and Rosati, 2001; Oudeyer et al., 2005]. Certains principes, même très abstrait, peuvent ainsi être sélectionnés par apprentissage bayésien. Il n'est alors plus nécessaire de supposer qu'ils soient innés comme le postule la théorie précédente et l'on peut supposer que le bébé découvre par lui-même les concepts d'objet, d'espace et de grammaire. L'approche bayésienne permet en fait un compromis entre connaissances innées et acquises.

La théorie de l'inférence bayésienne est une théorie mathématique simple qui peut être vue comme une extension de la logique en ce sens qu'elle permet de raisonner en présence d'incertitude. Le raisonnement peut ainsi s'étendre à un continuum de plausibilité plutôt que de se limiter à des conclusions binaires. A l'inverse de l'approche statistique, nécessitant un grand nombre d'observations, l'approche bayésienne permet des inférences rapides et sophistiquées en seulement quelques exemples. Cette particularité expliquerait, par exemple, la capacité des jeunes enfants à tirer des conclusions correctes à partir d'un petit nombre d'expériences [Gopnik et al., 2001]. Les inférences bayésiennes rendent aussi bien compte des processus de perception en proposant l'interprétation la plus probable à partir d'entrées ambiguës. Une expérience de Xu et Garcia montre que les enfants sont capables de telles inférences dès l'âge de 8 mois [Xu and Garcia, 2008]. De récents travaux suggèrent que des processus statistiques sont à l'oeuvre dans le cerveau et montrent comment certains codages en population implémentent les principes de bases de l'approche bayésienne. Afin de pouvoir réaliser ces inférences, l'enfant serait donc contraint de raisonner en termes de probabilités et serait également capable de se représenter les distributions de ces probabilités [Pouget et al., 2013]. Par ailleurs de nombreux travaux ont montré que les processus sensori-moteurs à l'oeuvre dans le système nerveux étaient bien modélisés par une approche bayésienne, aussi bien pour réduire les ambiguïtés de la perception [Kersten et al., 2004], combiner informations a priori et observations [Mamassian et al., 2002], lier les informations provenant de plusieurs modalités [Körding et al., 2004] et modéliser des raisonnements complexes [Doya, 2007].

Pratiquement toutes les méthodes de localisation robotique partagent la caractéristique commune d'aborder le problème de la navigation selon une approche probabiliste. Ces méthodes ont recours aux inférences bayésiennes afin de transformer les données brutes, extraites des capteurs, en information exploitable pour la localisation. Certains travaux font usage explicitement de méthodes d'inférences probabilistes, tandis que d'autres ne semblent pas, à première vue, reposer sur de telles méthodes mais celles-ci peuvent en réalité être interprétées comme telles sous certains aspects [Thrun, 2002b]. La raison de ce succès provient du fait que la robotique est caractérisée par des incertitudes et du bruit de mesure important. Les algorithmes probabilistes traitent le problème en modélisant explicitement différentes sources de bruit et leurs effets sur les mesures. Ce problème est sans doute le problème le plus complexe en robotique mobile et a contraint la communauté à adopter une vision singulière en utilisant une méthodologie mathématique unique pour le résoudre. Le principe de base qui sous-tend la plupart des algorithmes de SLAM est la règle de Bayes [Bayes, 1763]. Supposons que nous voulions en savoir plus sur une quantité X (e.g. un lieu), à partir de données de mesure D (e.g. données issues de capteurs ultrasons, odométrie, vision). La règle de Bayes nous dit que l'on peut déterminer la probabilité de X sachant D , si l'on connaît les probabilités de X , de D et de D sachant X :

$$p(X|D) = \frac{p(D|X)p(X)}{p(D)} \quad (3.1)$$

Le terme $p(X)$ est appelé probabilité à priori de X . Il représente l'état des connaissances sur X à un instant donné. De même, le terme $p(D)$ représente la probabilité à priori de D . Cette probabilité est un terme de normalisation de l'équation. Elle assure que la somme des probabilités soit égale à 1. Le terme $p(X|D)$ est appelé probabilité a posteriori de X sachant D . Cette probabilité représente l'inférence ou la "**conclusion**" que l'on peut tirer sur la variable X à partir des observations D . Le terme $p(D|X)$, pour un D connu, est appelé fonction de vraisemblance (likelihood) de X . Il spécifie la probabilité d'observer la mesure D sous l'hypothèse X . On peut comprendre la règle de Bayes comme :

$$Posteriori = \frac{Vraisemblance.Apriori}{Normalisation} \quad (3.2)$$

Autrement dit, si l'on connaît dans le détail les causes possibles d'une conséquence observée, l'observation des effets permet de remonter aux causes. A titre d'exemple, la simple observation de la tour Eiffel permet avec une grande probabilité d'en déduire que l'on se situe à Paris, puisque l'expérience nous aura appris que (1) Paris est une ville unique, (2) la tour Eiffel est également unique et enfin (3) la tour Eiffel n'est visible qu'à Paris (en omettant les copies présentes à Tokyo et Las Vegas !).

Les travaux de [Khatib, 1986; Thrun, 1995; Lebeltel et al., 2004] ont permis de développer l'approche bayésienne standard pour la navigation robotique (autour de Thrun dont le livre "Probabilistic robotics" fait autorité [Thrun, 2002a]). D'autres travaux proposent une comparaison entre approche neuronale et probabiliste [Ma et al., 2006; Rao et al., 2002; Pfister et al., 2003]. Ces travaux visent à comprendre le fonctionnement d'un réseau neuronal en termes probabilistes. Les travaux de [Filliat and Meyer, 2002] évoquent les relations entre SLAM et construction de carte cognitive. D'autres travaux [Torralba, 1999; Ullah et al., 2008] utilisent une approche probabiliste pour la reconnaissance sémantique de lieux. Il n'y a cependant pas de lien direct entre cellules de lieu et lieu sémantique, ces derniers ne faisant pas partie du cadre de la localisation.

De manière générale, on dénote peu de travaux sur une modélisation neuronale des mécanismes d'inférences. La plupart des auteurs réalisent en réalité l'inverse en montrant que leur modèle neuronal possède bien des propriétés probabilistes afin de pouvoir appliquer la règle de Bayes.

3.3 Notre modèle de cellules de lieu d'un point de vue bayésien

Nous présentons ici deux modèles probabilistes manipulant les mêmes notions que le modèle neuronal. Tout deux utilisent les mêmes entrées et visent à fournir des sorties équivalentes. Les modèles probabilistes de localisation ou de SLAM partent de l'hypothèse d'existence d'amers et d'une mesure de position. Nous partons de l'existence d'amers (caractérisés dans le modèle par la conjonction d'une signature et d'un azimut) et de lieux.

Notons :

- $\{s^k, \theta^k\}_{k \in \{1..K\}}$: l'ensemble des vues locales extraites des images aux positions k . Ce sont des variables aléatoires continues.
- $s^{(1..K)}$: la collection de toutes les signatures de 1 à K et s^k : la signature k .
- $a_l^i = \langle w_l, s^i \rangle$: la valeur d'activation de la cellule l pour la vignette i du vecteur d'amers.
- $a_\theta^i = \langle \theta, \theta^i \rangle$: l'activation de la cellule θ (les crochets ne décrivant pas un produit scalaire mais simplement une distance).
- \mathbf{a}_Θ^i , \mathbf{a}_L^i et \mathbf{a}_Y^i représentent les vecteurs $\{a_\theta\}_{\theta \in [1..M]}$, $\{a_l\}_{l \in [1..N]}$ et $\{a_y\}_{y \in [1..N][1..M]}$.

De façon générale, nous utiliserons les indices pour désigner les variables aléatoires et les exposants pour désigner une observation. Les variables aléatoires seront représentées en majuscule, leurs valeurs en minuscule. On fera une exception pour les variables d'entrée où variables et valeurs seront confondues afin de simplifier les notations.

3.3.1 Description générale du problème

Les entrées du système sont organisées autour d'un vecteur d'amers (landmarks) et d'un vecteur d'azimuts qui se combinent au sein d'une carte bidimensionnelle dont les sorties sont utilisées pour activer un vecteur de cellules de lieu (voir fig. 3.2). Les variables aléatoires associées à ces différents vecteurs ou cartes sont respectivement L , A , Y et X . Dans la suite de cette section, nous aurons recours au terme anglais "landmark" pour désigner un amer visuel.

Les entrées sont constituées d'une collection de vignettes vues selon différents azimuts. Elles sont représentées par la collection de variables $s^{(1..K)}, \theta^{(1..K)}$.

Les variables L et A sont assimilées aux valeurs des indices des cartes de landmarks et d'azimuts. Ce sont donc des variables discrètes. Il est alors possible d'utiliser $P(L = l | s^k)$, la probabilité que la variable L prenne la valeur l lorsque l'on sélectionne la signature s^k , c'est à dire la probabilité que la cellule l soit sélectionnée (ou activée) par la signature s^k . Dans le cas de l'azimut, la variable A peut prendre des valeurs d'angle θ en présence de la signature k vue sous l'azimut θ^k . La probabilité qu'elle prenne une valeur θ donnée est $P(A = \theta | \theta^k)$.

Ces deux variables L et A se combinent en une carte bidimensionnelle Y correspondant au groupe $\Sigma\Pi$ (voir PrPh : 2.5.1). Y est ensuite utilisée pour produire une variable aléatoire X correspondant aux cellules de lieu. Ces variables sont également discrètes. Y désigne donc les cellules de la carte $\Sigma\Pi$ et ses valeurs les indices de cette carte.

Il est possible d'envisager deux approches du traitement des observations successives. La première conduit à **factoriser dans un premier temps** en fonction des observations puis de tenter d'évaluer les termes intermédiaires. La seconde reporte la **factorisation dans une étape ultérieure**.

3.3.2 Factorisation immédiate

Nous proposons de déterminer la probabilité $P(X = x | \{S = s^k, \Theta = \theta^k\}_{k \in [1..K]})$ d'être dans un lieu x , connaissant un ensemble de signatures s^k et leur azimut θ^k , que nous simplifierons par l'expression $P(x, \{s^k, \theta^k\}_{k \in [1..K]})$. Afin de calculer cette probabilité, nous utiliserons un estimateur MAP (Maximum A Posteriori) [Duda et al., 2001] :

$$x^* = \underset{x}{\operatorname{argmax}} P(x | \{s^k, \theta^k\}_{k \in [1..K]}) \cdot P(\{s^k, \theta^k\}_{k \in [1..K]}) \quad (3.3)$$

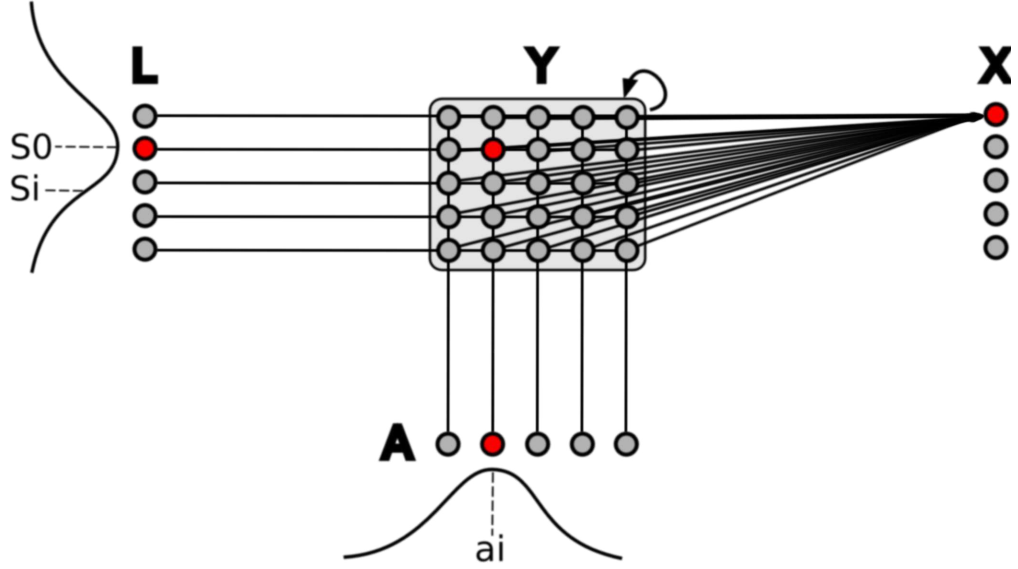


FIGURE 3.2 – Description du modèle de cellules de lieu. Le système prend comme entrées un vecteur d'amers (landmarks) **L** et un vecteur d'azimuts **A**. Les 2 vecteurs se combinent au sein d'une carte Sigma-Pi **Y** (PrPh) dont les sorties servent à activer un vecteur de cellules de lieu **X**. L'activité d'un neurone de landmark peut être interprétée comme le reflet de la probabilité $P(l^k/s^i)$ de percevoir le landmark l^k sachant la signature s^i . Cette probabilité est donnée par une mesure de distance entre la signature apprise s^0 et la signature perçue s^i . De même pour l'azimut θ^k de ce landmark, l'activité d'un neurone d'azimut reflète la probabilité $P(\theta^k/s^i)$ de percevoir l'azimut θ^k sachant la signature s^i . Les neurones de la carte **Y** peuvent être interprétée comme la probabilité $P(y/l^k, \theta^k)$ d'obtenir la conjonction du landmark l^k et de l'azimut a^k . Finalement l'activité d'une cellule de lieu est définie comme la probabilité $P(x/l^{1..K}, \theta^{1..K})$.

La question générale à résoudre est de calculer la probabilité $P(x|\{s^k, \theta^k\}_{k \in [1..K]})$. On considère dans un premier temps que les observations sont indépendantes. La règle de Bayes donne :

$$P(x|\{s^k, \theta^k\}_{k \in [1..K]}) = \frac{P(\{s^k, \theta^k\}_{k \in [1..K]}|x)}{P(\{s^k, \theta^k\}_{k \in [1..K]})} P(x) \quad (3.4)$$

Si l'on fait l'hypothèse d'indépendance statistique des données, nous pouvons écrire :

$$P(x|\{s^k, \theta^k\}_{k \in [1..K]}) = \prod_{k=1}^K \frac{P(s^k, \theta^k|x)}{P(s^k, \theta^k)} P(x) \quad (3.5)$$

Soit en appliquant à nouveau la règle de Bayes :

$$\begin{aligned} P(x|\{s^k, \theta^k\}_{k \in [1..K]}) &= \prod_{k=1}^K \left[\frac{P(x|s^k, \theta^k) \cdot P(s^k, \theta^k)}{P(s^k, \theta^k) \cdot P(x)} \right] P(x) \quad (3.6) \\ &= \prod_{k=1}^K \left[\frac{P(x|s^k, \theta^k)}{P(x)} \right] P(x) = \left[\frac{P(x)}{P(x)^K} \right] \prod_{k=1}^K P(x|s^k, \theta^k) = \left[\frac{1}{P(x)} \right]^{K-1} \prod_{k=1}^K P(x|s^k, \theta^k) \quad (3.7) \end{aligned}$$

On notera le terme $\ell [s^k, \theta^k]$ pour désigner la vraisemblance des observations :

$$\ell [s^k, \theta^k] = \prod_{k=1}^K \frac{P(s^k, \theta^k | x)}{P(s^k, \theta^k)} \quad (3.8)$$

Distributions intermédiaires :

Nous pouvons désormais nous intéresser aux termes du produit $\prod_{k=1}^K P(X = x | s^{(k)}, \theta^{(k)})$.

En introduisant pour chaque terme de ce produit les variables intermédiaires L_k et A_k :

$$P(x | s^k, \theta^k) = \sum_{l, \theta} P(x | L_k = l, A_k = \theta) P(L_k = l, A_k = \theta | s^k, \theta^k) \quad (3.9)$$

Dans ce modèle les variable L_k et A_k désignent les index des cartes de landmarks et d'azimuts. Moyennant l'indépendance de A_k et L_k :

$$= \sum_{l, \theta} P(x | L_k = l, A_k = \theta) P(L_k = l | s^k) P(A_k = \theta | \theta^k) \quad (3.10)$$

L'équation générale du modèle s'écrit alors (voir fig. 3.3) :

$$P(x | \{s^k, \theta^k\}_{k \in [1..K]}) = \left[\frac{1}{P(x)} \right]^{K-1} \prod_{k=1}^K \left[\sum_{l, \theta} P(x | L_k = l, A_k = \theta) P(L_k = l | s^k) P(A_k = \theta | \theta^k) \right] \quad (3.11)$$

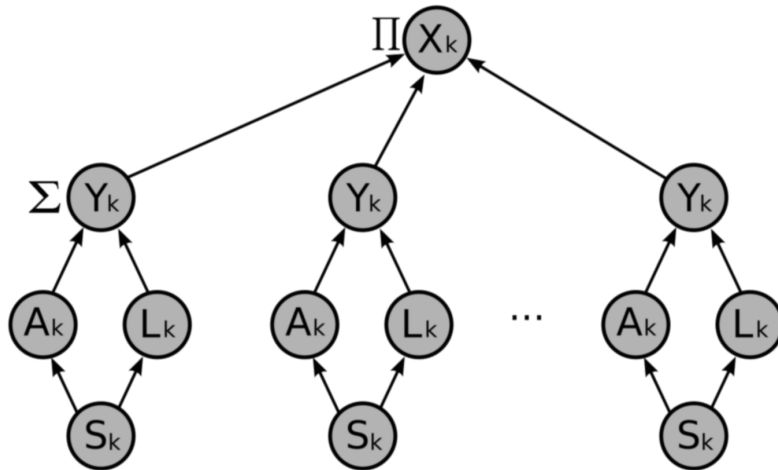


FIGURE 3.3 – Graphe de dépendance entre les différentes variables du modèle. la signature S_k d'un amer physique donne lieu à la mesure de son identité L_k et de son azimut A_k . Les variables intermédiaires Y_k de la carte Sigma-Pi dépendent de la co-activation (produit) du couple L_k et A_k . Un lieu X_k est défini comme la somme d'un ensemble de variables intermédiaires Y_k .

Nous allons considérer séparément les trois termes de l'équation 3.11.

Landmarks : Le fait de considérer L_k comme la variable aléatoire désignant l'index de la carte des landmarks conduit à la contrainte $\sum_l P(L_k = l | s^k) = 1$.

Il faudra donc utiliser une approche de type SoftMax¹ pour estimer cette probabilité soit :

$$P(L_k = l | s^k) = \frac{e^{w_l, s^k}}{\sum_j e^{w_j, s^j}} \quad (3.12)$$

Cette approche engendre une difficulté puisque le landmark associé à la signature courante appartient toujours à l'ensemble des landmarks considérés. L'approche ne fournit pas de méthode pour en créer de nouveaux.

Azimuts : Nous procéderons à une simplification ici en considérant que les azimuts sont distribués selon une loi gaussienne. En réalité, il s'agit d'une loi binomiale puisque les valeurs sont discrètes. Nous obtenons :

$$P(A_k = \theta | \theta^k, \sigma) = \mathcal{N}(\theta^k, \sigma) \quad (3.13)$$

Carte $\Sigma\Pi$ Elle correspond à l'évaluation du terme $P(x | L_k = l, A = \theta)$ de l'équation 3.9. Pour des raisons de simplification, nous introduisons une variable intermédiaire $Y_k \in \mathbb{N}^2$ dont les valeurs représentent les couples $(l, \theta) = y$ pour chaque observation. Il nous faut alors déterminer $P(Y_k = y | L_k = l, A_k = \theta)$ et $P(x | y)$. $P(Y_k = y | L_k = l, A_k = \theta)$ correspond à l'évaluation de la carte $\Sigma\Pi$ tandis que $P(x | y)$ correspond au calcul des cellules de lieux. Le calcul effectué par le réseau est déterministe : à un couple de neurones des cartes L et A correspond une unique unité active dans Y , de sorte que $P(Y_k = y | L_k = l, A_k = \theta) = 1$ pour $y = (l, \theta)$ et 0 sinon. Nous éliminerons donc cette étape de décomposition dans le modèle.

Cellules de lieux En partant de la supposition précédente, l'évaluation du terme $P(x | y)$ revient à évaluer $P(x | L_k = l, A_k = \theta)$ directement. Cette distribution est évaluée par apprentissage.

3.3.3 Factorisation retardée

Selon cette option, nous conservons la dépendance vis à vis de l'ensemble des observations et nous ferons apparaître les variables intermédiaires dans l'équation :

$$P(x, \{s^k, \theta^k\}_{k \in [1..K]})$$

Pour des raisons de simplification, on désigne par Ξ la variable aléatoire représentant le couple signature/azimut $\Xi = (S, \Theta)$.

$$P(x | \{\Xi_k\}_{k \in [1..K]}) = \sum_{l, \theta} P(X | \{A_k = l, \theta_k = \theta\}_{k \in [1..K]}) P(\{A_k = l, \theta_k = \theta\}_{k \in [1..K]} | \{\Xi_k\}_{k \in [1..K]}) \quad (3.14)$$

Cela nous conduit à évaluer les deux termes :

$$P(x | \{A_k = l, \theta_k = \theta\}_{k \in [1..K]}) \text{ et } P(\{A_k = l, \theta_k = \theta\}_{k \in [1..K]} | \{\Xi_k\}_{k \in [1..K]})$$

1. SoftMax est une variante normalisée du WTA : chaque entrée est pondérée de sorte que la somme de des activités soit égale à 1.

Nous ferons les mêmes hypothèses d'indépendance que précédemment
 Soit pour le **premier terme** :

$$P(x|\{A_k = l, \theta_k = \theta\}_{k \in [1..K]}) = \prod_{k=1}^K \frac{P(A_k = l, \theta_k = \theta_{k \in [1..K]}|x)}{P(A_k = l, \theta_k = \theta_{k \in [1..K]})} P(x) \quad (3.15)$$

$$= \frac{P(x)}{P(x)^K} \prod_{k=1}^K P(x|A_k = l, \theta_k = \theta_{k \in [1..K]}) = \frac{1}{P(x)^{K-1}} \prod_{k=1}^K P(x|A_k = l, \theta_k = \theta_{k \in [1..K]}) \quad (3.16)$$

Et pour le **second terme** :

$$P(\{A_k = l, \theta_k = \theta\}_{k \in [1..K]}|\{\Xi_k\}_{k \in [1..K]}) = \prod_{k=1}^K P(A_k = l, \theta_k = \theta|\Xi_k) \quad (3.17)$$

On parvient aux mêmes équations que par la factorisation directe.

Dans cette version, on ne prend pas en compte les activations sauf pour les entrées (landmarks et azimuts). Pour une observation donnée, il est possible de calculer la probabilité d'être dans un lieu. L'observation de deux couples amer/azimut seulement permet de déterminer le lieu dans lequel se trouve le système (aux erreurs près). Le modèle probabiliste intègre simplement les erreurs de mesure et d'appariement. Lorsque ce modèle réalise plusieurs observations, il effectue un produit des probabilités deux à deux pour l'ensemble des lieux (là où le modèle neuronal effectue un cumul).

3.4 Pondération statistique des amers

Dans cette section, nous allons présenter un modèle neuronal capable d'apprendre à estimer des densités de probabilité à partir de l'observation de l'évolution de variables. Ce modèle fait intervenir un simple conditionnement pavlovien. Nous montrerons deux applications du modèle dans le cadre d'expériences de navigation. La première expérience propose d'estimer la distribution des azimuts de chaque amer, dans un référentiel absolu, lors d'une navigation dans une salle et dans un couloir. L'expérience montre l'extraction par apprentissage des amers pertinents pour la tâche réalisée. La seconde expérience propose, de la même manière, d'estimer la distribution globale des azimuts de l'ensemble des amers, dans un référentiel relatif (distribution des écarts à l'azimut appris).

3.4.1 Un modèle neuronal de l'estimation de densité de probabilité

L'estimation d'une densité de probabilité à partir d'un ensemble donné d'observations représente un problème important qui peut être résolu par différentes méthodes. La méthode d'estimation de densité la plus ancienne et également la plus utilisée est la construction d'**histogramme**. On l'obtient en séparant la gamme des données en parts de tailles égales. Chaque nouvelle donnée vient s'accumuler dans un intégrateur représentant la part correspondante. L'histogramme fonctionne donc comme un compteur, intégrant chaque valeur différente prise par la variable considérée. Les valeurs sont normalisées en divisant chaque part par un facteur représentant le

temps de vie du système (un compteur global du nombre d'itérations réalisées depuis le début de l'expérience), de sorte que l'aire sous la courbe de l'histogramme soit toujours égale à 1.

Il est assez facile de modéliser un estimateur d'histogramme de manière neuronale par le simple ajout d'une carte neuronales avec des connections récurrentes en sortie d'une carte de Kohonen [Kohonen, 2001]. La carte de Kohonen nous donne un codage en population d'une valeur analogique (dans notre cas, l'azimut) de sorte que chaque état analogique, différent en entrée, n'active qu'un seul neurone de la carte. Une fois que les différentes classes de valeurs sont bien séparées, une seconde couche dotée de connections récurrentes réalise la somme des états issus de la carte de Kohonen (voir fig. 3.4). L'activité de cette couche représente l'histogramme des valeurs prises par la variable d'entrée. En appliquant le même principe pour chaque amer, puis en ajoutant une normalisation, nous obtenons une estimation de la fonction de densité de probabilité de l'azimut de chacun de ces amers. L'hypothèse est qu'après apprentissage, l'activité de reconnaissance de l'azimut reflète directement la probabilité de percevoir l'amer associé.

Ici, nous montrons qu'il est également possible d'estimer la fonction d'histogramme d'une manière contrôlée en utilisant la règle d'apprentissage de Widrow&Hoff, sans la moindre connexion récurrente. La carte de Kohonen représente l'entrée inconnue $S^d(n)$ que le système doit apprendre à prédire (la sortie désirée). Le système essaie donc d'adapter les poids w_n pour rendre la sortie $\hat{S}(n)$ aussi proche possible de $S^d(n)$:

$$e(n) = S^d(n) - \hat{S}(n) \quad (3.18)$$

$$\hat{S}(n) = w(n).l(n) \quad (3.19)$$

$$\frac{dw}{dn} = \epsilon.l(n).e(n) \quad (3.20)$$

Avec $e(n)$ l'erreur que le système tente de réduire, $l(n)$ l'amer (landmark) le mieux reconnu, ϵ le paramètre de pas de calcul contrôlant la vitesse de convergence.

Lorsque l'on tend vers un temps infini, les valeurs en sortie du système correspondent aux valeurs de l'histogramme, mais par contraste avec l'histogramme neuronal, l'estimation de densité est directement codée dans les poids synaptiques et non dans l'activité des neurones, ce qui semble plus plausible et permet de stocker l'apprentissage sur le long terme.

Cette méthode introduit 2 erreurs distinctes puisqu'elle cumule 2 estimations successives. La première estimation a lieu lorsque l'algorithme du LMS approxime les valeurs d'un histogramme réel. La seconde est que l'histogramme lui-même se trouve être une approximation de la loi de densité de probabilité réelle sous-jacente. Toutefois, la loi des grands nombres nous assure que l'histogramme tend à réduire son erreur d'estimation avec le temps, et puisque les valeurs du LMS convergent vers celles de l'histogramme, les 2 erreurs tendent à se minimiser.

3.4.2 Expérience : estimons la dispersion de l'azimut de chaque amer : P(A/S)

A partir de ce lien entre approche probabiliste et neuronale, nous proposons ici de mettre en application ce modèle dans une expérience réelle d'estimation de la dispersion de l'azimut de chaque amer. Dans cette expérience, le robot est libre de naviguer dans notre laboratoire pendant que le calcul en ligne de l'estimation de l'histogramme est réalisé. Le réseau tente d'approximer la loi de dispersion de l'azimut de chaque amer dans un référentiel absolu que nous fournit la boussole magnétique du robot.

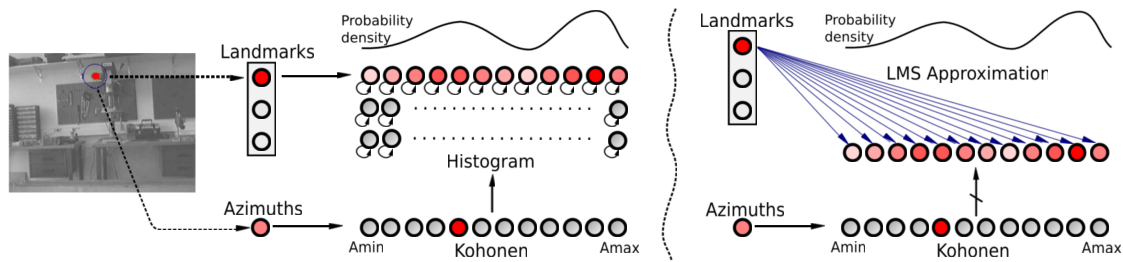


FIGURE 3.4 – Modèle neuronale d'estimation de densité de probabilité. **Gauche** : Estimateur neuronal d'histogramme obtenu par l'ajout de liens récurrents sur une carte de Kohonen. La carte de Kohonen permet un codage en population en projetant la valeur analogique de l'azimut sur un champ de neurone. Chaque neurone code pour un état de la variable analogique ($A_{min}=0$, $A_{max}=1$). Des liens récurrents permettent de compter le temps que la variable passe dans chaque état pour un amer (landmark) donné. Cette couche récurrente représente l'histogramme des azimuts perçus pour chaque amer. **Droite** : Cette même estimation peut être réalisée par un simple conditionnement entre amer et azimut (LMS). Chaque amer apprend à prédire l'état de l'azimut du champ de Kohonen. L'histogramme est cette fois encodé dans les poids synaptiques des liens (en bleu) et non dans l'activité des neurones.

Les résultats montrent plusieurs catégories de distribution (voir fig. 3.5) :

- Dans un premier cas, nous obtenons un ensemble d'amers dont la distribution est homogène et assez similaire à ce que donnerait une distribution uniforme. Ce résultat signifie que la signature correspondante désigne soit un amer redondant ou ambigu (e.g. un écran d'ordinateur), soit un amer mobile (e.g. une personne se déplaçant au cours de l'expérience), soit un amer proximal (e.g. une chaise autour de laquelle le robot peut tourner).
- Dans un second cas, certains amers saillants présentent des distributions étroites (dispersion azimutale faible), ce qui signifie que la signature correspondante n'est vue que pour une portion angulaire faible du panorama. Ce type de distribution étroite est obtenue pour des amers uniques et distaux (e.g. la lune).

De manière assez évidente, la forme d'une distribution est intimement liée au comportement du robot. Par exemple, un amer saillant présent au fond d'un couloir fournira une distribution azimutale étroite, les contraintes physiques du couloir imposant au robot de ne percevoir l'amer en question que sous une ouverture angulaire faible. Dans un couloir, le robot ne peut en effet pas tourner autour de l'amer. Ce type de distribution peut fournir une information utile que le robot peut afin d'apprendre, par lui-même, quels amers sont importants pour la réalisation de la tâche. La forme des distributions nous renseigne directement sur l'importance de tel ou tel amer dans la réalisation de cette tâche, à l'instar d'un mécanisme attentionnel. Ainsi, dans l'exemple du couloir, la saillance de la distribution de l'amer du fond indique au robot que cet amer est particulièrement pertinent dans la tâche de "**suivi de couloir**". À l'inverse, les amers présents sur les murs (poignées de porte, posters, interrupteurs) présentent une distribution d'azimuts assez uniforme et ne sont donc pas pertinents pour ce type de tâche.

Le système peut finalement apprendre à sélectionner les informations les plus pertinentes donnant la meilleure adéquation entre informations sensorielles et actions. Avant apprentissage, le système naviguerait dans le couloir sans à priori, c'est à dire en utilisant, de manière indifférenciée, l'ensemble des amers à sa disposition. Après apprentissage, il pourrait en référer

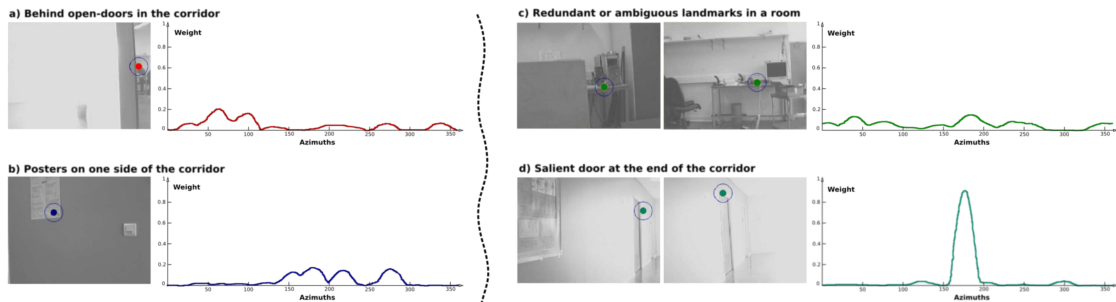


FIGURE 3.5 – Estimation de la densité de probabilité de l'azimut obtenue pour plusieurs amers. **A et B** - Certains amers ne sont pas représentatifs du lieu puisqu'ils sont perçus de manière uniforme dans le panorama (distribution d'azimut assez homogènes). **C** - Parfois, un amer peut être ambigu. La même cellule d'amer répond pour 2 objets physiques différents. **D** - Lors d'une navigation dans un couloir, un amer aligné avec le point de fuite présente une densité d'azimut très saillante et étroite. Cette information indique au système que l'amer en question est important pour la tâche de suivi de couloir.

presque uniquement à l'amer saillant en fond de couloir pour naviguer avec celui-ci. Le système serait ainsi capable d'inférer la reconnaissance d'un amer à partir de la perception de son azimut. L'avantage de ce mécanisme attentionnel est double : il permet d'améliorer la robustesse du comportement en réduisant l'information ambiguë et il pourrait également permettre de réduire le temps de calcul et/ou la mémoire nécessaire à la tâche en évitant de traiter l'information inutile dans une situation donnée.

3.4.3 Expérience : estimons la dispersion globale de l'azimut : $P(A)$

Dans notre modèle originel de cellules de lieu, chaque azimut perçu est projeté au sein d'un champ de 361 neurones puis convolué par un noyau gaussien d'une largeur arbitraire (50% de la taille du champ). Cette convolution permet une diffusion de l'activité centrée sur la direction privilégiée, ce qui donne au réseau des capacités de généralisation. En pratique, ce mécanisme permet de filtrer les mauvaises interprétations associées à des cellules de lieu non pertinentes dans le lieu courant. Si un amer est perçu sous un angle trop différent de l'angle sous lequel il a été appris, il peut s'agir (1) d'une erreur d'interprétation ou (2) d'un autre lieu, au delà d'un certain décalage. L'étendue de cette diffusion pourrait être apprise par le robot dans une tâche de navigation standard, en analysant au delà de quel décalage les amers ne sont plus suffisamment reconnaissables.

Le champ de neurones, utilisé pour le codage en population de l'azimut, est une version très simplifiée (sans lien récurrent, ni inhibition distale) des champs de neurones d'Amari. Le noyau d'interaction est gaussien et représente la distribution globale de l'azimut d'un amer perçu. Le choix d'un tel noyau implique l'hypothèse d'une distribution d'azimuts normale et cette hypothèse est supposée vraie peu importe l'amer traité. Cette distribution est figée et représente une connaissance "à priori" ou "innée" que possède le système sur l'environnement. Une telle méthode fait l'hypothèse que le monde au sein duquel évolue le robot suit cette loi normale.

Dans l'expérience suivante, nous montrons que notre réseau de neurones est capable d'apprendre cette distribution par l'observation. Cet apprentissage nous permet d'une part de vérifier que la loi de distribution définie a priori est valable et d'autre part de montrer qu'un agent est capable d'apprendre par lui même la loi qui reflète au mieux les observations issues de l'environnement

dans lequel il évolue.

Pour réaliser cet apprentissage, nous utiliserons une variante du modèle d'estimation de distribution utilisé précédemment : l'azimut n'est plus considéré dans un référentiel absolu mais dans un référentiel relatif, par calcul de l'écart entre l'azimut perçu et l'azimut appris. Afin d'obtenir une distribution relative globale, le système somme l'ensemble des distributions obtenues pour chaque amer au sein d'un champ global. La somme est réalisée au fur et à mesure des observations, en centrant les différentes distributions sur leur azimut appris respectif.

On notera que l'azimut appris lors de la première observation d'un amer n'est pas toujours un échantillon représentatif de la distribution des azimuts de cet amer. Il peut être décalé vis à vis de l'azimut réel, tout comme l'ensemble des azimuts perçus par la suite. Pour un amer particulier, il n'y a aucune garantie que l'azimut appris correspondant constitue une "vérité terrain". Néanmoins, l'opération de sommation des distributions obtenues pour l'ensemble des amers permet une réduction de l'erreur par lissage. En pratique, la somme des erreurs d'azimut appris pour chaque amer tend vers 0 si l'on considère un grand nombre d'amers.

Avant de mesurer la probabilité de dispersion globale lors d'une tâche de navigation, il convient de mesurer cette dispersion à l'arrêt afin de rendre compte de l'erreur statique que le système peut commettre. Dans cette première expérience, les roues du robot sont immobiles et la caméra effectue des panoramas successifs à 360°. Le système construit au fur et à mesure la distribution relative de l'azimut de chaque amer par rapport à son azimut appris, puis la somme de l'ensemble de ces distributions est mise à jour. La largeur de la distribution globale obtenue (voir fig. 3.6) nous donne un retour sur la précision maximale que le système peut atteindre.

Le résultat obtenu montre une dispersion étroite dont la largeur à mi-hauteur (hypothèse de distribution gaussienne) vaut un peu plus de 6°.

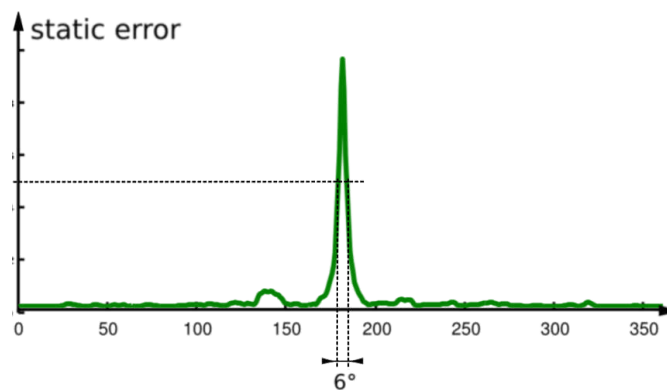


FIGURE 3.6 – Distribution globale des azimuts, obtenue lors d'une expérience statique (robot immobile). La largeur de la distribution reflète la précision maximale que le système peut atteindre. Cette dispersion statique provient de l'accumulation d'imprécisions dues aux défauts mécaniques du servomoteur, à l'erreur de la boussole magnétique, à la discrétisation de l'image et aux traitements logiciels.

Cette dispersion statique provient de l'accumulation d'imprécisions à différents niveaux de l'architecture. Ces imprécisions peuvent provenir de défauts mécaniques, de la discrétisation réalisée par l'acquisition d'image ou encore de traitements logiciels. Considérons l'imprécision éventuelle introduite par chaque membre de la chaîne :

- Boussole magnétique (CMPS03) : ± 3 à 4° .

- Servomoteur panoramique (Futaba S3003) : $\pm 0.9^\circ$ (200 positions de résolution angulaire pour encoder 180°).
- Discrétisation de l'image : $\pm 0.38^\circ$ (160 pixels de résolution en abscisse pour encoder 60° d'ouverture angulaire de l'optique).

On s'attend donc à une imprécision de $\pm 5^\circ$ autour de l'azimut appris, ce qui correspond à peu près à la mesure obtenue lors de l'expérience précédente.

Une fois l'erreur statique calculée, on réalise le même test lors d'une tâche de navigation dans une pièce. Dans cette expérience, nous comparons l'histogramme réel avec la réponse d'une batterie de 4 estimateurs neuronaux. Chaque estimateur possède sa propre vitesse d'apprentissage, de la plus lente à la plus rapide (respectivement 0.001, 0.01, 0.1 et 0.5).

Comme nous l'avons expliqué précédemment, la forme de la distribution est étroitement liée au comportement du robot pendant la tâche (voir fig. 3.7). Nous proposons de valider cette conjecture par l'observation en réalisant 2 tests pour lesquels le robot navigue en ligne droite (aller-retours sur une distance de 4m), et 2 tests pour lesquels il est libre de naviguer aléatoirement à l'intérieur d'un carré de 4x4m.

Sans surprise, les distributions obtenues sont plus étroites lorsque le robot navigue le long d'une ligne droite, et plus diffuses lorsqu'il navigue dans un espace bidimensionnel. On peut noter que la forme des distributions obtenues pour une navigation bidimensionnelle est finalement assez proche de la forme gaussienne choisie arbitrairement pour la diffusion des azimuts (50% de la taille du champ). Ce résultat montre que le système est capable d'apprendre la loi de dispersion azimutale par l'observation.

De manière évidente, la vitesse d'apprentissage d'un estimateur impacte grandement sur sa précision ainsi que sur le temps nécessaire à sa convergence. Ainsi une vitesse d'apprentissage faible induit une distribution proche de l'histogramme réel, au détriment d'une estimation plus lente. Et réciproquement, une vitesse d'apprentissage importante permet d'obtenir une estimation rapide de la loi de distribution, et donc une grande réactivité. En contrepartie, cette réactivité ne permet pas d'atteindre une bonne précision.

A titre de comparaison, le dilemme est semblable au principe d'incertitude formulé par Heisenberg [Heisenberg, 1927]. De manière plus générale, l'approche neuronale fournit des outils dynamiques et souvent réactifs, tandis que l'approche probabiliste se base sur des outils statistiques précis mais lents à converger sans l'utilisation de connaissances à priori. Ce constat amène à des questions concernant la façon dont les animaux pourraient résoudre un tel compromis entre précision et réactivité. Une hypothèse serait justement de considérer une batterie d'estimateurs fonctionnant en parallèle, dont les plus rapides permettraient de s'adapter à des événements soudains tandis que les plus lents permettraient une estimation statistique précise sur le long terme.

3.5 Une approche statistique de la reconnaissance de lieu

Dans la section précédente, nous avons pu voir comment un modèle neuronal utilisant un simple conditionnement entre amers et azimuts permettait d'apprendre à extraire les invariants

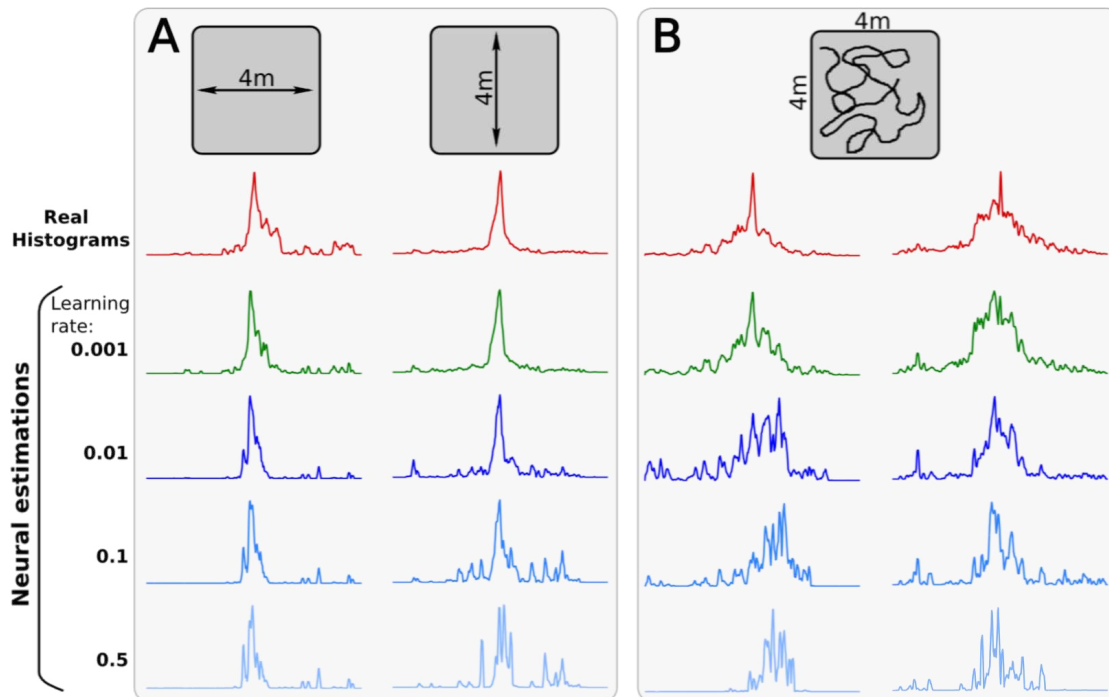


FIGURE 3.7 – Apprentissage de la loi de dispersion globale de l'azimut. Comparaison entre l'histogramme réel (rouge) et la réponse d'une batterie d'estimateurs neuronaux (vert et bleu) lors de 4 expériences. **A** - Dispersion obtenues lorsque le robot effectue des aller-retours en ligne droite sur une distance de 4m (cas mono-dimensionnel). Les trajectoires des 2 expériences sont orthogonales entre elles. Les distributions sont étroites. **B** - Dispersion obtenues lors de navigations aléatoires dans un carré de 4x4m (cas bidimensionnel). Les distributions sont plus larges et proches du noyau gaussien "a priori" utilisé dans le modèle original. On note que l'estimation est plus proche de l'histogramme réel pour une vitesse d'apprentissage faible (en vert), au détriment d'un temps de convergence plus grand. Une vitesse d'apprentissage plus grande (en bleu) permet d'obtenir une meilleure réactivité, au détriment d'une estimation moins précise.

statistiques de l'azimut d'un amer donné. Après apprentissage, le système pourrait inférer la reconnaissance d'un amer à partir de la perception de son azimut. Dans cette section, nous proposons d'utiliser ce même modèle afin d'apprendre à extraire les amers pertinents pour la reconnaissance d'un lieu. Après quoi, la perception d'un amer pertinent permettra d'inférer la reconnaissance du lieu associé.

3.5.1 Modèle de cellules de lieux probabilistes

Un simple conditionnement

Le modèle de cellules de lieu présenté au chapitre 2 propose un apprentissage instantané du motif présent sur le groupe PrPh (fusion amers-azimuts). Cet apprentissage de type "one-shot" recopie, dans les poids en amont du groupe de cellules de lieu, le motif présent dans l'activité du PrPh à la fin du panorama. A ce moment précis, l'activité d'un neurone du PrPh est de 1 pour chaque conjonction amer-azimut perçue et 0 partout ailleurs. Le motif appris est donc (1) binaire, puisque tout les amers ont le même poids, et (2) statique, puisque les poids ne sont plus mis à jour par la suite.

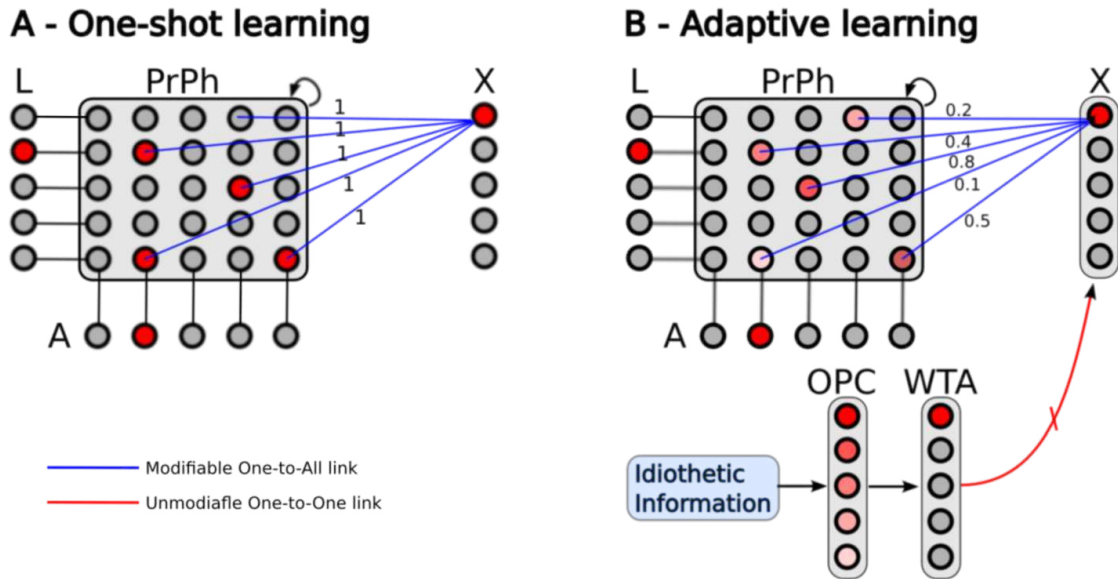


FIGURE 3.8 – Comparaison des 2 modèles neuronaux. **A** - Modèle d'apprentissage statique des lieux. Une cellule de lieu apprend "en un coup" (de type dirac) à reconnaître une constellation statique binaire d'amers-azimuts. **B** - Modèle d'apprentissage probabiliste. L'apprentissage n'est pas figé mais adaptatif au fil des observations. Le poids des couples amers-azimuts les plus stables se renforce tandis que celui des couples intermittents ou ambigus diminue. Ce second modèle extrait l'information statistiquement pertinente.

Ici, nous proposons de dépasser ces 2 limitations en utilisant un estimateur neuronal similaire à celui présenté précédemment (voir fig.3.8). Cette fois le groupe PrPh représente l'entrée conditionnelle utilisée pour apprendre à approximer le signal désiré (entrée inconditionnelle) provenant de cellules de lieu idiothétiques (odométriques). Nous reviendrons en détails sur la construction de ces cellules dans les chapitres 4 et 5. Pour le moment, considérons simplement que le système est capable de construire un code spatial, similaire à celui des cellules de lieu visuelles, mais basé uniquement sur des informations proprioceptives provenant d'un système d'intégration de chemin ou d'états particuliers de cellules de grilles. A l'instar de la carte de Kohonen fournissant un codage en population des valeurs d'azimuts, le groupe PrPh fournit un codage en population des conjonctions amer-azimut perçues. La modification des poids synaptiques des liens en provenance du PrPh permet d'estimer un histogramme des activités de ce groupe. Après apprentissage, les conjonctions amer-azimut les plus stables se voient attribuées un poids plus fort que les conjonctions épisodiques.

Un apprentissage stochastique basé sur la nouveauté

L'un des problèmes majeurs rencontrés lors de la mise en place de ce modèle a été la difficulté d'obtenir un apprentissage qui permette une discrimination équilibrée pour un ensemble de lieux. L'estimateur neuronal tel que décrit précédemment permet bien d'extraire l'information visuelle pertinente pour un lieu donné mais se heurte à un problème d'échantillonnage de l'apprentissage. Cet effet survient lors de l'utilisation d'un échantillonnage continu de l'apprentissage et introduit un biais favorisant les lieux dans lesquels le robot passe le plus de temps lors de la phase d'apprentissage (1). De plus, le dernier lieu appris est toujours mieux reconnu que

les autres (2).

Le premier phénomène s'explique par le fait que pour un apprentissage constant, la convergence d'une approximation (un lieu donné) dépend directement du temps passé à réaliser cette estimation. Obtenir une équiprobabilité de reconnaissance implique de partager le temps de manière égale entre les différents lieux à apprendre. Nous proposons de remédier à ce problème en modulant l'apprentissage du réseau en fonction de la **nouveauté** présente dans les entrées 3.9. Un détecteur de nouveauté rudimentaire peut être réalisé en apprenant à approximer la moyenne de l'activité de la cellule de lieu visuelle gagnante, puis de procéder au calcul d'erreur de prédiction par une simple différence entre l'activité courante et l'activité moyenne. Nous utilisons ici le même apprentissage par conditionnement entre la cellule visuelle gagnante (stimulus inconditionnel) et une entrée constante en guise de stimulus conditionnel (voir fig. 3.9). L'erreur de prédiction en sortie du système vient ensuite moduler la vitesse d'apprentissage du réseau. Ainsi, le système apprend uniquement lorsqu'il détecte un *delta* dans l'activité du lieu visuel gagnant et cesse d'apprendre lorsque l'activité reste constante (lorsque le robot stagne dans un lieu donné). Nous reviendrons en détail sur ce modèle de détection de nouveauté dans le chapitre 6. Dans un cadre différent de celui de la reconnaissance de lieu, Changeux et Dehaenne proposent un modèle neuronal similaire permettant d'adapter l'apprentissage en fonction de l'erreur de prédiction [Changeux et al., 2012]. Des solutions semblables ont été proposées par [Mumford, 1992], [Ballard et al., 1999] et [Seymour et al., 2005].

Le second phénomène est connu sous le nom d'*effet de récence* (recency bias) [Ebbinghaus

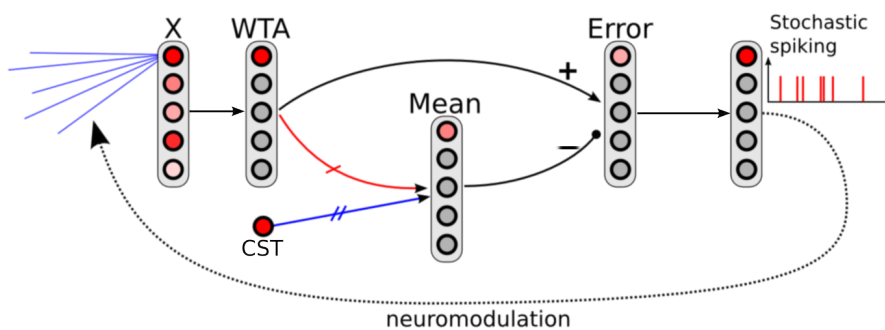


FIGURE 3.9 – Modèle d'apprentissage stochastique basé sur la nouveauté permettant d'assurer un échantillonnage non-biaisé de l'apprentissage. L'activité moyenne de la cellule de lieu gagnante (WTA) est apprise par le biais d'un conditionnement classique : le groupe de neurone "Mean" apprend à prédire le signal désiré "WTA" à partir d'une constante, à l'instar du modèle de la section 3.4.1. Une simple différence entre l'activité de la cellule gagnante et l'estimation de sa moyenne permet d'obtenir l'erreur de prédiction (Error). Cette erreur alimente un groupe de neurones déchargeants de manière aléatoire. Enfin, ces spikes stochastiques viennent neuromoduler l'apprentissage du lieu gagnant.

et al., 1913]. L'effet de récence désigne la facilité à rappeler les derniers éléments d'une liste de stimuli que l'on doit mémoriser. Cet effet introduit également un biais favorisant les stimuli appris en dernier. Ainsi, même si le mécanisme de neuromodulation basé sur la nouveauté permet de supprimer le biais temporel, le dernier lieu appris reste toujours mieux reconnu que les autres. Ce problème provient de l'aspect séquentiel de l'apprentissage des lieux, lié à un échantillonnage constant lors des phases de détection de nouveauté. Selon Dehaenne, il n'est possible de tirer une inférence correcte en statistique que si l'on a à faire à un échantillonnage aléatoire [Gweon et al., 2010].

[Postman and Phillips, 1964; Glanzer and Cunitz, 1966] ont présenté des listes de mots de différentes longueurs puis ont demandé à leurs sujets de rappeler les mots dont ils se souvenaient dans l'ordre qu'ils souhaitaient (rappel libre). Quand le rappel a lieu immédiatement, les premiers et les derniers éléments de la liste ont plus de chance d'être rappelés tandis que peu de sujets se souviennent du milieu de la liste. Quand le rappel a lieu après 15 ou 30 secondes, seuls les éléments du début de la liste ont une forte probabilité de rappel. Les auteurs interprètent ces expériences comme une preuve de l'existence d'une mémoire à court terme (short-term memory) responsable de l'effet de récence. Alors que le début de la liste serait rappelé de la mémoire à long terme (effet de primauté) et donc encore disponible après 30 secondes, la fin de la liste serait stockée en mémoire à court terme et disparaîtrait rapidement après la présentation.

Pour résoudre ce biais de récence dans la reconnaissance des lieux, nous proposons donc de raffiner le mécanisme de neuromodulation basé sur la nouveauté en ajoutant un processus stochastique en sortie (voir fig. 3.9). L'erreur de prédiction vient alimenter un groupe de neurones qui, au delà d'un certain seuil d'activation, se mettent à décharger par *spikes* de manière aléatoire. Ces *spikes* (signaux de type impulsionnel (Dirac)) viennent ensuite moduler l'apprentissage à intervalles irréguliers, permettant ainsi d'obtenir un échantillonnage aléatoire non-biaisé. Les *spikes* surviennent aléatoirement avec une moyenne d'un spike toute les 5 secondes, ce qui laisse le temps au robot de parcourir une certaine distance (20cm/s) entre chaque instant d'apprentissage et permet ainsi d'obtenir un échantillonnage non-biaisé. Il est également important que la vitesse d'apprentissage soit très faible (0.001 dans l'expérience suivante) afin de garantir un apprentissage statistique basé sur un grand nombre d'échantillons.

3.5.2 Expérience : estimation de la probabilité d'être dans un lieu sachant les observations

Dans cette expérience, nous mettons en application le modèle d'apprentissage adaptatif pour la reconnaissance visuelle de lieux lors de l'exploration d'une pièce. Afin de simplifier le problème, l'expérience se limite à plusieurs aller-retours de 4m en ligne droite dans une pièce du laboratoire 3.10. Le réseau de neurones recrute automatiquement une nouvelle cellule de lieu idiothétique (plus de détails au chapitre 4) lorsque l'activité de la cellule gagnante chute en dessous d'un seuil donné. Ce mécanisme assure un pavage régulier de l'espace. La valeur du seuil définit la résolution du pavage : un seuil élevé induit une fréquence spatiale de recrutement élevée. Le réseau tente d'approximer la cellule de lieu idiothétique gagnante en modifiant les poids en provenance du PrPh à chaque impulsion de neuromodulation qu'il reçoit. Dans cette expérience, le seuil de recrutement des cellules idiothétiques a été fixé à 0.7, ce qui implique un recrutement tous les mètres environ. Lors du premier aller (4m), le robot recrute donc 4 cellules de lieu.

Afin de procéder à une comparaison entre le modèle d'apprentissage statique originel et le nouveau modèle adaptatif, ces 2 systèmes fonctionnent en parallèle sans communication entre eux durant toute l'expérience. Les 2 modèles apprennent 4 cellules lors du premier passage, mais seuls les poids du modèle adaptatif sont mis à jour lors des autres passages. Les activités des cellules de lieu des 2 modèles ainsi que des cellules idiothétiques sont enregistrées simultanément. L'expérience s'achève après 9 trajets (4 aller-retours et 1 aller simple). Les résultats sont prometteurs puisque la dynamique d'activité des cellules adaptatives est bien plus élevée que

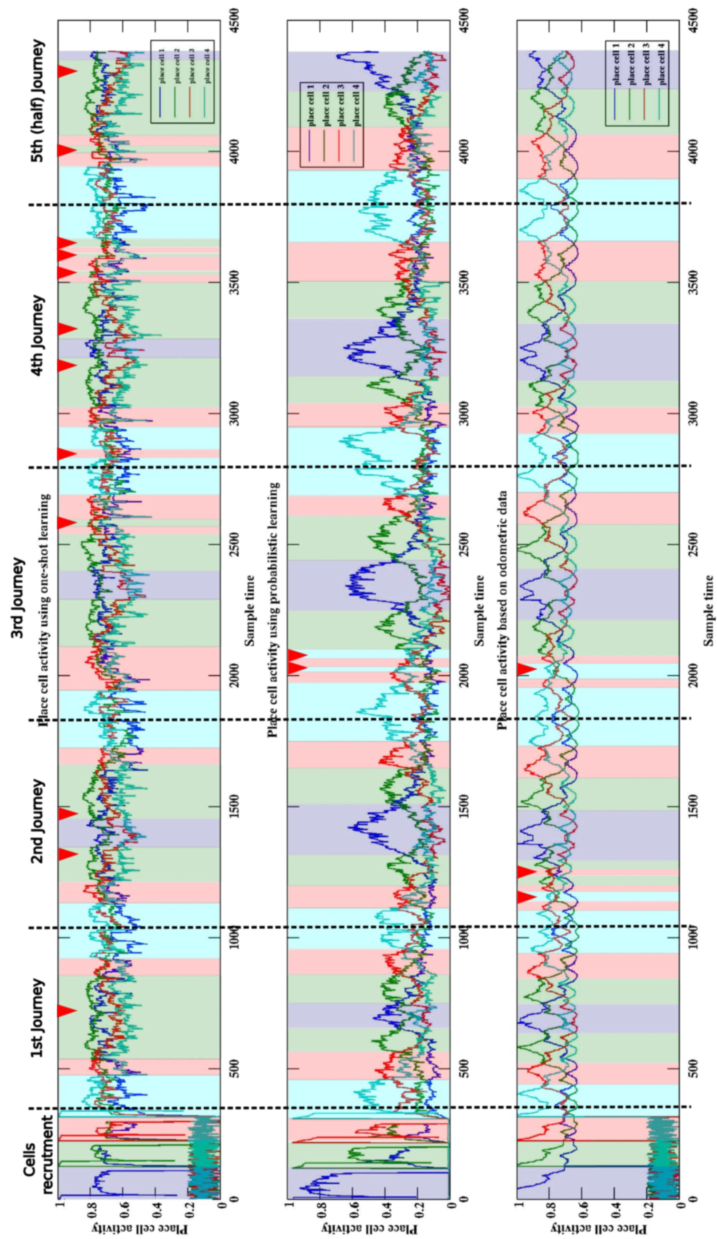


FIGURE 3.10 – Comparaison entre apprentissage statique et adaptatif lors de 4 aller-retours + 1 aller sur une trajectoire rectiligne. Les flèches rouges indiquent les erreurs de reconnaissances importantes. **Haut** - Activité de 4 cellules de lieux avec apprentissage statique. La dynamique est faible. On note beaucoup de faux positifs. **Milieu** - Activité de 4 cellules de lieux avec apprentissage adaptatif. La dynamique est plus grande et les champs de lieu sont mieux définis. Les lieux aux extrémités sont plus actif que ceux du centre. **Bas** - Activité de 4 cellules de lieux odométriques.

celle des cellules de lieu issues du modèle statique. Le nombre de fausses alarmes (ambiguïté dans la perception des lieux) est lui aussi plus faible pour le modèle adaptatif. On note tout de même un léger effet de récence qui transparaît toujours progressivement dans l'activité des cellules de lieux probabilistes. En effet, les lieux 1 et 4 sont plus actifs que les lieux 2 et 3.

3.6 Discussion

Dans ce chapitre nous avons présenté une formalisation de notre architecture de reconnaissance de lieu visuel selon un point de vue probabiliste. Les limitations constatées du modèle d'apprentissage "statique" nous ont poussé à adopter ce point de vue probabiliste afin d'aller vers une capacité de navigation autonome sur de longues périodes temporelles et au travers d'environnements fortement dynamiques. Nous avons montré qu'un simple apprentissage par conditionnement pouvait être utilisé à différents niveaux pour réaliser des estimations de densité de probabilité et ainsi agir comme un filtre attentionnel ne sélectionnant que l'information pertinente, que ce soit au niveau des amers visuels ou des cellules de lieu.

Les résultats obtenus montrent que le système est capable de procéder à un tri dans les amers perçus en estimant les distributions de leur azimut respectif pour une tâche donnée. Le système est capable d'une part, de découvrir par lui-même quelles amers sont pertinents pour la tâche qu'il réalise et d'autre part, d'apprendre des lois de distribution de l'azimut qui étaient autrefois définies a priori dans le modèle.

Lors du passage à l'échelle de l'estimateur pour l'apprentissage probabiliste de lieu, nous avons constaté 2 problèmes liés à l'échantillonnage de l'apprentissage : un biais temporel (les lieux dans lesquels le robot passe le plus de temps sont favorisés) et un effet de récence (le dernier lieu appris est toujours favorisé). L'utilisation d'un détecteur de nouveauté rudimentaire, basé sur le même type de conditionnement que l'estimateur, a permis au système de s'affranchir du biais temporel en évitant d'apprendre lors d'une stagnation dans un lieu. De plus, l'ajout d'un mécanisme stochastique d'apprentissage impulsif a permis de réduire l'effet de récence constaté dans l'apprentissage.

Les cellules de lieu issues d'un apprentissage adaptatif présentent une dynamique d'activité plus grande et moins d'ambiguïtés visuelles que celles issues d'un apprentissage statique. Des tests ultérieurs viseront à rendre compte de leur capacité à maintenir une activité correcte lors de modification importante de l'environnement et sur le long terme. Il sera également intéressant d'étudier leur impact sur le comportement du robot dans des tâches de navigation complexes, incluant un environnement dynamique (e.g. une foule de personnes) et/ou redondant.

Du point de vue développemental, nous avons montré qu'il était possible d'apprendre des lois sur le fonctionnement du monde physique et sur la dynamique des interactions liant l'agent et son environnement. Ces lois, construites de manière progressive par l'expérience, se formalisent sous l'aspect de distributions de probabilité encodées dans les poids synaptiques du réseau. Cette approche permet d'apprendre des lois qui restaient jusque là choisies arbitrairement. Les connaissances autrefois données au réseau de neurones sont désormais apprises par l'expérience et constitue un pas de plus vers l'autonomie développementale en robotique.

A un autre niveau, cette comparaison a permis de constater que les notions manipulées par nos architectures neuronales ne sont pas si éloignées des notions couramment utilisées par les théories bayésiennes. La comparaison devrait permettre d'améliorer le dialogue entre les communautés neuronales et probabilistes en débouchant à terme sur un formalisme unificateur.

Du point de vue de l'autonomie, on peut se poser la question de la facilité/difficulté de construire un modèle d'auto-évaluation en termes bayésiens. Les approches probabilistes nous fournissent le moyen d'évaluer une hypothèse du point de vue du sujet. L'approche bayésienne permet de formuler une conclusion $P(h | C)$ sur une hypothèse h à partir d'un ensemble de prémisses C . La mesure de la probabilité $P(h | C)$ représenterait alors l'opinion que peut avoir le système sur l'hypothèse h connaissant un ensemble d'indices C .

Du point de vue robustesse, l'apprentissage progressif d'une différenciation entre informations pertinentes et informations sans-intérêt pour la tâche a permis d'obtenir une discrimination des lieux plus stable face aux environnements dynamiques. Cette approche probabiliste est une première voie permettant de garantir la robustesse du système.

Dans les 2 chapitres suivants, nous allons présenter une autre voie pour l'amélioration de la robustesse : la multimodalité, c'est à dire l'utilisation de modalités de natures différentes pour la reconnaissance des lieux. Dans le chapitre 4, nous détaillerons l'architecture permettant la construction de cellules de lieu idiothétiques, introduites dans ce chapitre. Puis nous proposerons, dans le chapitre 5, un modèle de fusion d'information multimodale en se basant sur un type de conditionnement inspiré de celui présenté dans ce chapitre.

Localisation à partir d'informations idiothétiques

Dans le chapitre précédent, nous avons montré comment un système pouvait apprendre à sélectionner les informations visuelles pertinentes afin de se localiser de façon robuste et assurer une partie du passage à l'échelle de l'architecture de reconnaissance visuelle à de grands environnements. Néanmoins, il peut devenir crucial de recourir à un système de navigation basé sur des modalités de natures différentes pour faire face à des environnements très uniformes comme un couloir, une forêt dense, une large prairie ou simplement lorsque le robot est plongé dans le noir. Dans de telles situations, l'information visuelle n'est pas ou peu exploitable et d'autres modalités de nature différente, comme l'information idiothétique, peuvent devenir primordiales. De nombreux modèles tentent d'expliquer comment les animaux parviennent à intégrer leur chemin à partir de ces informations idiothétiques [Mittelstaedt, 2000; Hartmann and Wehner, 1995; Wittmann and Schwegler, 1995; Issa and Zhang, 2012].

Outre les mécanismes de retour au nid par intégration de chemin (voir chapitre 2, section 2.5.2) utilisés abondamment dans le règne animal, les mammifères semblent avoir recours également à des informations idiothétiques pour estimer leur position. Cette estimation pourrait être réalisée à l'aide d'un code spécifique fourni par les **cellules de grille** du cortex entorhinal, cellules que nous avons brièvement évoquées dans le chapitre 2. Tandis que les cellules de lieu déchargent à un endroit unique dans l'environnement, les cellules de grilles présentent des champs d'activité multiples correspondant aux vertex d'un maillage hexagonal virtuel.

Dans ce chapitre, nous étudions les propriétés de ces cellules en présentant de prime abord un modèle capable de rendre compte, au travers d'expériences réalisées sur robot réel, de plusieurs résultats neurobiologiques obtenus chez le rat. Les cellules de grille permettraient entre autres choses d'alimenter un système de localisation précis construit majoritairement à partir d'informations idiothétiques. L'hypothèse est qu'une localisation spatiale non-ambiguë requiert la conjonction des activités d'un ensemble de cellules de grille de périodes différentes. La résolution spatiale ainsi que la taille de l'espace utile codable dépend du choix des bons rapports entre les périodes des différentes grilles utilisées.

Notre modèle se base sur le système d'intégration de chemin présenté au chapitre 2. Ce mécanisme d'intégration de chemin est soumis au cumul d'erreurs inhérents à tout système dépendant uniquement d'informations idiothétiques. Nous proposons ici un mécanisme de calibration de ce mécanisme à partir d'informations allothétiques basées sur un simple conditionnement entre

intégration de chemin et cellules de lieu visuelles. Ce mécanisme permet de mettre à jour le champ d'intégration de chemin afin de limiter ce cumul d'erreurs.

Nous présenterons ensuite de nombreuses expériences sur robot réel permettant de mettre en avant l'impact des différents paramètres sur la structure des grilles obtenues. Puis, nous proposerons un modèle de cellules de lieu idiothétiques obtenues par l'apprentissage de conjonctions entre différentes activités de grille. Enfin, nous terminerons ce chapitre en discutant de la plausibilité de ce modèle et des paramètres permettant la génération de champs de lieu de différentes largeurs.

4.1 Les cellules de grilles

Des études éthologiques sur la navigation des animaux montrent qu'une grande variété de modalités sensorielles peuvent être utilisées pour naviguer et se localiser dans des environnements complexes et inconnus. La découverte des cellules de lieu (PC) dans la structure hippocampique du rat [O'Keefe and Nadel, 1978] a permis de mettre en évidence le codage d'une représentation neuronale de l'espace allocentrique utilisé par les rats, et plus généralement les mammifères (voir Chapitre 2, section 2.3.2). Plus tard, il a été montré que d'autres structures en lien avec l'hippocampe exhibaient des propriétés de représentation spatiale, ainsi que le cortex entorhinal (EC) dont les cellules présentent des champs de lieu larges et flous [Quirk et al., 1992; Sharp, 1999a]. Plus récemment, le cortex entorhinal a été le centre d'intérêt de nombreuses études. Des études anatomiques ont montrés que différents secteurs du cortex entorhinal se projettent à plusieurs niveaux dans l'hippocampe : la face dorsale d'EC se projette sur la partie dorsale de l'hippocampe, la face ventrale sur sa partie ventrale. Les neurones de CA1 (Corne d'ammon) présentent une activité spécifique du lieu, persistante même lors de la suppression de toute entrée intrahippocampique provenant de CA3, et ce même lorsque les champs de lieu sont moins précis que leurs homologues [Brun et al., 2002]. Par déduction, des signaux spatiaux doivent donc être calculés en amont de l'hippocampe. Puisque les entrées directes du cortex entorhinal sont de loin les plus importantes de l'hippocampe [Amaral and Witter, 1989; Dolorfo and D.G., 1998], il apparaît important de comprendre leurs propriétés de décharge spatiale. Ainsi, Fyhn et collègues [Fyhn et al., 2004] ont découvert des activités présentant de multiples champs étroits en enregistrant la réponse de cellules des couches II et III de EC. Ce qui a mené en 2005 à la découverte frappante des cellules de grille (grid cells) dans la bande dorso-latérale du cortex entorhinal médian (dMEC) [Hafting et al., 2005]. Lorsqu'elles sont enregistrées dans des environnements suffisamment larges, les cellules de grille présentent des activités spatialement périodiques, formant un pavage régulier de l'espace, à l'instar d'une grille couvrant l'environnement. Chaque cellule génère une grille à elle toute seule lorsque le rat explore l'environnement (voir fig. 4.1). Les différentes couches de dMEC contiennent toutes des cellules de grilles bien que leurs attributs diffèrent légèrement. La couche externe abrite uniquement des cellules de grilles pures, insensibles à l'orientation, tandis que les couches plus profondes présentent des cellules hybrides par une conjonction d'information de grille et de direction de la tête. Une cellule hybride présente une activité identique aux grilles habituelles mais ne décharge que lorsque l'animal fait face à une direction particulière.

Les cellules de grilles physiquement proches les unes des autres génèrent des grilles d'activité de même espacement et de même orientation mais leur champs de lieu sont décalés (phases différentes), tandis que des cellules lointaines varient en orientation et en espacement. Chaque grille

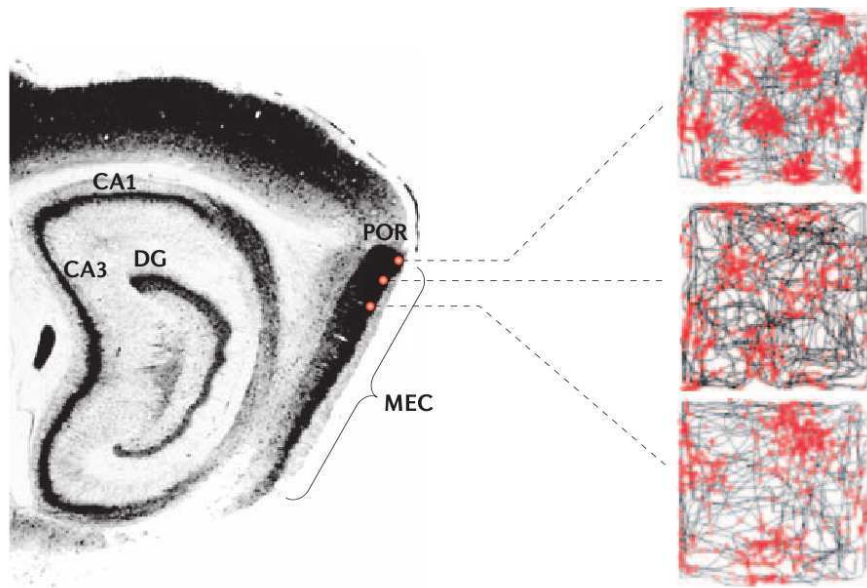


FIGURE 4.1 – Image extraite de [McNaughton et al., 2006]. Représentation d’une coupe transversale de la région hippocampique du rat et activités des cellules de grilles enregistrées lors d’une expérience dans un enclos carré. Le schéma montre les différentes structures de l’hippocampe, à savoir le gyrus dentelé (DG), la corne d’Ammon (CA1 à CA3) et le cortex enthorinal médian (MEC) où sont situées les cellules de grilles. **A gauche** : Les points rouges montrent l’emplacement des 3 cellules enregistrées lors de l’expérience. **A droite** : L’activité obtenue pour chaque cellule est une tessellation triangulaire projetée sur l’environnement, formant une grille dont l’orientation, l’espace et la phase sont caractéristiques. On remarque que l’espace de la grille obtenue dépend de la couche des cellules enregistrées dans MEC.

est donc définie par 3 paramètres : sa **période** (l’espace entre 2 pics d’activité), sa **phase** (le décalage en x et y relatif à un point de référence extérieur) et son **orientation** (l’angle de la grille par rapport à une référence allocentrique). L’activité des cellules de grille est présente même en l’absence de stimuli visuels. Les activités de grille persistent lors d’une navigation dans le noir, même si l’activité de grille tend à se dégrader avec le temps [Hafting et al., 2005]. La raison provient du fait que les cellules de grilles semblent prendre en compte les informations idiothétiques (internes à l’animal) disponibles. En l’absence de vision, l’animal n’a plus de référence externe et les pics d’activité deviennent de moins en moins étroits en raison du cumul d’erreurs lié au processus d’intégration de chemin. A l’inverse, lorsque des repères visuels sont accessibles, ils exercent un fort contrôle sur l’alignement des grilles : le fait de déplacer ces repères d’un angle donné a pour effet de pivoter les grilles du même angle [Hafting et al., 2005]. Les caractéristiques des grilles apparaissent à la première entrée de l’animal dans l’environnement et restent stables par la suite. Cependant, si les repères sont perçus comme étant mobiles par le rat, alors les grilles ne subissent pas de rotation et l’animal se réfère alors uniquement à des informations idiothétiques [Jeffery, 1998]. Cependant, le mécanisme par lequel l’intégration de chemin est influencée par des stimuli externes n’a encore pas été déterminé.

L’intérêt premier des cellules de grilles serait de coder l’emplacement de l’individu de manière précise sur de grandes distances puisque, malgré leurs activités individuelles périodiques, la conjonction de plusieurs grilles de caractéristiques différentes n’autorise qu’un seul emplacement dans l’environnement et permet ainsi de générer des cellules de lieu proprioceptives (voir fig.

4.2). Si des travaux de neurobiologie ont permis de mettre en évidence la nature multimodale du traitement opéré par le cortex entorhinal, les mécanismes régissant ces interactions entre modalités restent encore méconnus [Deneve S, 2004].

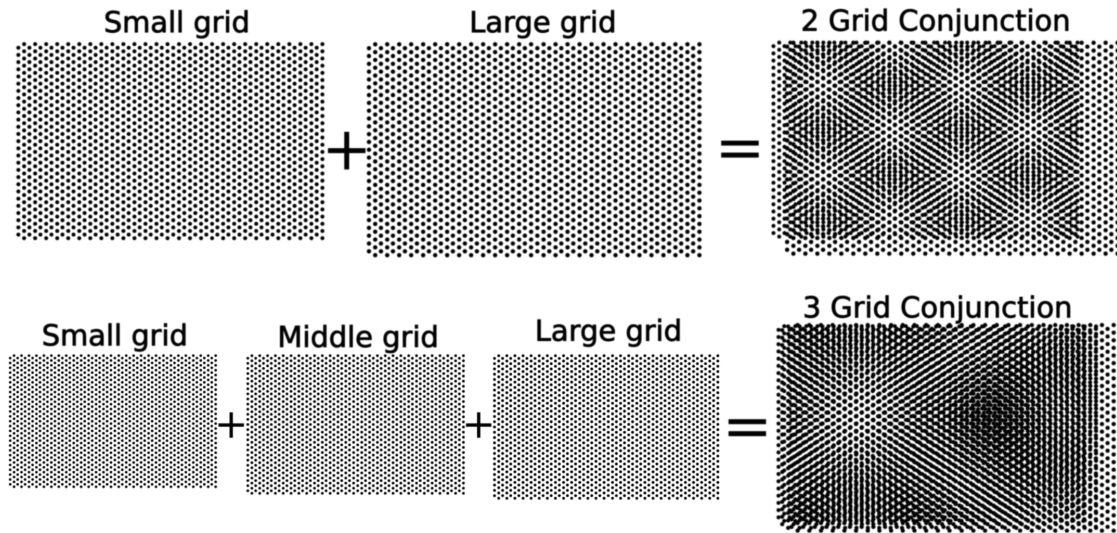


FIGURE 4.2 – Représentation intuitive du codage d'un lieu obtenu par la conjonction de cellules de grille. La combinaison de plusieurs grilles de périodicités spatiales différentes peut expliquer l'apparition de champs de lieu unique. **Haut** : La conjonction de 2 grilles de fréquences différentes fait émerger une structure de fréquence inférieure par effet de moiré (interférences spatiales). Des maxima locaux apparaissent de manière périodique lorsque les grilles se synchronisent. La fréquence résultante est égale à la différence des 2 fréquences fondamentales. **Bas** : La conjonction de 3 grilles de fréquences différentes permet de réduire la fréquence des synchronisations et ainsi présenter un champ de lieu unique au sein de l'espace considéré.

Trois classes de théories ont été proposées pour tenter d'expliquer la façon dont un réseau de neurones pourrait produire des formes d'activités spatiales hexagonales et régulières : (1) les modèles de cellules de grille à réseaux récurrents, basés sur la dynamique d'attracteurs continus [Fuhs and Touretzky, 2006; McNaughton et al., 2006], (2) les modèles d'interférence oscillatoire [Burgess N. Barry C., 2007] et enfin (3) les modèles utilisant un système de résidus numériques [Fiete et al., 2008; Gaussier et al., 2007].

Certains modèles récents expliquent l'activité des cellules de grille par le pliage d'une représentation cartésienne bidimensionnelle de l'environnement physique. Cette représentation est en fait celle d'un tore, ce qui permet de s'affranchir des problèmes d'effets de bords de la carte [McNaughton et al., 2006]. Cette hypothèse soulève plusieurs problèmes, comme par exemple la détermination de la taille du tore, ou encore la manière dont sont assemblées les cartes provenant de plusieurs environnements au sein d'une seule et unique carte cohérente.

Malgré le grand nombre de modèles théoriques proposés à ce jour, seule une infime partie [Milford and Wyeth, 2008] a été testée en environnements réels sur plate-forme robotique. On en sait assez peu sur les exigences nécessaires à la reproduction de l'activité hexagonale caractéristique dans une expérience réelle, ni le gain que pourrait procurer un tel système (facteur de compression de l'information et qualité du code généré), ni quelle est leur rôle réel dans le traitement de l'information. De récents travaux suggèrent même que ces cellules pourraient réaliser un processus général de réduction du bruit qui ne soit pas nécessairement lié à des traitements spatiaux [Sreenivasan and Fiete, 2011].

4.2 Un modèle plausible de cellules de grille

Dans cette section, nous nous proposons d'étudier une nouvelle architecture permettant la génération de cellules de grille. En suivant une approche constructive [Mataruna and Varela, 1980; Varela et al., 1993], nous présenterons, en premier lieu, une implémentation neuronale d'un modèle minimaliste capable d'exhiber des cellules dont l'activité est similaire à celle des cellules de grille biologiques. Ce modèle minimal se base sur un système de résidus numériques et à d'ors et déjà été testé en simulation [Gaussier et al., 2007]. Les résultats expérimentaux soulignent le rôle clé joué par les signaux visuels pour le maintien d'une activité de grille sur de longues durées. Nous montrerons qu'en l'absence d'indices visuels, l'activité de grille est amenée à se dégrader avec le temps.

Nous proposerons ensuite un mécanisme simple de recalibration de l'intégration de chemin à partir d'informations visuelles afin de maintenir une erreur suffisamment faible pour pouvoir conserver l'activité de grille. Plusieurs expériences seront présentées pour étudier l'impact des paramètres du modèle et l'effet des différentes sources d'erreur sur la forme de ces grilles.

Enfin, nous montrerons comment nous parvenons à obtenir une activité localisée, de type cellule de lieu, à partir d'un apprentissage hebbien et comment certains paramètres peuvent contrôler la taille du champ de lieu généré (et donc la capacité de généralisation du modèle).

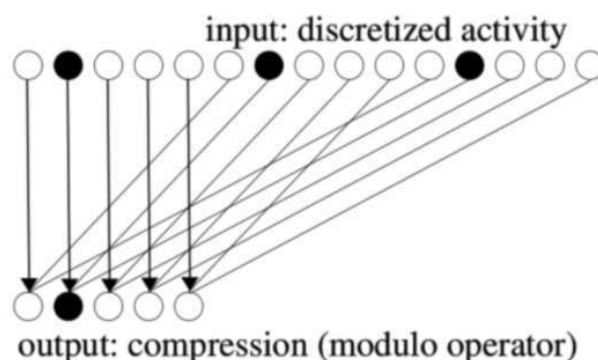


FIGURE 4.3 – Image extraite de [Gaussier et al., 2007]. Opération de modulo neuronal, réalisé par la projection d'une première couche de neurones (haut) sur une seconde couche (bas) contenant moins de neurones. Cette projection réalise une compression d'information avec pertes. Ainsi, l'activation de neurones espacés d'un facteur de compression dans la première couche n'active qu'un unique neurone dans la couche de sortie.

4.2.1 Description du modèle

A l'inverse de la plupart des modèles de cellules de grille existants, nous proposons un modèle d'intégration de chemin global indépendant des cellules de grille (externe à EC) ainsi qu'un modèle de cellules de grille résultant de la simple projection et fusion de l'activité d'intégration de chemin sur des neurones de dMEC (voir fig. 4.3).

Notre modèle d'intégration de chemin est directement inspiré de [Wittmann and Schwegler, 1995] et [Etienne, 1998]. La direction du robot est codée sur un champ de neurones uniforme et circulaire. Le champ couvre 360° de sorte que chaque neurone code pour une direction donnée θ_i . La précision angulaire est donnée par le rapport $360^\circ/N$ pour un champ de N neurones. Le neurone le plus actif de ce champ d'intégration de chemin correspond à la direction globale

du mouvement de l'animal depuis la dernière recalibration. Le niveau d'activité de ce neurone correspond à la distance cartésienne parcourue depuis le point de départ (ou de recalibration). Considérons maintenant le neurone du champ associé à la direction $\theta_i = 0$. Son activité est donnée par $\alpha \sum_{t=t_A}^{t_B} \cos(\Theta(t))$, avec $\Theta(t)$ la direction du mouvement au temps t . Si l'on considère maintenant le neurone associé à la direction $\theta_i = \frac{\pi}{2}$, son activité est donnée par :

$$\alpha \sum_{t=t_A}^{t_B} \cos(\Theta(t) + \frac{\pi}{2}) = \alpha \sum_{t=t_A}^{t_B} \sin(\Theta(t))$$

L'activité de ces 2 neurones correspond aux coordonnées (x, y) de l'agent au temps t dans le plan cartésien dont l'origine est donnée par le point de départ de la trajectoire. D'autres paires de neurones peuvent être utilisées pour exprimer le même déplacement à partir d'axes non-orthogonaux.

Notre modèle le plus minimaliste de cellules de grille est basé sur plusieurs opérations de modulo appliquées sur l'intégration de chemin (voir Fig.4.4). Ce modèle repose en grande partie sur l'hypothèse que la fonction générique des couches superficiels de EC est de compresser et fusionner des informations multimodales [Gaussier et al., 2007]. L'activité D_i d'un neurone appartenant au champ d'intégration de chemin (associé à la direction θ_i) est discrétisée sur un nouveau champ $E_j^i = \text{round}(\frac{D_i \cdot N_E}{D_{max}})$ où D_{max} est l'activité maximum calculable (c'est à dire la distance maximum encodable), N_E est le nombre de neurones de chaque champ utilisés pour discrétiser chaque activité analogique du champ d'intégration de chemin. Puis, une opération de modulo est appliquée pour compresser le champ E_j^i par projection sur un champ plus petit.

$$M_k^n = \begin{cases} 1 & \text{if } k = \text{Argmax}(E_j) \text{ mod } G^n \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

$$(4.2)$$

avec G^n la valeur du modulo utilisé pour générer la grille n . Les valeurs du facteur G^n représentent une distance dans une entité abstraite correspondant à $\frac{D_{max}}{N_E}$. L'espacement de la grille est déterminé par la valeur du modulo et le rayon du champ d'activité par le facteur de discrétisation. Cette opération peut être vue comme une compression de l'activité d'intégration de chemin avec perte d'information et peut être générée par des connections récurrentes régulières entre 2 couches de neurones.

Notre modèle se base sur 2 de ces projections modulo (correspondant à 2 neurones du champ d'intégration de chemin) pour générer 1 grille particulière. La différence absolue entre θ_1 et θ_2 détermine l'orientation de la grille, et le facteur de compression de projection (le rapport entre taille du groupe d'entrée sur taille du groupe de sortie) détermine la période de cette grille. Chaque neurone de la couche de sortie d'un modulo donné génère une grille de même orientation et même période mais possédant une phase différente. 3 facteurs de modulo différents sont suffisants pour couvrir un grand espace et la taille de la carte torique équivalente correspond au produit de ces facteurs (et non la somme). Ces 3 facteurs doivent être premiers entre eux pour supprimer toute redondance d'information.

Ainsi, ce modèle de cellules de grille se prête bien à la navigation robotique puisqu'il permet d'encoder une position sur de longues distances à partir d'un minimum d'information. Par exemple, 3 grilles de respectivement 4, 7 et 11 neurones peuvent générer un espace de $4 \cdot 7 \cdot 11 = 308$ positions à partir de seulement $4 + 7 + 11 = 22$ cellules.

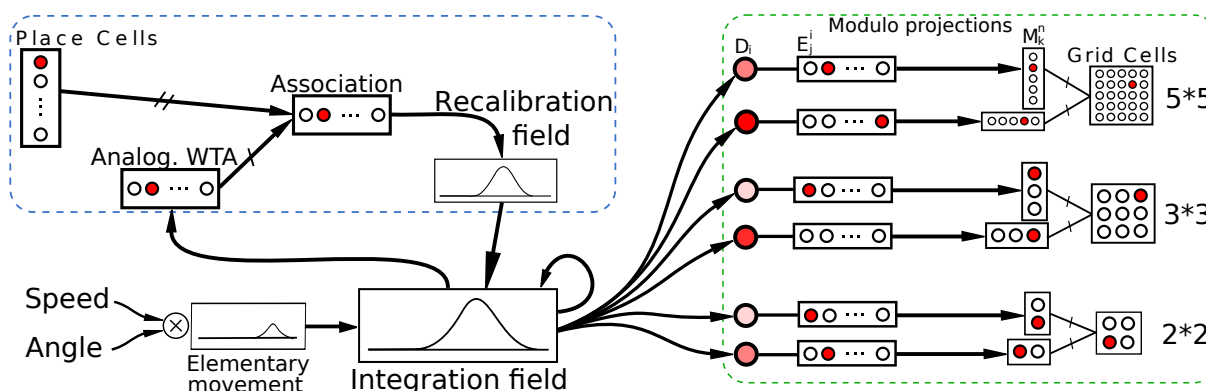


FIGURE 4.4 – Modèle de cellules de grilles. La vitesse linéaire et l’orientation absolue du robot sont utilisées pour définir un mouvement unitaire (en bas à gauche). Un champ d’intégration de chemin somme les mouvements unitaires du robot au fil du temps. Un mécanisme de recalibration basé sur l’information visuelle permet au système de limiter le cumul d’erreur de ce champ (en haut à gauche). Ce mécanisme apprend à associer un état du champ d’intégration avec la cellule de lieu gagnante. Cela permet au système de charger l’état mémorisé dans le champ lorsque la cellule de lieu correspondante gagne à nouveau. Le champ d’intégration de chemin est utilisé pour générer des cellules de grille sans besoin d’une carte cartésienne. L’activité (D) d’une paire de neurones du champ, choisie aléatoirement sur d’autres champs (E à droite). Ces champs sont ensuite compressés par projection modulo sur des champs plus petits (M). La conjonction de 2 de ces champs M est suffisante pour obtenir une matrice de cellules de grille.

Cependant, ce premier modèle n’est pas suffisant pour générer des activités de type grille à partir d’information idiothétique uniquement puisqu’il est soumis à la dérive classique de l’intégration de chemin. Les premiers résultats montrent que l’activité de grille se dégrade progressivement en se diffusant sur les régions voisines, si bien que l’activité globale apparaît comme aléatoirement distribuée sur tout l’espace après un certain temps [Gaussier et al., 2007]. Une solution à ce problème consiste en l’ajout d’un mécanisme de recalibration du champ d’intégration de chemin se basant sur des indices externes (visuels) et donc non-soumis à la dérive.

Il est possible de réaliser cela en apprenant à associer un état d’activité du champ d’intégration de chemin à la cellule de lieu gagnante, à l’instar d’un conditionnement classique. A partir de là, si la cellule de lieu correspondante est de nouveau gagnante, elle déclenche la recalibration. Un tel mécanisme est utilisé par [Strösslín et al., 2005; Fuhs and Touretzky, 2006; Arleo and Gerstner, 2000; Herrmann et al., 1999; Mataric, 1990]. Dans les expériences suivantes, l’apprentissage de l’association entre l’activité du champ d’intégration de chemin et la cellule de lieu gagnante est réalisé en un coup. Cette méthode peut induire de petites imprécisions lors de la recalibration si la cellule gagnante possède un champ de lieu large et/ou si le robot pénètre dans la cellule de lieu par une orientation différente de celle réalisée lors de l’association. Il serait intéressant de remplacer cette association immédiate par un apprentissage plus lent (<1). De cette façon il serait possible d’apprendre une moyenne des vecteurs proposés par le champ lorsque la cellule devient gagnante et donc d’obtenir un système de calibration plus précis.

Le système de reconnaissance de la cellule de lieu gagnante pour la recalibration dépend de 2 seuils : un seuil sur le niveau absolu de l’activité de la cellule (ce seuil vaut 0.9 dans nos expériences) et un seuil relatif (lorsque la différence entre le niveau de la cellule gagnante et le niveau

moyen de l'ensemble des cellules de lieu visuelles est supérieur à 0.4). Cette recalibration consiste simplement en un remplacement de l'activité du champ d'intégration de chemin par celle apprise lors du conditionnement. Ainsi, on autorise le robot à calibrer sa position idiothétique à l'aide d'information allothétique.

4.2.2 Structure des expériences

Toutes les expériences présentées dans ce chapitre ont été réalisées sur la plate-forme robotique Robulab de Robosoft (40cm de large pour 1m de haut), exceptée l'expérience n°3 qui a été réalisée en simulation. Chaque expérience (sauf expérience n°3) a été réalisée dans un enclos hexagonal comparable aux expériences menées sur les rats [Hafting et al., 2005]. Typiquement, les expériences réalisées sur les rongeurs consistent en l'enregistrement de l'activité des cellules de grille de dMEC pendant que l'animal (long de 20cm) explore un enclos circulaire (de 2m de diamètre). Dans le but de réaliser nos expériences dans des conditions relativement similaires, notre robot, large de 40cm, navigue aléatoirement dans un parc hexagonale de 4m de diamètre (voir Fig. 4.5)). La position du robot ainsi que l'activité des cellules de grille simulées sont enregistrés simultanément pendant l'expérience. La stratégie réflexe d'évitement d'obstacle est prioritaire et permet au robot de rester à l'intérieur de son enclos.

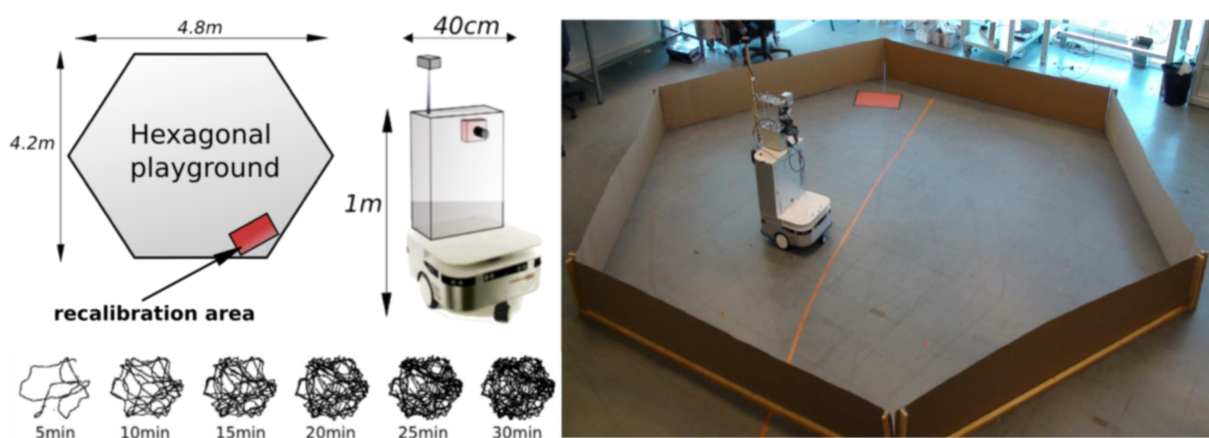


FIGURE 4.5 – Setup expérimental : le robot, large de 40cm, navigue librement dans un enclos hexagonale de 4m de diamètre. Le réflexe d'évitement d'obstacle le force à rester à l'intérieur de l'enclos. Le robot apprend à retourner à un point de recalibration (marqué d'un rectangle rouge au sol) à l'aide de quelques associations Lieu/Action. Un compteur interne déclenche un comportement de retour au nid toutes les X minutes (X dépendant de l'expérience), poussant le robot à revenir à ce point de recalibration afin de limiter le cumul d'erreur de l'odométrie.

4.3 Résultats

4.3.1 Expérience 1 - exploration aléatoire, sans recalibration

Dans une première expérience, le comportement exploratoire du robot est uniquement basé sur des mouvements aléatoires et le mécanisme de calibration n'est pas utilisé (voir Fig. 4.6 - A). Ce test permet de mettre en évidence l'effet du cumul d'erreur de l'intégration de chemin sur l'activité de grille obtenue. Il permet de souligner le besoin d'un mécanisme basé sur une

modalité externe pour corriger cette erreur et ainsi maintenir une activité de grille stable au cours du temps.

On observe une dérive rapide de l'activité de grille due à l'erreur cumulative de l'intégration de chemin. Cette erreur provient en fait d'une imprécision introduite par la discrétisation angulaire du champ, mais également de l'erreur angulaire commise par la boussole magnétique et les glissements du robot sur le sol. Chaque cellule présente donc une activité qui semble aléatoire et exempte de toute tessellation (pavage régulier de l'espace). Pourtant, elles présentent une activité périodique et plus cohérente lorsque l'on ne considère que des portions de 5min de l'expérience. Ainsi, un mécanisme basé sur une modalité externe est nécessaire pour limiter cette dérive. L'expérience suivante propose donc un mécanisme de recalibration simple à partir des cellules de lieu visuelles.

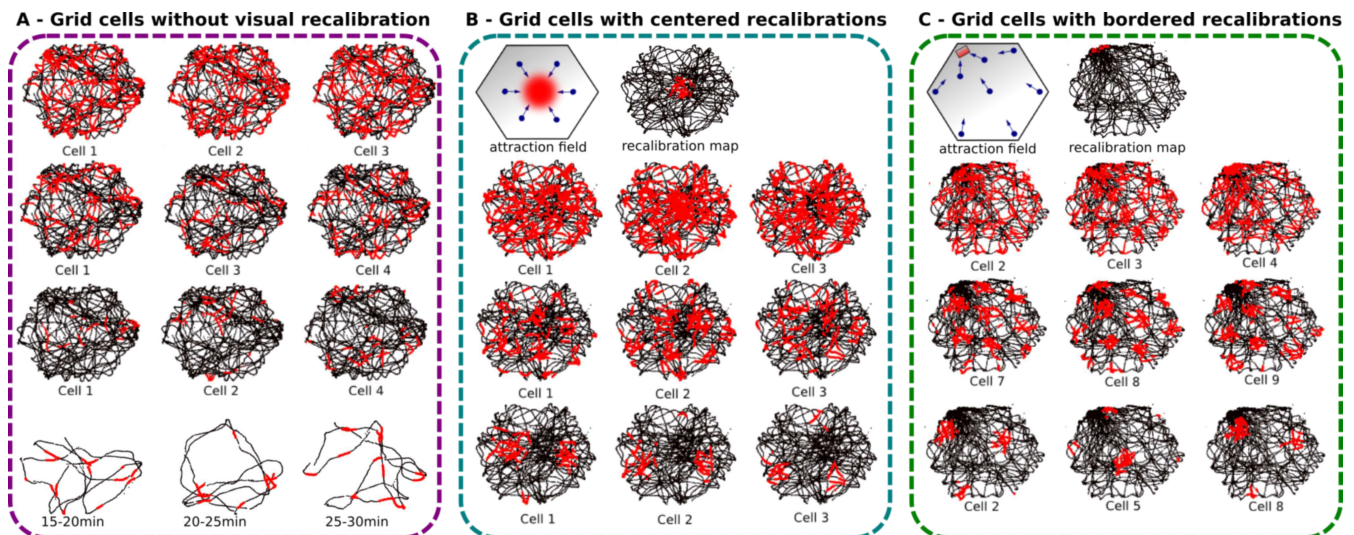


FIGURE 4.6 – Influence du lieu de recalibration sur la précision de l'activité de grille. L'activité des cellules de grille (en rouge) est superposée à la trajectoire du robot (en noir). **A** - En haut : Résultats pour 30min de navigation aléatoire sans recalibration. Chaque ligne correspond à un modulo particulier (projection de 60 neurones sur respectivement 4 (haut), 9 (milieu), 25 (bas)). Chaque colonne correspond à 3 cellules choisies aléatoirement (une pour chaque modulo). L'activité de grille est complètement brouillée à cause du cumul d'erreur de l'intégration de chemin sur une période de 30min. En bas : Des portions de 5min de l'expérience laissent pourtant deviner une activité périodique et regroupée sous forme de grappes. **B** - Même expérience, mais une zone de recalibration est cette fois ajoutée au centre de l'enclos. Un comportement de retour au nid, déclenché chaque minute, pousse le robot à se recalibrer périodiquement pour limiter le cumul d'erreur. **Haut** : Un champ d'attraction est appris autour de la zone de calibration. **Gauche** : 6 associations lieux actions permettent de converger vers le but. **Droite** : La carte des recalibrations montre une zone assez large. **Bas** : Des tâches d'activité apparaissent de manière périodique. L'activité de grille obtenue est plus cohérente cette fois. Son imprécision est directement liée à la taille de la zone de recalibration. **C** - Cette fois la zone de calibration est définie dans un coin de l'enclos. Les capteurs ultrasons sont utilisés comme modalité "tactile" pour augmenter la précision. La carte des recalibrations montre une zone beaucoup plus petite. L'activité de grille est mieux définie.

4.3.2 Expérience 2 - recalibration de l'intégration de chemin à l'aide d'un retour au nid périodique.

Dans les expériences suivantes, l'erreur de l'intégration de chemin est limitée par une recalibration périodique à un endroit donné de l'enclos. En premier lieu, un mécanisme de conditionnement apprend à lier une cellule de lieu particulière à un état du champ d'intégration de chemin. Après quoi l'activation future de la cellule de lieu concernée force le champ d'intégration de chemin à prendre la valeur de l'état appris. Le robot navigue à l'aide d'un comportement d'exploration aléatoire comme précédemment, mais un comportement de retour au nid (homing) est utilisé pour forcer le robot à retourner au point de recalibration au bout d'un temps donné. Un bassin d'attraction est appris autour de ce point de recalibration à l'aide de quelques associations Lieu/Action, ce qui permet au robot de converger naturellement vers ce point. Un besoin primaire déclenche périodiquement le comportement de retour au nid. A l'instar du besoin de manger, ce comportement ne cesse que lorsque le besoin primaire de retour au point de recalibration est satisfait. Après quoi le robot reprend son comportement d'exploration aléatoire.

Expériences montrant l'impact du lieu de calibration

Nous étudions ici l'impact du lieu de recalibration à l'intérieur de l'enclos en comparant les activités de grilles obtenues pour deux de ces lieux. Puisque le champ de réponse d'une cellule de lieu est assez large en comparaison de la taille de l'enclos, nous avons également utilisé d'autres informations accessibles au robot afin de réduire la zone de recalibration. Chaque expérience dure 30 minutes et le comportement de retour au nid est déclenché chaque minute. Un papier rouge est collé au sol pour montrer le lieu de recalibration, mais il n'est pas utilisé par le robot pour reconnaître ce lieu.

Dans la première expérience, le lieu de recalibration est situé au centre du parc, et un champ d'attraction est appris tout autour pour permettre au robot de converger vers ce lieu (voir Fig. 4.6 - B).

La seconde expérience est identique à la précédente mais le lieu de recalibration est cette fois défini dans un coin de l'arène formé par 2 murs (voir Fig. 4.6 - C). Un champ d'attraction est appris pour permettre au robot de converger vers ce lieu. Notons que dans cette expérience, le profil d'activité des capteurs ultrasons dans le lieu de recalibration est appris au même titre que la cellule de lieu visuelle. Cette information supplémentaire permet de restreindre la zone de recalibration à un lieu plus précis (recalibration = lieu visuelle + contact tactile).

La recalibration empêche l'accumulation d'erreurs au delà d'une minute, mais elle introduit une erreur statique directement liée à la taille de la zone de calibration. La précision de l'activité de grille dépend en grande partie de la taille de cette zone de calibration. Intuitivement, la grille n'est visible que si sa période est supérieure à la taille de la zone de calibration. La première expérience, avec calibration au centre de l'enclos, montre une zone de calibration assez large. La zone représente environ 20% de la taille du parc.

La seconde expérience tente de résoudre ce problème de précision en utilisant la localisation précise des murs de l'enclos. La calibration est réalisée dans le coin formé par 2 murs contigus formant une singularité naturelle de l'environnement vers laquelle il est facile de converger. Cette technique utilise la conjonction de l'activité des capteurs ultrasons (stratégie simili-tactile) et la reconnaissance de lieu. La zone de calibration est donc sensiblement plus précise et le robot est largement aidé par la présence des murs pendant la phase de retour au nid. Les résultats mon-

trent cette fois des activités de grille plus précises pour les différents facteurs de modulo. Les grilles partagent la même période pour le même modulo et la même orientation pour le même facteur de discrétisation. Néanmoins, au sein d'un groupe de cellules partageant le même modulo et la même orientation, chaque cellule produit une grille de phase différente.

Expérience montrant l'impact de la fréquence de calibration.

Dans ces expériences, nous testons l'impact du temps écoulé entre chaque recalibration sur l'activité de grille (voir Fig. 4.7). Le robot explore l'arène pendant 30min. La vitesse linéaire du robot est de 20cm/s et l'arène fait 4m de diamètre, ce qui signifie que le robot à besoin de 20 secondes pour parcourir l'environnement de part en part. Ces 20 secondes nous donnent la limite basse du temps de recalibration. A l'inverse, les expériences précédentes ont montré que l'activité de grille était complètement brouillée si la recalibration n'a lieu que toutes les 5 minutes. Ce temps nous donne une estimation de la limite haute de la période de calibration, à ne pas dépasser si l'on désire obtenir une activité de grille exploitable. Nous avons donc testé successivement les temps de recalibrations suivants : 30sec, 1min, 2min, 4min.

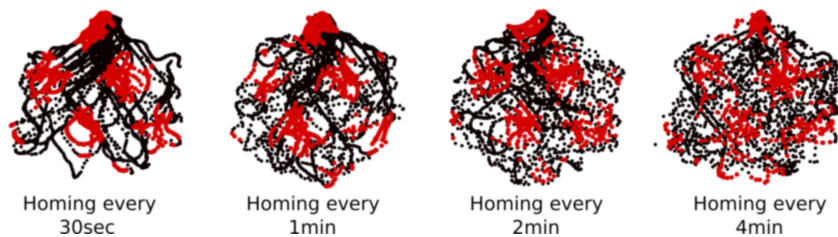


FIGURE 4.7 – Influence de la fréquence des calibrations sur l'activité de grille. L'activité d'une cellule de grille (en rouge) est superposée à la trajectoire du robot (en noir) pour 4 périodes de recalibration différentes. De gauche à droite : le robot déclenche un comportement de retour au nid toutes les 30sec, 1min, 2min et 4min. Plus la période entre chaque calibration est courte, plus l'activité de grille est précise. Mais le robot à besoin d'une période d'au moins 1min afin de couvrir l'environnement uniformément.

La fréquence de calibration nécessaire dépend de la précision du système d'intégration de chemin et de la taille de l'environnement. Les résultats montrent qu'une recalibration toute les 30sec permet d'obtenir une activité de grille précise mais l'environnement n'est alors pas couvert uniformément. Puisque le temps d'exploration est très court, le robot restreint son exploration à un espace proche de la zone de calibration. A l'inverse, l'environnement est uniformément couvert pour une calibration toute les 4min, mais l'activité de grille commence à se détériorer. Dans les expériences suivantes, nous avons choisi de déclencher un comportement de retour au nid toutes les minutes. Cela s'avère être un bon compromis entre précision et surface couverte.

Expérience d'exploration aléatoire, sans calibration.

De même que pour les expériences précédentes, les premières 30 minutes de l'expérience suivante permettent au robot d'explorer l'enclos et de retourner au nid se recalibrer chaque minute. Au delà de cette période de 30min, les informations visuelles ainsi que le comportement de recalibration sont volontairement désactivés. Le robot continue à naviguer aléatoirement pendant une nouvelle session de 30min (voir Fig. 4.8). L'activité de grille est mesurée à

différents moments, à savoir au bout de 33, 36, 39, 42, 45, 48, 51, 54, 57 minutes puis finalement 1 heure. Cette expérience diffère de l'expérience 1 par le fait que le robot ne navigue aléatoirement qu'après une exploration complète de l'environnement et l'établissement d'une activité de grille stable. Les résultats montrent la dégradation de la grille sur 30min de mouvements aléatoires sans recalibration. La carte des recalibrations montre une petite zone bien définie dans un coin pour les 30 premières minutes. la dégradation apparaît progressivement comme un bruit gaussien en fonction du temps. La grille est difficilement distinguable au delà de 15min.

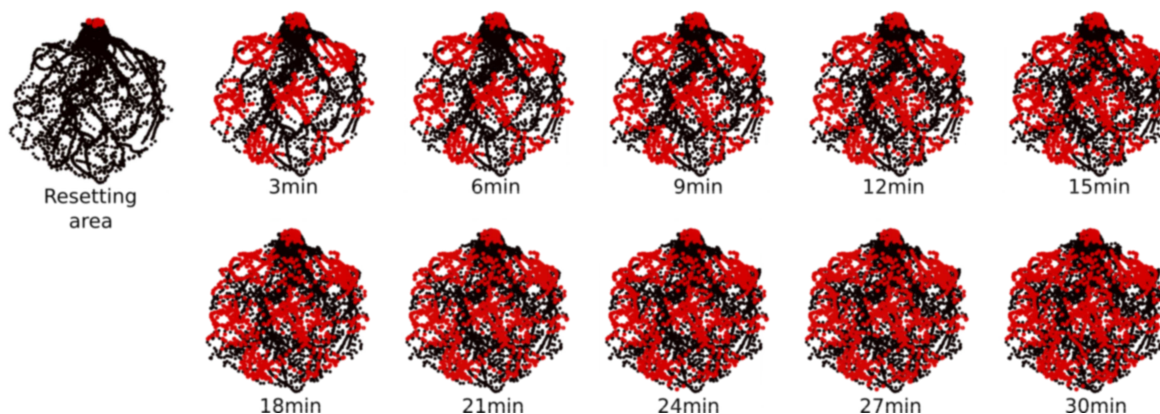


FIGURE 4.8 – Influence du cumul d'erreur sur l'activité de grille au cours du temps sans recalibration. Comme pour les expériences précédentes, le robot explore l'environnement pendant 30min, se recalibrant toutes les minutes dans un coin de l'enclos. Après 30min, la stratégie de retour au nid est inhibée et le robot continue son exploration pendant à nouveau 30min sans recalibration. L'activité de grille se dégrade alors progressivement en se diffusant dans le voisinage comme sous l'effet d'un bruit gaussien.

Expérience de calibration en aveugle.

Comme précédemment, le robot explore l'environnement pendant 30min. Pendant cette phase, le robot peut se recalibrer grâce au comportement de retour au nid mais la cellule de lieu codant pour le lieu de recalibration est construite à l'aide d'un simple conditionnement entre l'information visuelle et l'activité des cellules de grille. La cellule de lieu peut ainsi être activée par un état particulier des cellules de grille si la vision n'est pas accessible. Cela permet au robot de pouvoir se recalibrer en aveugle grâce à l'estimation de sa position et par l'information d'un contact tactile, au même titre qu'un animal pourrait reconnaître un coin de l'environnement dans le noir.

Après quoi le robot navigue pendant une nouvelle session de 30min dans le noir complet (voir Fig. 4.9). L'information visuelle n'est plus accessible et le retour au nid est réalisé à l'estime, via les cellules de grille. Les résultats soulignent le rôle de la vision dans le maintien d'une activité de grille précise et cohérente, puisqu'en l'absence d'indice visuel, la grille perd en précision angulaire. Le fait est que le robot tente encore d'aller se recalibrer en se basant uniquement sur des informations non-visuelles (l'intégration de chemin et le sens "tactile" fournis par les capteurs ultrasons), mais la dérive rapide de l'intégration de chemin fausse sa localisation et finit par créer une ambiguïté de perception de la zone de calibration. L'erreur atteint une amplitude qui finit par faire se tromper de coin le robot. A cause de la symétrie de l'enclos, chaque coin est identique et une erreur suffisamment forte dans la phase de retour au nid peut amener à une

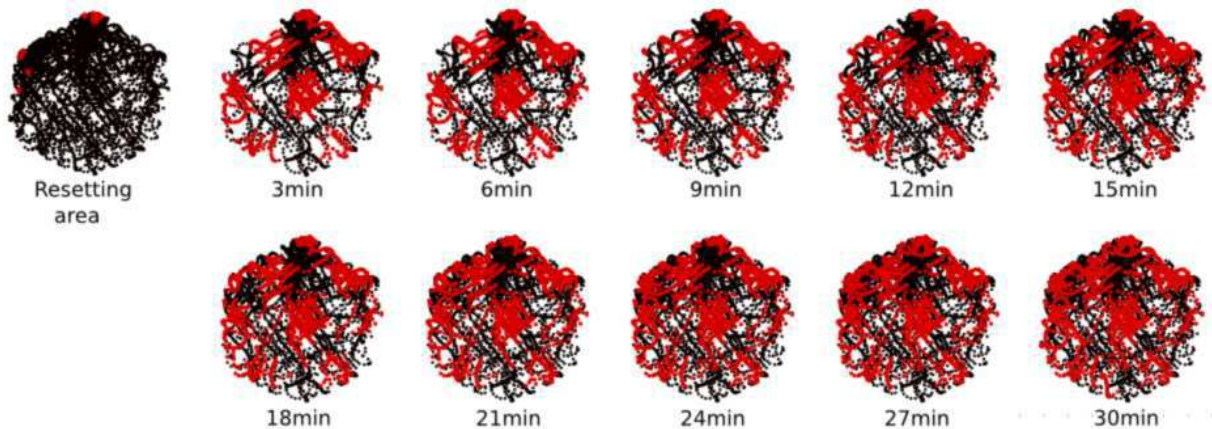


FIGURE 4.9 – Influence du cumul d'erreur sur l'activité de grille sans vision. Après 30min d'exploration standard, l'expérimentateur bande les yeux du robot de sorte qu'il soit aveugle pendant les 30 prochaines minutes. Le comportement de calibration n'est pas désactivé pour autant et le robot tente encore de se recalibrer chaque minute mais uniquement à l'estime (intégration de chemin + sens tactile). Due à l'accumulation d'erreur et à la symétrie de l'enclos, le point de calibration est perçu sur un autre coin au bout de 10min. L'activité de grille se dégrade alors sur les bords (le centre est toujours bien défini). C'est le résultat d'une rotation de 60° du repère. Après 20min, l'erreur est telle qu'elle affecte aussi la perception de la distance. Le robot ne parvient alors plus à se recalibrer puisque le point de calibration est perçu hors de l'enclos.

confusion entre 2 d'entre eux. C'est ce qui se produit ici au bout d'un certain temps (environ 10min), ce qui induit un rotation de 60° de l'activité de grille autour du centre de l'enclos. Cette rotation se traduit par une dégradation circulaire de l'activité. Les bords sont dégradés, mais le centre reste bien défini.

Par ailleurs, après 20 minutes d'expérience, le cumul d'erreur est devenu si important qu'il affecte également la distance de perception du but, et plus uniquement son angle. Le robot ne parvient plus à se recalibrer pendant les 10 dernières minutes tout simplement parce que la zone de calibration est perçue en dehors de l'environnement.

En résumé, bien que moins efficace qu'une recalibration multimodale (vision, intégration de chemin, perception tactile), la recalibration en aveugle retarde le cumul d'erreur et permet de maintenir une activité de grille relativement stable en comparaison d'une exploration aléatoire sans calibration. Elle pourrait être importante lorsque l'agent est privée de vision. Cependant, la difficulté réside dans le fait de choisir des points de recalibration non-ambigus. Ceux-ci doivent correspondre à des singularités de l'environnement, comme un coin ou le passage d'une porte étroite. Si le lieu de recalibration est ambigu, cela peut conduire à mettre en défaut le système. Lors de l'expérience précédente, la présence de plusieurs lieux de calibration (coins de l'enclos) partageant le même profil perceptif (distance au centre et profil tactile identique) peut amener le robot à les confondre.

4.3.3 Expérience 3 - De la cellule de grille à la cellule de lieu

Ici, nous étudions comment un ensemble de cellules de grille peut donner naissance à des cellules de lieu dépendantes uniquement d'informations idiothétiques. L'expérience est cette fois réalisée en simulation et se déroule dans un environnement carré (500x500 états discrets). Afin de conserver une approche minimaliste, nous utilisons un simple apprentissage associatif des

cellules de grille pour générer des cellules de lieu. Nous étudions ici les propriétés de généralisation de ces cellules, c'est à dire la taille du champ de lieu généré.

Dans notre modèle, chaque grille présente des champs binaires (activés ou non) de sorte que l'activité générée par la conjonction de 3 de ces grilles a une forme d'escalier à 3 marches. De plus, les 3 facteurs de modulo différents sont premiers entre eux, ce qui implique que l'activité de la conjonction change du tout au tout pour un petit mouvement. Des cellules de lieu au champ d'activité étroit se comportent comme si elles n'avaient aucune relation de proximité. Chaque cellule de grille est considérée comme une entrée orthogonale de sorte que le réseau de grille ne bénéficie d'aucune topologie (voir Fig. 4.10). Cependant, au moins 2 paramètres peuvent impacter sur la taille du champ de lieu : le facteur de discrétisation ainsi que le nombre de cellules de grille prise en compte. Nous avons étudié l'effet de ces paramètres sur le champ d'activité de la cellule de lieu générée.

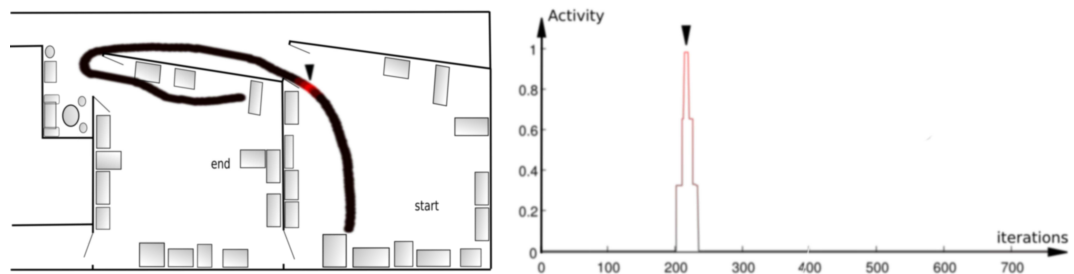


FIGURE 4.10 – Cellule de lieu odométrique (OPC) obtenue par la conjonction de 3 cellules de grille issues de modulo différents. A gauche : activité de la cellule de lieu sur la trajectoire du robot. A droite : Activité de la cellule dans le temps. Le champ de lieu a la forme d'un escalier à 3 marches très étroit. Il n'y a pas ou peu de généralisation de l'activité puisque l'opération de modulo supprime toute topologie.

Les résultats obtenus avec un grand nombre de cellules de grille montrent que le facteur de discrétisation a un impact direct sur la largeur du champ de lieu (voir Fig. 4.11). Le nombre de cellules de grille utilisées pour générer la cellule de lieu affecte la résolution du champ. Intuitivement, le nombre de cellules de grille reflète le nombre de marche de l'escalier d'activité du champ de lieu.

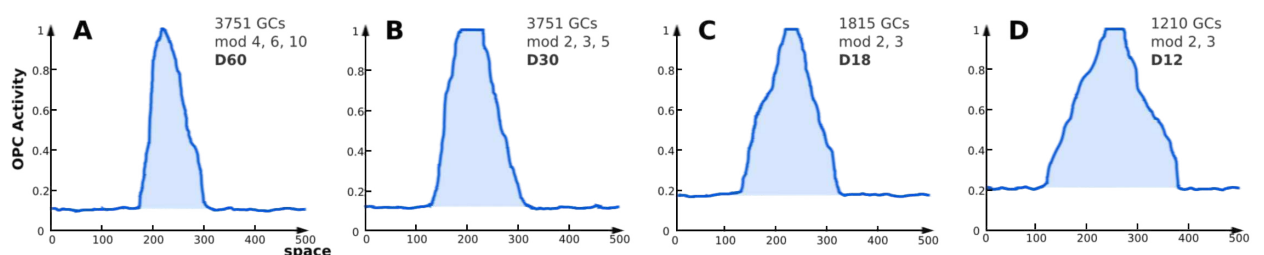


FIGURE 4.11 – Influence du facteur de discrétisation sur le champ d'activité de la cellule de lieu. La figure montre 4 cellules de lieu odométriques (OPC) obtenues en simulation par conjonction d'un ensemble de cellules de grille. Chaque cellule de lieu possède ses paramètres propres (*GCs* indique le nombre de cellules de grille utilisées, *mod* indique les modulo et *D* indique le facteur de discrétisation). Le facteur de discrétisation a un impact direct sur la largeur du champ de lieu.

4.4 Discussion

En partant de nos travaux précédents, nous avons proposé un modèle de cellule de grille qui ne nécessite pas de construire, a priori, une carte cartésienne de l'environnement. Ce modèle diffère d'autres travaux par plusieurs aspects. Premièrement, l'activité des cellules de grille observée est le résultat d'une simple projection puis d'une fusion de l'information d'intégration de chemin sur des neurones du cortex entorhinal. Ces projections d'une population donnée sur une autre plus restreinte agissent comme une opération de modulo appliquée à l'intégration de chemin. Elles n'impliquent aucune opération complexe et en ce sens s'avèrent relativement plausibles.

Nous avons réalisé des simulations de neurones de EC apprenant automatiquement à partir d'un nombre limité et aléatoire de connections entrantes. Les résultats montrent qu'une grande variété d'activités périodiques peut émerger lorsque l'on relâche certaines contraintes dans le modèle (les neurones extraits du champ d'intégration de chemin ainsi que les facteurs de projection modulo sont choisis aléatoirement : voir Fig.4.12).

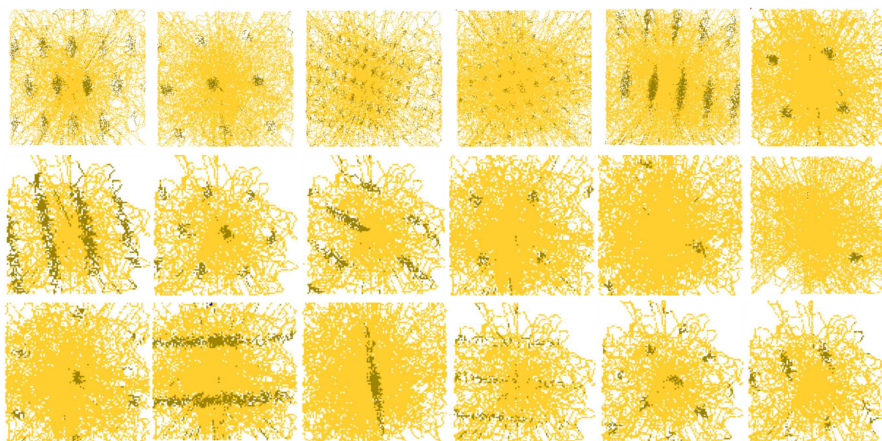


FIGURE 4.12 – Variété d'activité de grille obtenue dans un environnement simulé. Les cellules de grille résultent de projections modulo aléatoires. La figure montre l'activité de 27 cellules choisies aléatoirement parmi une population de 30000. L'agent simulé explore l'environnement pendant 20000 itérations. Les grilles obtenues présentent des caractéristiques très différentes (période spatiale, phase, orientation, angle séparant les axes de la grille).

Deuxièmement, notre modèle contraste avec la plupart des autres modèles de cellules de grille puisqu'il ne fait pas l'hypothèse d'une intégration de chemin réalisée par les cellules elles-mêmes à partir de l'information des cellules d'orientation de la tête (head direction cells) ou d'autres informations. A l'inverse, nous faisons l'hypothèse que l'intégration de chemin sur de grandes distances pourrait être stockée dans d'autres aires cérébrales, par exemple dans le cortex pariétal comme proposé par Etienne et collègues [Etienne and Jeffery, 2004].

Les expériences robotiques précédentes montrent que l'ajustement de quelques paramètres (fréquence des calibrations, position du but, rapport entre vitesse du robot et taille de l'environnement) permet d'obtenir des résultats assez proches de ceux obtenus lors d'expériences sur les rats. Les résultats de la figure 4.13 présentent une mesure de la qualité géométrique de la grille (gridscore) assez proche de celles enregistrés chez les animaux [Langston et al., 2010].

Cependant, il n'a pas été montré que les indices visuels présents exercent un contrôle fort sur l'alignement des grilles : La rotation d'un indice sur la surface d'un cylindre cause une rotation

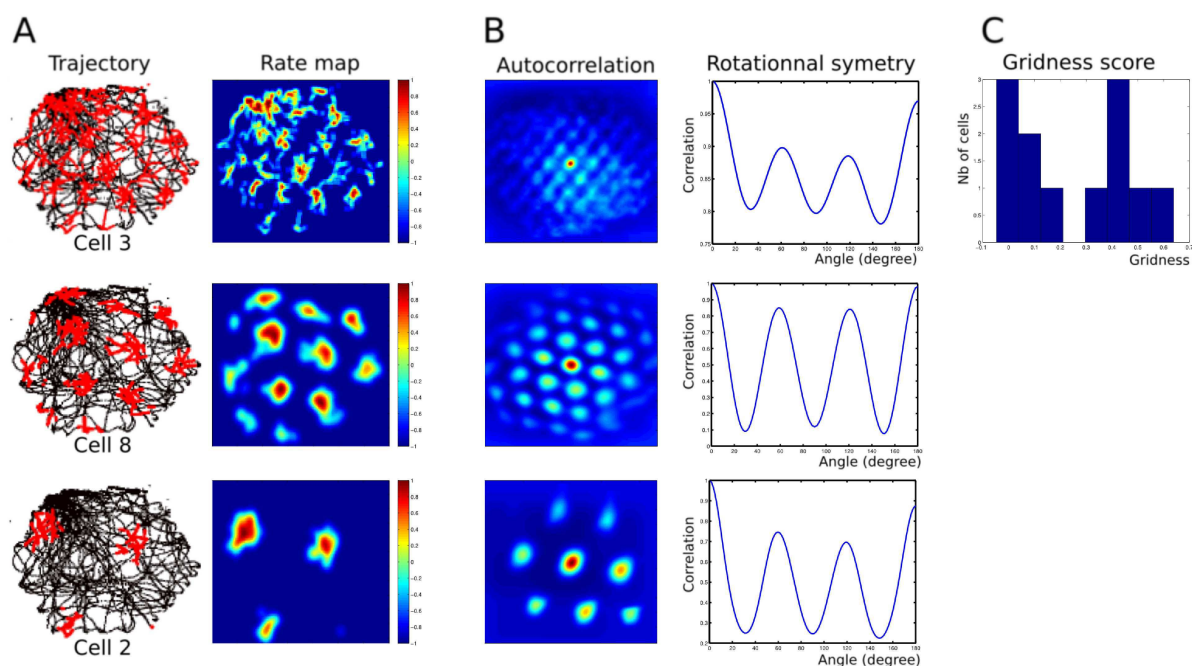


FIGURE 4.13 – A - Activité de grille pour 3 périodes spatiales différentes (une cellule par modulo). **Gauche** : l'activité de la cellule (en rouge) est superposée à la trajectoire du robot (en noir). **Droite** : carte de fréquence moyenne de décharge : rate-coding map (obtenue par convolution gaussienne), rouge=1, noir=0. **B** - Structure périodique de l'activité montrée en (A). **Gauche** : matrice d'autocorrélation spatiale de la rate-coding map. Bleu=-1, vert=0, rouge=1. L'échelle des distances est 4 fois plus grande que pour la rate-coding map. **Droite** : Autocorrélation par rotation (périodicité angulaire) : Calcul du coefficient de corrélation de Pearson entre la matrice d'autocorrélation original et sa version pivotée de 0 à 180° par pas de 1°. A cause de la tessellation hexagonale des grilles, le calcul fait apparaître la version modulée d'une sinusoïde, dont les maxima correspondent à des angles multiples de 60°. **C** - Histogramme du score de grille (gridness score) pour 12 cellules enregistrées dans les 3 couches de notre modèle de dMEC.

de la grille du même angle [Hafting et al., 2005]. Notre modèle ne peut pas reproduire directement ces résultats. Ceci est en fait dû à l'utilisation d'une boussole magnétique par le robot pour s'orienter, mais nous avons de fortes raisons de penser que le fait de remplacer cette boussole magnétique par une boussole visuelle [Giovannangeli and Gaussier, 2007] devrait permettre d'obtenir des effets similaires.

De la nécessité d'utiliser des lieux de calibration

Comme attendu, l'accumulation d'erreur de l'intégration de chemin, due aux glissements des roues du robot et aux erreurs de discrétisation, nuit rapidement (après 4 ou 5 minutes) au maintien d'une activité de grille cohérente, d'où la nécessité de prendre en compte d'autres informations. Une possibilité serait d'utiliser des cellules de bords (border cells) dont la découverte est récente [Solstad et al., 2008]. Ces cellules possèdent la particularité de décharger uniquement lorsque l'animal se trouve en bordure d'environnement (un coin ou un mur, typiquement). Une solution intéressante a été proposée par de récents travaux qui présente un moyen d'utiliser l'activité des cellules de bords pour réduire la dérive de l'intégration de chemin [Cheung et al., 2012]. Ces cellules de bord pourraient correspondre à l'information de reconnaissance tactile des

coins de l'enclos que nous avons mis en place de manière ad-hoc dans les expériences précédentes. [Hafting et al., 2005] a montré que l'activité des cellules de grille est fortement contrôlée par les indices visuels présents. Ainsi, nous avons choisi d'exploiter la modalité visuelle à l'aide d'un simple mécanisme de conditionnement entre cellules de lieu visuelles et cellules de grille pour limiter l'erreur de l'intégration de chemin. Il serait également plausible de réaliser la recalibration par la reconnaissance d'un amer saillant dans l'environnement, sans passer par la reconnaissance complète d'une cellule de lieu.

Ce que nous avons appris de ces expériences

Toutes ces expériences en lien avec un mécanisme de calibration peuvent nous donner plusieurs indices sur la manière de paramétrer le modèle afin de réaliser des tâches de navigation sensori-motrices (comme les séquences d'association Lieu/Action). Par exemple, nous avons souligné la question du paramétrage de la fréquence de calibration. Les résultats de la section 4.3.2 montrent que notre robot doit recalibrer son intégration de chemin après avoir parcouru au plus 48m. Ainsi, dans un environnement plus long, comme une pièce ou un couloir, la recalibration n'a pas de raison d'être réalisée pour chaque cellule de lieu, toutefois le robot devrait s'inquiéter et réduire le niveau de confiance attribué à l'intégration de chemin si aucune recalibration n'a eu lieu sur une distance de plus de 50m. Cela nous mène à la définition d'un niveau de confiance relatif aux modalités visuelles et proprioceptives. Les expériences relatives à l'importance de l'emplacement du lieu de recalibration soulignent également le même point clé, puisque l'activité de grille est mieux définie pour une zone de calibration étroite. De plus, la dernière expérience montre comment l'activité de grille peut être utilisée dans la génération de cellules de lieu. La taille du champ de lieu résultant peut être modulé par le facteur de discrétisation et le nombre de cellules de grille impliquées. Pouvoir moduler la largeur du champ de lieu est un paramètre important pour la navigation sensorimotrice, puisqu'il permet de définir les propriétés de généralisation du modèle.

Dans ce chapitre, nous avons décrit une architecture neuronale qui peut rendre compte sur robot réel des résultats obtenus dans le cadre de l'étude de cellules de grille sur l'animal. Nous avons démontré la nécessité d'un mécanisme de calibration, contrôlé par un besoin périodique de retour au nid, pour maintenir une activité de grille précise et cohérente. Nous avons également montré comment un ensemble de cellules de grille pouvait être fusionné pour générer des cellules de lieu par une simple règle d'apprentissage hebbienne. Notre modèle montre l'intérêt que pourrait avoir les cellules de grille pour la navigation de nos robots puisque notre système est capable de couvrir un environnement large en utilisant un petit nombre de résolutions de grille premières entre elles (seulement 3 modulo). D'un autre côté, les cellules de lieu obtenues ne permettent que d'obtenir un champ de lieu très étroit sans aucune capacité de généralisation. Cela pose la question de la façon dont un tel mécanisme pourrait être utile à l'animal pour pouvoir naviguer en pratique. Nous proposons l'hypothèse que ces capacités de généralisation pourraient émerger grâce au nombre important de cellules de grille présentes dans les différentes couches de dMEC. Lorsque nous simulons un grand nombre de ces cellules, nous parvenons ainsi à contrôler la taille du champ de lieu en fonction de certains paramètres. Le facteur de discrétisation a été identifié comme le principal paramètre contrôlant la taille du champ.

Publications personnelles

- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2012a). Multimodal integration of visual place cells and grid cells for navigation tasks of a real robot. In *From Animals to Animats 12*, number 7426 in LNAI, pages 136–145, Odense, Denmark. Springer
- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2011). Toward long range robot navigation. In *The International Workshop on Bio-Inspired Robots.*, Nantes, France
- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2012b). Multimodal integration of visual place cells and grid cells for robots navigation. In *Second Symposium on Biology of Decision Making*, Paris, France

Fusion d'informations multimodales

Précédemment, nous avons vu que la vision pouvait s'avérer ambiguë dans certains cas et ne pas suffire pour fournir une localisation robuste dans un environnement large, redondant ou encore pauvre en amers visuels. Nous avons également montré comment exploiter des informations proprioceptives pour générer des cellules de grilles cohérentes et plausibles biologiquement. L'un des rôles de ces cellules serait d'alimenter une représentation cartésienne de l'environnement en l'absence d'indices externes accessibles. Elles pourraient jouer un rôle clé dans la navigation spatiale chez l'animal. Nous avons montré comment les exploiter pour se localiser en proposant un modèle de cellules de lieu idiothétiques. Cependant, un problème persiste quant à la correspondance des informations entre vision et proprioception. Comment un système peut-il construire une représentation cohérente du lieu dans lequel il se trouve à partir d'information de nature complètement différente ?

Dans ce chapitre, nous proposons d'étudier cette question puis de proposer un modèle de fusion de ces 2 informations afin de générer des cellules de lieu multimodales, c'est à dire des cellules dont l'activation dépend à la fois d'informations allothétiques et idiothétiques. Ce modèle se base sur un simple conditionnement classique entre vision et proprioception. Nous verrons que ce mécanisme permet d'exploiter la synergie émergeant de ce couplage bi-modal en exploitant les bonnes propriétés de chacune de ces modalités et ainsi générer des cellules de lieu multimodales robustes. La robustesse de la reconnaissance de lieu sera vérifiée au cours de plusieurs expériences de navigation multi-salles sur robot réel. Nous montrerons notamment comment le système parvient à résoudre simplement le problème du kidnapping (transport passif du robot dans un autre lieu) ainsi que celui de la fermeture de trajectoire (lorsque le robot termine une boucle et revient au point de départ après un cumul d'erreur d'intégration de chemin).

Pour finir, nous ouvrirons la discussion sur la fusion de modalités en proposant un modèle plus générique d'intégration d'information multi-modales ne se limitant pas à 2 modalités. Ce modèle générique est une extension du modèle de fusion bi-modale et se base sur l'apprentissage de prédiction intermodales : en apprenant les corrélations existantes entre les régularités de plusieurs flux sensoriels, le modèle est capable de fournir un signal de substitution pour combler les manques d'une modalité défaillante.

5.1 Cohérence et conflits sensoriels

Afin de modéliser un système capable de résoudre des conflits entre différents flux sensoriels, il apparaît nécessaire d'étudier en détails les conflits eux même et la manière dont le cerveau animal les traite. La théorie du conflit sensoriel explique l'apparition de tels symptômes par le fait que le cerveau est confronté à des messages sensoriels contradictoires [Godefroid, 2001].

Pour rappel, le corps humain est constitué de sens externes tels que la vision, l'ouïe, l'odorat, le toucher ou le goût, mais également de sens internes comme la proprioception qui nous renseigne sur la position relative de nos membres, le système vestibulaire, situé dans l'oreille interne, permettant de percevoir la gravité et l'accélération, ou encore la somesthésie, c'est à dire l'ensemble de sensations internes au corps (peau, tendons, articulations, viscères) telles que pression, chaleur ou douleur [Marieb, 1993].

Il est important de bien faire la différence entre ces deux classes sensorielles dont une distinction fondamentale réside dans le fait qu'une modalité externe peut facilement être biaisée par un élément extérieur au corps (e.g. privation visuelle causée par l'ajout d'un foulard devant les yeux), tandis que les modalités internes sont, par nature, rarement mises en défaut (sauf cas pathologique et/ou cumul d'erreur).

Ainsi, comme nous le verrons par la suite, le cerveau s'affranchit relativement bien des situations de conflit entre différentes modalités externes, mais est bien souvent incapable de résoudre, de manière rationnelle, les conflits faisant intervenir une modalité interne. Il en résulte la création d'une nouvelle cohérence "irrationnelle" (voir section 5.1.2 et 5.1.2) voir même dans certains cas, une rupture perceptive accompagnée de douleurs et crampes [Ramachandran, 1998].

Le cerveau n'est généralement pas confronté à ce genre de cas de manière naturelle puisque les modalités internes possèdent des caractéristiques fondamentalement intimes, à tel point que la proprioception serait à la base de la conscience de soi. La proprioception permettrait d'apprendre à faire une distinction entre l'individu et le reste du monde, une notion que le nouveau-né ne possède pas encore [Damasio, 1999].

5.1.1 Conflits bénins entre modalités externes

L'être humain est bien adapté à la résolution de conflits perceptifs mineurs, survenant fréquemment au cours de la vie. Ces conflits se produisent uniquement entre modalités sensorielles externes et font référence à des situations dans lesquelles une modalité se trouve décorrélée des autres [Godefroid, 2001]. Un bon exemple de ce type de conflit est la privation sensorielle : l'obscurité ou à l'inverse, l'éblouissement total. Dans ce type de situation, il paraît évident de ne pas prendre en compte la vision qui se trouve être non-informative, et de privilégier les autres sens.

L'être humain est capable de détecter les sens défectueux ou n'apportant aucune information utile pour la tâche à réaliser. Celui-ci peut alors réaliser un changement radical de stratégie afin de s'adapter à la situation.

La notion de "situation", ou même de "tâche" va prendre tout son sens pour résoudre ce genre de conflit, puisque les modalités utiles à un certain moment donné dépendent fortement du contexte dans lequel est effectuée l'action. Ainsi Gentaz affirme que chaque modalité excelle dans le traitement de certaines propriétés tout en étant moins performante dans d'autres [Gentaz et al.,

2001; Hatwell et al., 2003]. Cette idée est défendue par [Freides, 1974; Hatwell, 1987; Pick, 1965] qui montrent la spécialisation fonctionnelle des modalités sensorielles. Ainsi, la modalité haptique (tactile) est très performante dans la perception de la texture et la dureté des matériaux, alors qu'elle l'est moins dans celle de l'espace, manifestement réservée à la vision. Dans les situations de fonctionnement bisensoriel, les deux modalités appréhendent en même temps un objet et se coordonnent pour le percevoir de façon cohérente. Lorsque le test est visuel, pour les aspects spatiaux, l'exploration bi-modale n'est pas différente de l'exploration unimodale visuelle. L'"effet ventriloque" est un bon exemple de telles situations dont nous faisons l'expérience régulièrement lorsque nous regardons la télévision : la voix semble émaner des lèvres de l'acteur et non de la source sonore réelle [Alais and Burr, 2004]. De même, l'effet McGurk montre une interférence entre l'audition et la vision lors de la perception de la parole : La vue de lèvre prononçant le phonème "ga", tandis que la bande sonore diffuse l'enregistrement du phonème "ba", induit l'illusion de la perception d'un phonème intermédiaire "da" [McGurk and MacDonald, 1976]. L'intégration multi-sensorielle peut également améliorer la reconnaissance : la compréhension d'un discours se trouve améliorée si la vision du mouvement des lèvres est accessible et synchrone avec le flux sonore correspondant [Schroeder et al., 2008]. Pour un test tactile, l'information visuelle améliore également la reconnaissance [Walsh, 1973] : comme la perception visuelle est plus riche et plus économique, l'exploration haptique ne serait pas mise spontanément en oeuvre (puisque'elle n'apporte pas d'informations supplémentaires). Lors de conflits perceptifs, la vision est dominante pour le traitement spatial et les informations haptiques ne sont utilisées que dans le cas d'une rupture forte de la cohérence [Power and Graham, 1976].

Delorme et Flückiger décrivent, dans le cadre d'études effectuées sur l'animal, des structures cérébrales qui répondent à l'excitation de deux ou trois modalités simultanées ([Delorme and Flückiger, 2003]). Ces structures multimodales suivent des règles d'intégration particulières : l'effet d'une stimulation bimodale n'est pas additif mais multiplicatif, c'est à dire qu'une réponse neuronale faible correspondant à un stimulus visuel peut être amplifiée jusqu'à 12 fois si celle-ci est déclenchée par un stimulus bimodal (visuel + auditif) [Stein and Meredith, 1993]. Des stimulations bimodales simultanées augmentent la réponse, tandis que des stimulations décorréliées, spatialement ou temporellement, la diminuent. Cependant certaines réponses peuvent être renforcées même si les deux stimuli sont décalés dans le temps d'une fenêtre de 200ms, comme dans le cas d'un stimulus audio-visuel qui s'explique aisément par la différence de rapidité de propagation entre les deux sources. [Stein and Meredith, 1993] parlent également d'une "règle d'efficacité inversée" qui stipule que plus un stimulus unimodal est efficace et moins il produit de réponse en combinaison avec un autre. Inversement, deux stimuli peu efficaces peuvent augmenter la réponse neuronale une fois combinés.

Le cerveau est également capable de substitution sensorielle et permet ainsi de compenser un handicap perceptif en substituant un sens par un autre [Bach-y Rita et al., 1969]. Les personnes non-voyantes utilisent généralement une canne pour se déplacer, c'est à dire une extension directe du sens du toucher. D'autres préféreront un système auditif directif renseignant l'utilisateur sur la présence éventuelle d'obstacles proches [Malika Auvray, 2003]. Les travaux de [Bach-y Rita et al., 1969] proposent un système de perception tactile de l'espace en recréant une image en temps réel par la transformation des signaux d'une caméra numérique en 400 impulsions électriques réparties sur une plaque de 20 lignes par 20 colonnes directement

appliquée sur le dos du sujet. Ce dispositif permet à des aveugles de "voir" par l'intermédiaire du touché. On parle dans ce cas de substitution visuo-tactile.

5.1.2 Conflits majeurs et ruptures perceptives

Toutefois, si l'incohérence est trop grande ou si elle implique une modalité interne, le sujet en prend alors conscience et ressent une sensation de mal être important. Selon la théorie de [Festinger, 1957], l'individu en présence de perceptions incompatibles entre elles, éprouve un état de *dissonance cognitive* se traduisant par une tension désagréable. Dès lors, le sujet met en oeuvre des stratégies inconscientes visant à restaurer l'équilibre cognitif. Ce principe est connu sous le nom de "Reduction of cognitive dissonance", qui affirme que les organismes sont motivés pour réduire l'écart entre la prédiction des effets d'une action et la situation réellement perçue [Festinger, 1957; Kagan, 1972]. C'est le cas notamment de la cinétose, plus connu sous le nom de "*mal de mer*" ou "*mal des transports*", qui désigne en réalité le mal du mouvement (motion sickness) dans sa globalité.

Le phénomène se produit lorsqu'une personne est en mouvement dans un véhicule quelconque, abandonnant ainsi le mode de déplacement naturel. Les accélérations et décélérations répétées perçues par l'oreille interne entre en conflit avec la vue immobile de l'intérieur du véhicule. Le conflit continu entraîne nausées et maux de tête. Treisman a suggéré que cette réaction au conflit sensoriel se serait développée au cours de l'évolution, afin de protéger l'organisme contre certains poisons absorbés par mégarde. La contradiction entre différentes informations sensorielles amènerait l'organisme à réagir par des nausées et des vomissements afin de se débarrasser de substances potentiellement toxiques [Treisman, 1977]. La cinétose est présente en voiture, en train, en avion, en bateau, dans l'espace mais également dans les environnements de réalités virtuels [So and Lo, 1999].

Le cas le plus étudié, en raison du danger qu'il représente, reste sans doute la désorientation spatiale des pilotes qui survient lors d'une mauvaise visibilité et à laquelle on attribue 5 à 10% des accidents dans le domaine de l'aviation [Melchor J. Antunano, 2005].

Lors d'un vol, le pilote peut perdre très rapidement son orientation par rapport au sol à cause de ce que l'on nomme la verticale subjective (l'axe verticale du corps). En effet, en cas de perte de repères visuels, la verticale subjective correspond alors au référentiel gravitaire [Luyat and Gentaz, 2002]). L'erreur d'un pilote navigant en aveugle en raison de mauvaises conditions climatiques est de tenter de garder son plateau horizontal en se fiant uniquement à ses perceptions. Une autre situation de rupture perceptive intéressante est obtenue en stimulant le système vestibulaire humain par le biais d'impulsions électriques. Ce processus affecte sensiblement l'équilibre du sujet et permet, dans une certaine mesure, de contrôler ses déplacements à distance. Les stimulations induites donnent au sujet l'impression de tomber, ce qui le pousse à se déplacer pour retrouver son équilibre. Cette technique encore à l'étude est appelée GVS pour Galvanic Vestibular Stimulation [Day and Fitzpatrick, 2004].

Un autre exemple de rupture perceptive concerne l'expérience appelée "illusion de Pinocchio" [Lackner, 1988; Ehrsson et al., 2005; Thyrión and Roll, 2009]). En appliquant sur les muscles du bras des vibreurs électrique, il est possible de déclencher des signaux factices d'élongation ou de contraction de ces muscles, créant chez la personne une illusion de mouvement. Partant de ce fait, on place le doigt du sujet en contact avec le bout de son nez de sorte qu'il puisse sentir ce contact de manière tactile.

Lorsque l'on simule un mouvement du bras par vibration, il y a apparition d'un conflit perceptif entre proprioception et toucher. Le sujet a alors recours à un processus de rationalisation pour réduire la dissonance : puisque le contact entre le doigt et le nez est toujours présent, le sujet perçoit des variations de la taille de son nez. Le nez semble s'allonger si le mouvement illusoire éloigne l'index du visage et réciproquement.

5.2 Un couplage bi-modale par simple conditionnement

Dans cette section, nous proposons un modèle minimaliste de couplage entre informations visuelles et proprioceptives pour la génération de cellules de lieu multimodales robustes. Nous démontrerons les capacités de ce modèle par le biais d'expériences de reconnaissance de lieux entre plusieurs pièces de notre laboratoire.

On peut imaginer différentes façons de réaliser une fusion de sources, et ce point est connu comme étant un problème difficile, particulièrement lorsque la nature des sources est complètement différente. Les modalités sensorielles doivent être recodées au sein d'un format commun avant de pouvoir être fusionnées, puisqu'elle n'utilise pas la même représentation ("*recoding problem*") [Pouget et al., 2002].

Une première méthode consiste en la simple fusion des 2 modalités par le biais d'une catégorisation de l'ensemble [cellules de grille, PrPh] considéré comme une seule entrée (voir fig. 5.1-A). Cependant, cette solution n'est pas satisfaisante dès lors que le nombre de neurones représentant chaque modalité n'est pas équivalent. Dans le cas présent, la modalité visuelle (PrPh) comprend 1500 neurones tandis que la modalité proprioceptive (cellules de grille) ne comprend que 819 neurones (169 pour la grille de petite période, 289 pour la grille de période intermédiaire et 361 pour la plus grande). La quantité d'énergie diffère également, puisque le groupe PrPh contient en permanence plus de 40 neurones actifs, contre 3 seulement pour les cellules de grille brutes (une cellule de grille active à la fois pour chacun des 3 modulo).

Une seconde méthode serait de catégoriser chaque modalité indépendamment dans un premier temps, puis de procéder à la fusion des 2 catégories dans un second temps (voir fig. 5.1-B). Un apprentissage hebbien pourrait ensuite être utilisé comme mécanisme de fusion en apprenant les corrélations entre catégories visuelles et proprioceptives, le réseau renforçant alors le poids des neurones co-activés. Cette méthode est utilisée par [Arleo and Gerstner, 2000; Arleo et al., 2001] pour l'apprentissage non supervisé de cellules de lieu multimodales à partir de cellules visuelles et d'un système idiothétique basé sur l'intégration de chemin. Une troisième méthode possible serait de considérer une nouvelle étape de catégorisation. Le système apprendrait alors des configurations de cellules de lieu visuelles et odométriques. Mais cette solution ne permet pas de moduler la pondération attribuée à chaque modalité.

Le modèle de fusion que nous avons choisi d'utiliser est celui d'un conditionnement classique entre cellules de lieu visuelles et cellules de grille. L'apprentissage du conditionnement est réalisé par une variante normalisée de l'algorithme des moindres carrés (NLMS) [Haykin, 2002] qui permet entre autre une neuromodulation locale. Les cellules de lieu visuelles constituent la sortie désirée (stimulus inconditionnel) tandis que les cellules de grille représentent l'entrée conditionnelle. Les cellules de lieu odométriques obtenues après association possèdent l'avantage de correspondre avec leurs homologues visuelles, puisqu'elles sont conditionnées par

celles-ci. Une somme pondérée des 2 modalités permet ensuite d'obtenir des cellules de lieu multimodales. Ce modèle propose une méthode de fusion simple et efficace puisqu'elle permet d'obtenir des cellules cohérentes sans besoin de recourir à une quelconque catégorisation intermédiaire. De plus, la somme pondérée permet d'ajuster facilement l'importance d'une modalité vis à vis de l'autre sur l'activité finale. On notera que le mécanisme utilisé est le même que pour la recalibration du champ d'intégration présentée au chapitre 4. Une solution similaire est proposée par Strosslin et al. [Strösslin et al., 2005].

Selon notre modèle, l'activité MPC_k d'une cellule multimodale k est donc donnée par :

$$MPC_k = \alpha.VPC_k + (1 - \alpha).PredVPC_k$$

avec VPC_k l'activité de la cellule de lieu visuelle correspondante, $PredVPC_k$ l'activité de la cellule de lieu odométrique (cellule construite à partir des grilles pour prédire la cellule de lieu visuelle correspondante), $\alpha \in [0;1]$ le facteur de pondération de chaque modalité (ce paramètre est égale à 0.5 dans la suite de ce chapitre, par souci de simplicité) en enfin $k \in [0;K]$ l'indice de la cellule multimodale parmi un ensemble de K cellules.

L'activité de la cellule de lieu odométrique est donnée par :

$$PredVPC_k = \sum_{k=0}^{K-1} w_k.G_k$$

où w_k représente le poids de la synapse provenant de la cellule de grille G_k correspondante. La loi d'évolution de ce poids est donnée par l'algorithme du NLMS :

$$\frac{dw_k}{dt} = \frac{\lambda_k.(VPC_k - PredVPC_k).G_k}{\sum_{k=0}^{K-1} G_k^2 + \epsilon}$$

avec $\epsilon = 0.01$ un scalaire empêchant toute division par zéro. $\lambda_k = \gamma . \beta_k . \lambda$ représente un facteur de vigilance locale dépendant d'une constante λ , d'une neuromodulation binaire γ et d'un mécanisme de compétition WTA β_k de sorte que :

$$\beta_k = \begin{cases} 1 & \text{si } k = \text{Argmax}(VPC) \\ 0 & \text{autrement} \end{cases}$$

Les deux modalités considérées sont de natures assez complémentaires. Le système visuel seul permet une reconnaissance de lieu assez grossière mais possède de bonnes capacités de généralisation (champs de lieu larges). Il présente des erreurs de perception dues à certains indices redondants ou pauvres dans les environnements trop complexes. Au contraire, la reconnaissance proprioceptive propose une localisation très fine (champs de lieu étroits et singuliers) mais est soumise au cumul d'erreurs inhérent au système d'intégration de chemin basé sur des informations idiothétiques.

Grâce à la nature complémentaire (interne et externe) de ces 2 modalités, leur fusion permet une localisation robuste. Le modèle de fusion permet d'exploiter la synergie des 2 informations, chacune comblant les défauts de l'autre.

5.3 Robustesse de la localisation : une expérience sur robot réel

Afin de tester la robustesse de la reconnaissance de lieu de notre modèle de fusion, nous avons réalisé une expérience de navigation passive en intérieur dans un environnement de 25x15m

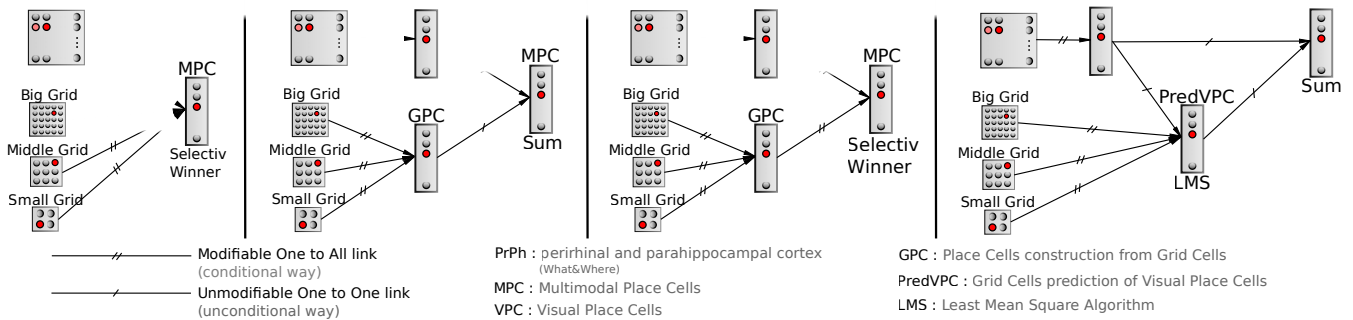


FIGURE 5.1 – Modèles possibles de fusion des informations visuelles et proprioceptives. **A** - Fusion par catégorisation du tout : Le poids de l'activité des grilles est très faible devant celui du PrPh. **B** - Catégorisation puis fusion. Aucune correspondance n'existe alors entre les champs de lieu des 2 modalités. **C** - Méta-catégorisation. Pas de gestion de la pondération possible. **D** - Fusion par conditionnement classique. Le réseau réalise une association entre l'activité du PrPh (stimulus inconditionnel) et celle des cellules de grille (stimuli conditionnels). La fusion est ensuite réalisée par simple somme. La pondération de la somme est facile à réaliser. Aucune catégorisation intermédiaire n'est nécessaire.

(une portion de notre laboratoire). À des fins d'analyse, nous avons supervisé l'apprentissage du robot pour tester la reconnaissance de 19 lieux régulièrement espacés (un lieu appris tout les 1.5m) sur une trajectoire multi-pièces (voir fig. 5.2). La trajectoire débute dans une première pièce du laboratoire riche en amers visuels, elle traverse ensuite un couloir pauvre en amers pour enfin finir dans une seconde pièce très similaire à la première. On teste ensuite les capacités de reconnaissance de ces lieux en faisant suivre au robot 5 trajectoires parallèles à celle apprise. Ces 5 trajectoires permettent de tester les capacités de généralisation des cellules de lieu multimodales en fonction de l'écart aux lieux appris.

Afin de valider la nature déterministe des résultats, nous avons répété l'expérience une douzaine de fois dans un environnement dynamique (changement de luminosité ambiante, déplacement d'objets, mouvement de personnes). Les résultats montrent une caractéristique intéressante et pour le moins imprévue puisque le mécanisme de fusion parvient à garder une localisation cohérente même lorsque des erreurs apparaissent simultanément sur chaque modalité (voir fig. 5.3). Le fait de fusionner les modalités permet de mettre en avant les votes unanimes tout en réduisant l'impact des décisions contradictoires (un vote divisé). Par exemple, si les 2 modalités reconnaissent un même lieu à 50%, l'activité de ce lieu se verra renforcé après fusion. L'activité de ce lieu sera supérieure à celle d'un lieu reconnu à 90% par une modalité mais à 0% par l'autre. Au final, le modèle permet d'obtenir des cellules de lieu multimodales plus robustes qu'une localisation basée uniquement sur une seule information. Certaines fausses alarmes peuvent toutefois se produire dans l'éventualité d'une reconnaissance ambiguë (mauvais lieu reconnu) partagé par les 2 modalités (l'erreur d'identification du lieu commise par les modalités est identique). Cependant la probabilité d'apparition de telles fausses reconnaissances devrait être faible étant donnée la différence de nature des modalités. Cette probabilité décroît avec le nombre de modalités sensorielles impliqués. Si les modalités sont indépendantes, elles ne sont donc pas soumises aux mêmes perturbations. Il sera intéressant de valider cette hypothèse dans de futurs travaux en caractérisant plus finement la probabilité d'apparition de telles erreurs dans différentes situations.

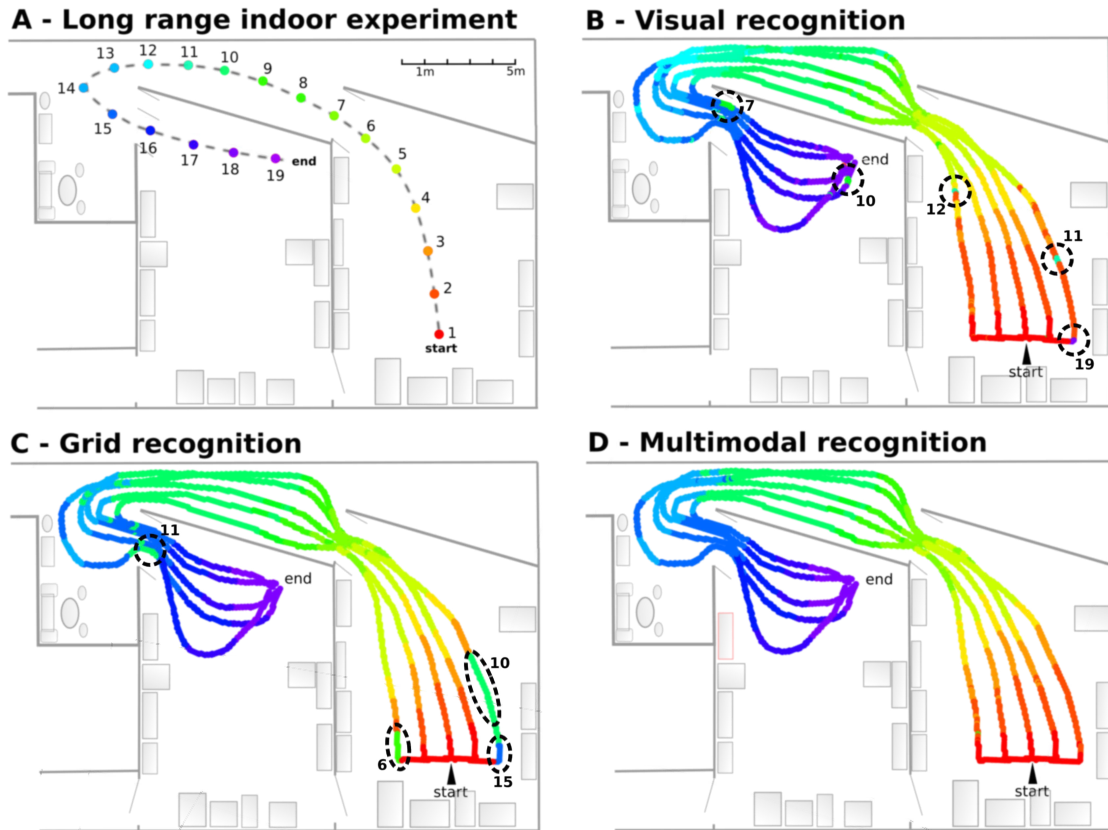


FIGURE 5.2 – Comparaison des performances de localisation monomodale et multimodale lors d’une expérience de navigation en intérieur. **A** - Setup expérimental : le robot a appris 19 lieux régulièrement espacés de 1.5m. La trajectoire commence dans une pièce, traverse un couloir et se termine dans une seconde pièce assez similaire à la première. **B** Reconnaissance visuelle testée sur 5 trajectoires parallèles à celle apprise. Chaque couleur est associée à une cellule de lieu particulière. Les résultats montrent une bonne capacité de généralisation mais souligne certaines ambiguïtés (cercles en pointillés. Le numéro correspond à l’indice du lieu appris en A). **C** Reconnaissance proprioceptive (cellules lieu odométriques issues des cellules de grille) lors des mêmes trajectoires. Les champs de lieu sont plus étroits mais ne présentent aucune ambiguïté avec un lieu lointain. Cependant, le bon lieu n’est plus reconnu lorsque le robot quitte la zone de généralisation du lieu appris. L’activité du lieu gagnant est alors faible et correspond à un second maximum issu du modulo de certaines grilles (6, 10, 15). La proprioception est aussi sensible au rotation, ce qui induit des erreurs de reconnaissance dépendantes de la trajectoire du robot (11). **D** Reconnaissance multimodale (cellules de lieu issues de la fusion des cellules visuelles et proprioceptives). La synergie des 2 modalités permet d’obtenir une reconnaissance plus robuste en conservant de bonnes capacités de généralisation. De légères erreurs subsistent (7, 15) mais impliquent un lieu proche du lieu réel.

5.3.1 A propos de la topologie des cellules de grille

Nous avons discuté au chapitre 4 le fait que notre modèle crée des champs de cellules de grille binaires (une cellule est active ou non) ce qui entraîne la génération de cellules de lieu odométrique au champs très étroit (escalier à n marches lorsque l’on fusionne n grilles). Le fait que les facteurs de projections modulo soit premiers entre eux supprime toute topologie entre cellules de grille. En d’autres termes, l’activité d’une cellule de lieu odométrique est nulle si le robot ne se trouve pas dans un rayon très proche du lieu appris. Ce rayon dépend du champ d’activité de la cellule de grille dont la période est la plus large.

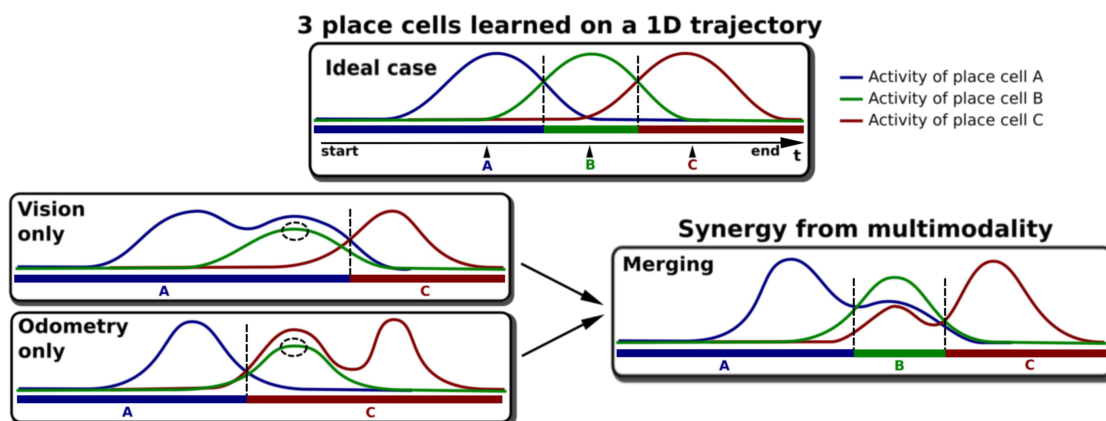


FIGURE 5.3 – Comment la synergie émerge t'elle de la corrélation : illustration du mécanisme de fusion dans un cas schématique simple (trajectoire mono-dimensionnelle). 3 cellules de lieu ont été apprises respectivement aux lieux A, B et C. Dans le cas idéal (en haut) les champs de lieux sont bien définis. Dans le cas d'une ambiguïté visuelle différente sur les 2 modalités (en bas à gauche) il y a contradiction. Ni la modalité visuelle, ni la modalité proprioceptive ne reconnaît le lieu B (vert) comme gagnant. Mais, puisque la fusion renforce les activités corrélées (votes unanimes) et diminue les décisions contradictoires, le lieu B est mis en avant, tandis que les lieux A et C se voient diminués. La cellule de lieu multimodale B reste cohérente même lors de fausses reconnaissances de la part des 2 modalités, pour peu que le même lieu ne soit pas gagnant pour ces 2 modalités. Des cas problématiques pourraient apparaître si des perturbations amènent les modalités à reconnaître le même lieu ambigu. Cependant, la probabilité d'apparition de tels cas est faible étant donnée la différence de nature des modalités. Cette probabilité décroît avec le nombre de modalités sensorielles impliqués.

Notre hypothèse était de dire que le cerveau de l'animal contient en réalité un très grand nombre de cellules de grille, ce qui pourrait expliquer l'observation de cellules de lieu au champs d'activités larges, même en l'absence d'indice visuel. Les simulations que nous avons présentées, impliquant 30000 cellules de grille, semblent aller dans ce sens (voir chap. 4, fig. 4.11). Néanmoins, ces simulations étaient réalisées hors ligne. Un nombre si important de cellules pose des problèmes évidents de coût de calcul lorsque l'on désire contrôler un robot en temps réel.

Afin de pallier ce problème, nous avons choisi, dans nos travaux, de ne conserver que les cellules de grille issues de 3 modulo différents (générant le champ étroit en forme d'escalier à 3 marches) et de recréer une topologie virtuelle à l'aide d'une méthode ad-hoc. Cette méthode consiste en la convolution de chaque cellule de grille par un masque gaussien, à l'instar d'un noyau d'interaction entre cellules proches. Ce noyau gaussien recrée une topologie entre cellules de grille et permet ainsi d'obtenir de meilleures capacités de généralisation. La topologie obtenue est celle d'un tore et s'inspire du modèle de [McNaughton et al., 2006].

Cependant, la généralisation obtenue dépend entièrement de la taille du noyau d'interaction utilisé. Lorsque le robot quitte le champ de généralisation du lieu proprioceptif appris il n'est plus capable de reconnaître ce lieu. Le lieu proprioceptif gagnant résultant correspond alors à un second maximum issu de la synchronisation d'un sous-ensemble de cellules de grilles (voir Fig. 5.2 - C : ambiguïtés 6, 10 et 15). Les lieux construit à partir des cellules de grilles possèdent des maxima secondaires parasites dont l'activité est inférieure à celle présente à l'endroit de l'apprentissage de ce lieu. En aucun cas ces maxima secondaires ne peuvent être utiles au robot pour se repérer ni pour effectuer une remontée de gradient (pour plus de détails, on pourra se

référer à [Gaussier et al., 2007]).

5.3.2 Robustesse de la reconnaissance des lieux aux ambiguïtés visuelles

Les 2 pièces du laboratoire utilisées pour notre test partagent un grand nombre de caractéristiques spatiales et visuelles, ce qui induit certaines ambiguïtés dans la reconnaissance des lieux issue du système visuel seul. Ce problème n'est pas particulièrement gênant lorsque l'ambiguïté existe entre 2 lieux spatialement proches ou lorsque les lieux ambiguës sont associés à des actions similaires (même direction). Néanmoins, cela peut devenir dramatique lorsque l'ambiguïté survient entre 2 lieux dont l'action associée est différente. Dans notre cas, une ambiguïté forte existe entre les portes d'entrée des 2 pièces (voir Fig. 5.2 lieux 7, 16 et Fig. 5.4 - A). Le robot a appris à sortir de la première pièce et, à l'inverse, à entrer dans la seconde (directions contradictoires). L'environnement visuel proche des 2 portes est très similaire.

De plus, la dynamique des activités des cellules de lieu visuelles se réduit de manière drastique lorsque le système doit apprendre un grand nombre de lieux (voir chapitre 3) et induit une augmentation de la probabilité de confusion.

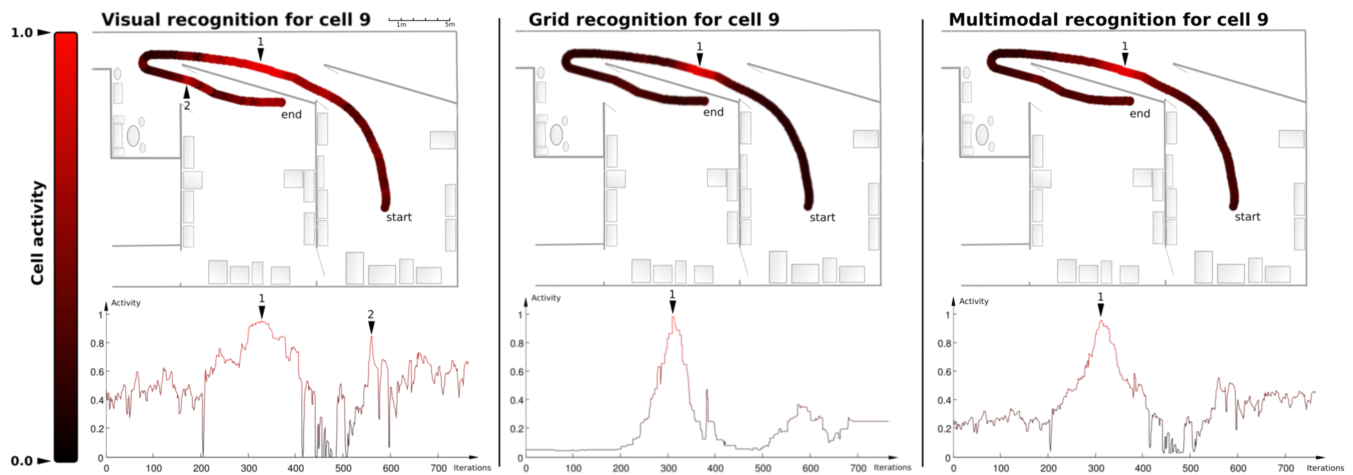


FIGURE 5.4 – Comparaison entre les cellules de lieu unimodales (visuelles et odométriques) et la cellule de lieu multimodale lors d'un trajet multi-pièces. Le lieu n°9 à été choisi pour illustrer son caractère ambigu. **Haut** : Activité de la cellule superposée à la trajectoire du robot (rouge=1, noir=0). **Bas** : activité de la cellule dans le temps. **A** - La cellule de lieu visuelle montre une activité maximum au lieu d'apprentissage (1), mais le lieu est aussi reconnu en (2) à cause de la redondance d'indices visuels. **B** - La cellule de lieu odométrique (issue des cellules de grille) montre une activité nette et bien localisée sur le lieu appris (1). **C** - La cellule de lieu multimodale tire avantage des 2 informations pour générer un champ d'activité robuste, large et bien défini.

5.3.3 A propos du problème du kidnapping

Précédemment, nous avons vu que l'ajout de la modalité proprioceptive pouvait permettre d'augmenter la robustesse de la localisation. Mais qu'arrive t'il si le robot est soudainement transporté les yeux bandés dans un autre lieu ? Si ce lieu est bien reconnu par le système visuel, le robot déclenchera une recalibration du champ d'intégration de chemin à l'état associé précédemment à ce lieu. Nous avons réalisé plusieurs expériences de kidnapping pour tester la

capacité du modèle à retrouver une cohérence entre vision et proprioception. Après un kidnapping, le système proprioceptif conserve la reconnaissance du lieu précédent. Si la trajectoire prend une orientation différente de celle apprise avant kidnapping alors l'amplitude de l'activité de reconnaissance des lieux décroît rapidement jusqu'à atteindre le niveau de bruit de cette modalité (au delà du champ de généralisation artificiel des cellules de grilles) (voir fig. 5.5). Le système visuel quant à lui reste insensible au kidnapping puisqu'il se base sur des informations externes. L'activité des lieux visuels reste forte et le système déclenche une recalibration de l'intégration de chemin lorsqu'il atteint un lieu fortement reconnu. Après recalibration, les 2 modalités sont de nouveau cohérentes entre elles et le robot poursuit sa tâche normalement.

Dans la majorité des cas, la reconnaissance multimodale reste relativement cohérente durant la phase critique (lorsque le système n'a pas encore déclenché de recalibration) puisque l'activité de reconnaissance proprioceptive décroît tandis que celle de la reconnaissance visuelle reste forte. Le poids de la vision est alors plus grand que celui de la proprioception (voir fig. 5.6).

Un cas problématique peut toutefois se produire si l'orientation de la trajectoire après kidnapping est similaire à celle apprise avant kidnapping. Dans une telle situation, la modalité proprioceptive conserve une activité aussi forte que la vision. Le système n'a alors aucun moyen de privilégier une modalité ou une autre. La reconnaissance multimodale reste incohérente jusqu'à la prochaine recalibration.

Jusqu'à présent, nos modèles de fusion par conditionnement ainsi que de recalibration donnaient une préséance à la vision sur la proprioception. Cet a priori permettait de maintenir une cohérence entre les 2 modalités, mais il peut s'avérer dangereux si la vision ne fonctionne pas correctement pour une raison ou pour une autre. Ce point pose le problème de la confiance à accorder à chaque information. Actuellement notre modèle ne possède pas de capacité d'évaluation de la pertinence d'une modalité en fonction de la situation. Dans de futurs travaux, il sera intéressant de mettre en place un système d'évaluation capable de moduler l'importance d'une modalité en fonction de sa pertinence dans la réalisation de la tâche. Nous reviendrons sur ses aspects d'auto-évaluation au chapitre 6.

5.4 Discussion

Dans ce chapitre, nous avons montré comment il était possible d'exploiter l'information de notre modèle de cellules de grille en présentant un modèle de fusion des modalités visuelles et proprioceptives par le biais d'un simple conditionnement. Ce modèle permet la génération de cellules de lieu multimodales autorisant une reconnaissance de lieu plus robustes qu'une reconnaissance unimodale. Par ailleurs, une propriété intéressante de cette fusion est apparue, mettant en avant la synergie des 2 modalités. La fusion bi-modale peut être vue comme un vote des 2 participants (vision et proprioception) pour élire le lieu dans lequel le robot se trouve. Cette fusion donne plus de poids à un vote unanime qu'aux votes contradictoires. Ainsi la reconnaissance multimodale est globalement plus robuste que celle de la modalité la plus fiable. Des expériences futures devraient permettre de quantifier précisément cette différence de fiabilité.

En psychologie cognitive, une expérience de discrimination de forces visuelles réalisées sur l'homme tend à montrer qu'un effet de synergie émerge de la décision collective. Une décision issue d'un consentement mutuel entre 2 participants donne de meilleurs résultats de discrimination que la décision unilatérale de l'individu possédant la meilleure acuité visuelle. La

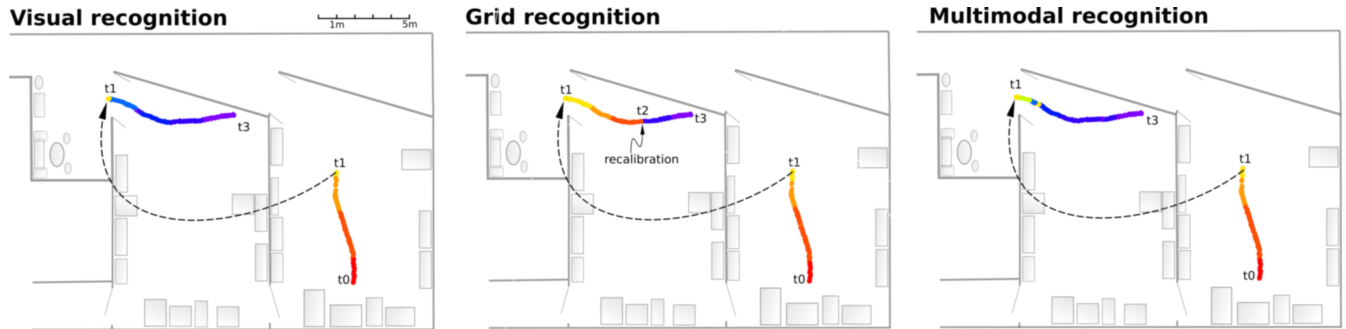


FIGURE 5.5 – Comment le système parvient à résoudre le problème du kidnapping : le robot navigue de (t_0) à (t_1) lorsqu’il est soudainement kidnappé, foulard sur la caméra, puis transporté dans une autre pièce (t_1). Il parvient à reconnaître les lieux puis termine sa trajectoire comme il a appris à le faire de (t_1) à (t_3). **A** - Reconnaissance visuelle superposée à la trajectoire. Le système visuel reconnaît instantanément le lieu après le kidnapping. Le lieu est fortement reconnu en (t_2) et autorise le système à envoyer un signal de recalibration. **B** - Reconnaissance proprioceptive. Après le kidnapping, le système proprioceptif est perdu de (t_1) à (t_3) et croit revenir en arrière. Le signal provenant de la vision à (t_3) permet de recalibrer l’intégration de chemin et de retrouver ainsi une reconnaissance proprioceptive cohérente. **C** - La reconnaissance multimodale permet d’obtenir une localisation correcte peu de temps après le kidnapping, avant même que le signal de recalibration n’apparaisse.

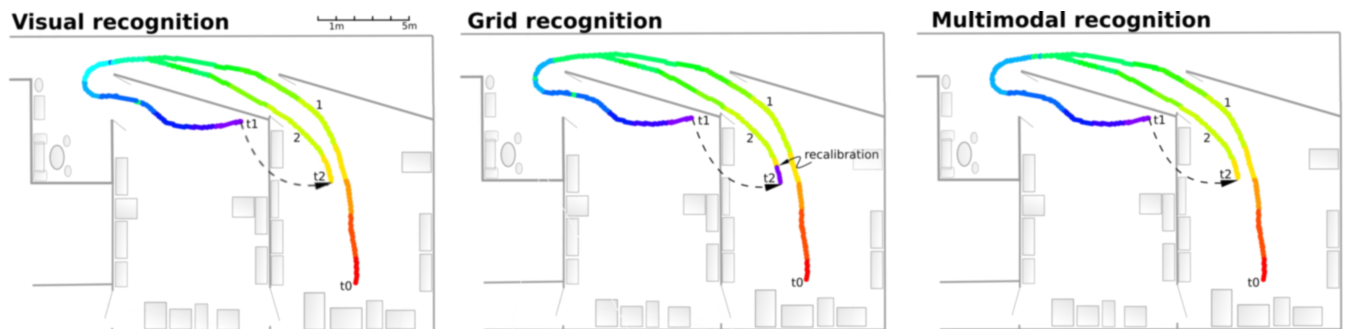


FIGURE 5.6 – Un autre exemple de kidnapping. Le robot navigue de (t_0) à (t_1) en empruntant la trajectoire (1). Il est kidnappé en (t_1) et transporté dans l’autre en pièce (t_2). Il parvient à reconnaître les lieux puis converge vers sa trajectoire apprise en empruntant un chemin légèrement décalé (2) grâce aux bonnes capacités de généralisation du modèle.

performance conjointe dépasse en sensibilité celle du meilleur des deux [Bahrami et al., 2010].

Même si notre modèle de fusion permet d’augmenter la robustesse de la reconnaissance de lieux, nous n’avons pas traité dans ces travaux de la pondération à attribuer à chaque modalité. Une des raisons qui justifie le choix de notre modèle de fusion est qu’il permet une pondération aisée des modalités. Cependant dans les expériences précédentes, les 2 modalités possédaient un poids égal figé (0.5 chacune) de participation dans l’activation des cellules de lieu multimodales. Cette rigidité implique l’hypothèse d’une importance équivoque de chaque modalité dans la réalisation de la tâche, et ce quelque soit la situation. Il serait intéressant d’améliorer ce mécanisme de fusion en pondérant chaque modalité en fonction de sa pertinence pour la tâche dans un contexte donné. Par exemple, une privation visuelle (navigation dans le noir) devrait induire un poids nul de la modalité visuelle pour la reconnaissance de lieu.

Afin de réaliser cette pondération, le système aura besoin d’un mécanisme supplémentaire d’évaluation le renseignant sur la pertinence de ses modalités. Cette pertinence peut être apprise par

renforcement dans un contexte donné, mais également par l'ajout de modalités supplémentaires. Nous proposons ici l'ébauche d'un premier modèle plausible d'évaluation des sensations basé sur la prédiction intermodale.

5.4.1 Vers un modèle générique de prédiction intermodale

Le fonctionnement du cerveau n'est pas uniquement réactif, il est très souvent pro-actif [Berthoz and Petit, 2006]. En détectant les régularités dans un flux sensoriel donné et ses corrélations avec d'autres flux sensoriels, il est possible d'apprendre quelle aurait du être la valeur d'un signal donné grâce à un signal de substitution. Par exemple, le flux optique (mouvement apparent dans l'image) est souvent corrélé aux consignes de déplacement du robot. Nous proposons ici l'ébauche d'un modèle simple et générique de l'apprentissage de cette capacité à prédire l'activité de chaque modalité sensorielle à partir du flux sensoriel des autres.

Ce modèle présente une architecture régulière permettant l'apprentissage de compensations mutuelles des modalités sensorielles. Dans la section précédente, nous avons conclu que l'apprentissage d'un conditionnement entre cellules de lieu visuelles et cellules de grilles permettait d'obtenir des cellules de lieu proprioceptives cohérentes. Ce conditionnement a permis de résoudre de manière simple et élégante le problème de la fusion d'information multimodale. En partant de ce principe, nous proposons ici de généraliser le modèle à l'ensemble de stimuli sensoriels en apprenant de la même manière de simples conditionnements entre modalités (voir fig. 5.7).

Puisque nous désirons réaliser des prédictions sur la base de simples conditionnements, il est nécessaire de s'assurer que l'activité des entrées à prédire soit linéaire. L'algorithme de réduction des erreurs quadratiques moyennes (LMS), utilisé dans ces conditionnements, n'est pas capable de traiter des non-linéarités complexes puisqu'il ne permet d'estimer qu'un unique hyperplan séparant les données dans l'espace considéré. A l'instar du modèle de fusion bi-modale précédent, ce modèle requiert une étape de catégorisation afin de linéariser le problème. Chaque stimulus sensoriel est donc projeté sur une carte de Kohonen (ou un k-means) fournissant un codage en population de son activité analogique. Un neurone du champ de Kohonen code pour une valeur de l'entrée.

La sortie de chaque champ est ensuite utilisée comme stimulus inconditionnel par l'ensemble des autres champs.

Après apprentissage des corrélations intermodales, ce système bas-niveau devrait permettre d'augmenter la robustesse au bruit, de lever certaines ambiguïtés et de compenser les limitations éventuelles d'une modalité sensorielle de sorte que la couche de sortie fournisse la plupart du temps les signaux cohérents requis par les modules de plus haut niveau. Le système permettrait ainsi une entraide entre modalités à bas niveau afin de rendre invisibles, autant que possible, les insuffisances sensorielles aux yeux des modules de traitement présents en aval. Bien qu'il n'ait pas été testé formellement lors de nos travaux, ce modèle semble prometteur et pourrait permettre d'accroître la robustesse à tout niveau dans l'architecture : que ce soit au niveau des signaux sensoriels brutes, des traitements intermédiaires (carte de saillance, extraction d'amers), des cellules de lieu ou encore des stratégies sensori-motrices.

Dans le chapitre suivant, nous développerons cette idée de prédiction intermodale dans le cadre de l'auto-évaluation de stratégies sensori-motrices. Nous proposerons un modèle comparable

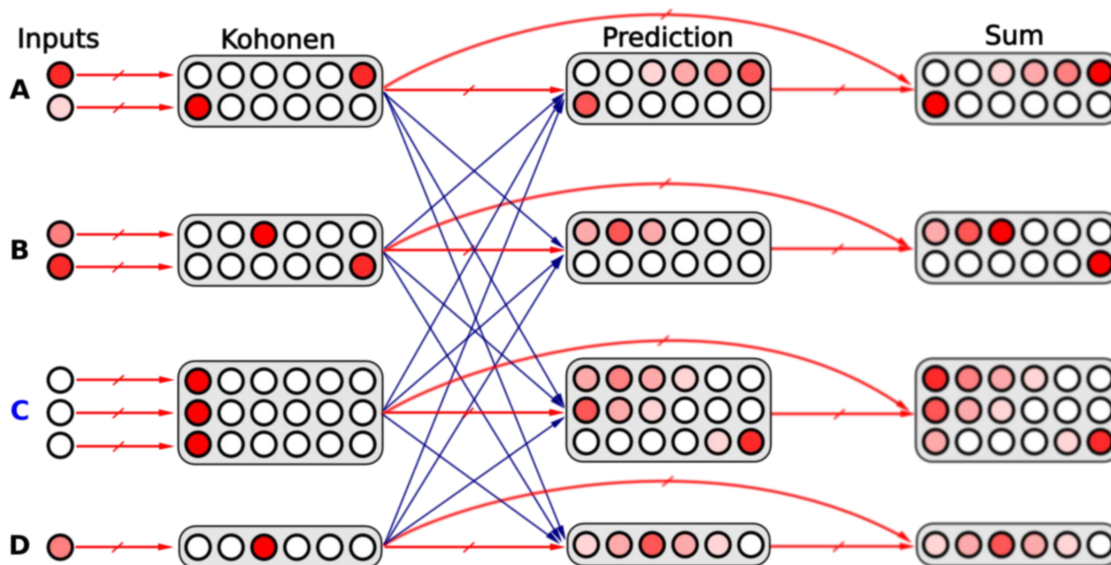


FIGURE 5.7 – Modèle générique de prédiction intermodale. L’architecture prend en entrée différents stimuli sensoriels dont la nature et la dimension peuvent varier (à titre d’exemple, le schéma présente 4 entrées, respectivement A, B, C et D). Chaque stimulus est projeté sur un champ de Kohonen (codage en population) afin de linéariser le problème. Puis, de simples mécanismes de conditionnements permettent aux différents champs de Kohonen de se prédire mutuellement. Nous ferons l’hypothèse d’une orthogonalité des entrées sensorielles au sein d’une même modalité. La modalité C renvoie ici des valeurs erronées dont l’impact est réduit par l’étage de prédiction intermodale, les autres modalités ayant appris à se compenser entre elles. La couche de sortie reste cohérente grâce aux signaux de substitution fournis par l’ensemble des modalités fonctionnelles.

exploitant non seulement l’information sensorielle mais également la notion d’action.

Publications personnelles

- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2012a). Multimodal integration of visual place cells and grid cells for navigation tasks of a real robot. In *From Animals to Animats 12*, number 7426 in LNAI, pages 136–145, Odense, Denmark. Springer
- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2011). Toward long range robot navigation. In *The International Workshop on Bio-Inspired Robots.*, Nantes, France
- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2012b). Multimodal integration of visual place cells and grid cells for robots navigation. In *Second Symposium on Biology of Decision Making*, Paris, France

- Tu sens quelque chose de bizarre toi ?
- Bizarre, peut-être pas. Je dirais plutôt cocasse ou simplement inattendu, voire impromptu, quoiqu'un tantinet insolite disons-le, dans la mesure du saugrenu, tout en étant singulièrement fantasque, presque excentrique si l'on y songe, et qui sait ? Extravagant en diable...

– A. Damasio "La Horde du Contrevent"

CHAPITRE 6

Un modèle générique de détection de nouveauté pour l'auto-évaluation

L'autonomie en robotique est encore un problème largement ouvert, pour lequel des concepts restent à inventer. Par autonome, nous entendons un système capable de développer et évaluer ses compétences afin de décider si son comportement est adéquat ou non au vue de la situation et de la tâche à accomplir. Ainsi, il faut inclure dans la notion d'autonomie comportementale, la capacité à apprendre des comportements à long terme, mais également le fait de pouvoir les évaluer afin de les adapter en fonction de la situation.

Les concepts de développement à horizon ouvert (*open-ended development*) et d'apprentissage cumulatif (*cumulative learning*) ont été étudiés depuis des années en psychologie, en robotique et en apprentissage automatique (*machine learning*). Ces capacités dépendraient en partie de ce qu'on appelle les *motivations intrinsèques*, un concept qui fait suite à la théorie présentée par [Deci, 1975] et enrichie par [Deci and Ryan, 1985, 2002]. Selon cette théorie, les motivations intrinsèques se distingueraient des motivations extrinsèques en ce qu'elles impliquent une "*action conduite uniquement par l'intérêt et le plaisir que l'individu trouve à l'action, sans attente de récompense externe*". Ces motivations seraient notamment impliquées dans l'exploration spontanée ou encore la curiosité.

L'étude des motivations intrinsèques est un champ de recherche qui prend de l'ampleur et attire de plus en plus l'attention ces dernières années, au vue des problèmes d'apprentissage cumulatif auxquels se heurtent les systèmes autonomes actuels. Bon nombre de modèles computationnels ont été proposés pour résoudre ces problèmes. Ces modèles sont divisés en 2 catégories : Les modèles basés sur l'accumulation de connaissances (*knowledge based*) et ceux basés sur l'accumulation de compétences (*competence based*).

Le premier modèle basé sur l'apprentissage de **connaissances**, proposé par Schmidhuber [Schmidhuber, 1991], consiste en un modèle du monde apprenant à prédire la perception future, connaissant les sensations actuelles ainsi que l'action entreprise par l'agent. Le progrès en prédiction est alors utilisé comme signal de récompense intrinsèque. Plus récemment, un modèle similaire de curiosité artificielle, proposé par Kaplan et Oudeyer [Kaplan and Oudeyer, 2007], permet à un système de se focaliser sur les stimuli nouveaux pour continuer d'apprendre à caractériser des situations de plus en plus complexes, en évitant les situations trop nouvelles (effrayantes) ou trop familières (ennuyeuses).

Le premier modèle de motivations intrinsèques basé sur l'apprentissage de **compétences** a été

proposé par [Barto et al., 2004]. Selon ce modèle, la récompense intrinsèque est attribuée en fonction de la capacité (ou l'incapacité) de l'agent à atteindre son but. De la même façon, Baldassarre et Mirolli [Schembri et al., 2007; Santucci et al., 2012] ont proposé une architecture d'apprentissage par renforcement qui implémente des compétences, ou savoir-faire sur la base d'un panel de modules experts. On pourra se référer à [M. Mirolli, 2013] pour un état de l'art complet sur ces 2 approches.

Dans nos travaux, nous ne nous intéressons pas directement au problème de l'apprentissage de nouvelles compétences dans un cadre *open-ended* (à horizon ouvert) et de manière totalement autonome. Nous proposons dans un premier temps d'étudier le développement à long terme dans le cadre de l'interaction avec l'humain. Selon ce point de vue, l'agent a besoin d'un professeur pour pouvoir apprendre une tâche. Cette tâche est apprise par démonstration, mais pas de manière prescriptive. Depuis longtemps, nous développons des modèles permettant à des robots d'apprendre de manière semi-autonome différentes tâches de navigation telle que la construction d'une carte cognitive pour être capable de rejoindre plusieurs buts [Gaussier et al., 2002; Giovannangeli et al., 2006; Gaussier et al., 2010; Cuperlier et al., 2007; Hasson et al., 2011; Hirel et al., 2013] ou l'utilisation de récompenses implicites ou explicites pour apprendre différents types de comportements sensori-moteurs [P. Andry, 2002; Andry, 2004]. Cependant, il apparaît inévitable qu'à certains moments le robot échoue dans la réalisation de sa tâche à cause d'un apprentissage incomplet ou bien de changements dans l'environnement. Pour des tâches compliquées, être autonome signifie également être capable de se rendre compte d'un comportement défaillant afin de demander de l'aide et ainsi apprendre d'autrui. Cela peut être réalisé de plusieurs façons : par un filtrage adéquat au niveaux des stimuli sensoriels (sélection de l'information pertinente : voir chapitre 3), par l'intégration d'information multimodale (voir chapitre 5) ou encore par l'ajout de différents niveaux d'imitation.

Nous proposons dans un premier temps de nous focaliser sur la question du robot amené à travailler pendant un certain temps sous la supervision d'un professeur afin de découvrir efficacement la solution à un problème complexe ou bien dans le but de compléter son apprentissage. Afin de répondre à la problématique des interactions naturelles de bas-niveau, l'apprentissage est réalisé par le biais d'une laisse. De la même manière que réagirait un chien ou encore un cheval, notre robot ne peut s'empêcher de tourner la tête dans la direction de la force appliquée à la laisse, ce qui permet de le guider assez intuitivement. La différence principale avec un système supervisé classique réside dans le fait que le robot doit évaluer quand et comment prendre en considération le signal de supervision. Pour des raisons de simplicité, nous supposons qu'il n'y a pas de contradiction entre les motivations du robot et les exigences du professeur (le robot est docile).

Nous ne nous focaliserons pas sur la question de la construction de carte cognitive à partir d'apprentissages latents, ni sur l'exploitation de divers signaux de renforcement modifiant les associations sensori-motrices du robot (voir [Hirel et al., 2013] pour plus de détails sur l'utilisation d'une architecture pour résoudre ce genre de problèmes).

Ajouter des capacités d'auto-évaluation aux robots semble être une solution intéressante dans le cadre de la robotique sociale, dans lequel l'humain joue un rôle clé dans le développement cognitif du robot. Cela peut amener le robot à rechercher plus d'interactions avec le professeur quand nécessaire, comme dans le système de contrôle collaboratif de Fong et al. [Nourbakhsh and Fong, 2003]. Le robot pourrait, par exemple, communiquer son incapacité à améliorer ses compétences. Cela signifie que ce dernier a non seulement besoin de coder ses

compétences mais également d'être capable de déterminer leurs limites. Cette capacité est d'autant plus importante pour les systèmes robotiques qui ont à prendre des décisions basées sur des observations provenant de plusieurs modalités [Zillich et al., 2011].

Évaluer les performances d'un comportement requiert, avant tout, (1) la capacité à prédire ce comportement. Ensuite, (2) cela requiert le fait de détecter des problèmes potentiels en prenant en considération des aspects de nouveauté dans ces prédictions. La **nouveauté** est donc un signal important à considérer puisqu'il fournit un premier retour sur la qualité du comportement. Selon [Téglás et al., 2011], le cerveau humain est capable d'inférer un modèle des interactions qu'il entretient avec le monde extérieur. Ce modèle est utilisé pour générer des anticipations et attentes sur les futures entrées. Un effet de surprise liée à l'erreur de prédiction apparaît alors lorsque ces attentes sont violées.

Selon [Wacongnea et al., 2011], ces prédictions sont organisées de manière **hiérarchique** dans le cortex. Dans une expérience auditive, les auteurs montrent que les erreurs de prédiction peuvent être locales ou globales : la présentation d'une note musicale anormale, au sein d'une suite de notes identiques, entraîne une erreur liée à la nouveauté perçue. Par contre, la présentation répétée de cette séquence peut être apprise comme étant normale, après quoi la présentation d'une séquence de notes sans anomalie (toutes notes identiques) déclenchera une erreur. Dans ce cas, la nouveauté globale est précisément l'absence de nouveauté locale, ce qui implique un niveau *méta* de prédiction. C'est l'absence de surprise locale qui constitue la surprise globale.

Le problème de l'auto-évaluation implique donc une forme de prédiction hiérarchique et inclue la notion de surprise, sensiblement proche de la classe des problèmes de détection de nouveauté. Dans ce chapitre, nous allons passer en revue un état de l'art sur les modèles de détection de nouveauté existants, après quoi nous tenterons de définir les briques fondamentales nécessaires à un système de détection de nouveauté dans le cadre de l'approche sensorimotrice. Nous proposerons un modèle générique et montrerons ses capacités de généralité au travers de tests impliquant deux comportements sensorimoteurs de natures différentes : la stratégie de navigation basée sur l'apprentissage d'associations lieu/action, présentée précédemment, et une stratégie bio-inspirée de suivi de route basée sur la reconnaissance du point de fuite présent dans l'image (A).

6.1 Modèles computationnels de détection de nouveauté

La détection de nouveauté est une technique couramment utilisée pour détecter qu'un stimulus est différent, à certains égards, des stimuli précédents. La capacité à reconnaître une situation perceptive inattendue est un outil essentiel pour les animaux, puisque cette situation peut représenter un prédateur potentiel ou une proie éventuelle. L'importance de la nouveauté dans les processus émotionnels a été suggérée par la théorie de l'évaluation (*appraisal theory*) [Lazarus and Kemper, 1992; Scherer, 1994; Lewis, 2005]. La nouveauté est liée de près à la surprise (qui peut être une émotion positive comme négative) mais est aussi déterminante dans les processus d'évaluations pour beaucoup d'autres émotions [Grandjean and Peters, 2011]. L'un des buts premiers de l'auto-évaluation est l'auto-protection. L'auto-évaluation est fortement utilisée dans des situations d'échec ou de menace survenue par le passé, où bien dans la prédiction d'une menace ou d'un défi futur [Taylor et al., 1995]. Elle permet de réduire la quantité importante d'information reçue par l'animal pour qu'il puisse se focaliser sur des stimuli inhabituels (voir [Marsland, 2002] pour un état de l'art global sur le sujet).

Certaines études proposent que la région hippocampique, outre son implication dans la navigation spatiale, pourrait être impliquée dans la détection de nouveauté, puisque le fait même d'identifier un événement nouveau implique de mémoriser les situations usuelles et de construire des attentes basées sur ces situations [Knight, 1996; Lisman and Otmakhova, 2001]. La modulation ACh (acetylcholine) du septum contrôlerait l'apprentissage de l'hippocampe et permettrait ainsi de décider de l'apprentissage d'une nouvelle forme [Hasselmo et al., 1995].

Un ensemble de filtres pour la détection de nouveauté a d'ores et déjà été proposé. Le fonctionnement de la majorité d'entre eux consiste en 2 étapes : (1) l'entraînement par l'apprentissage d'une base d'exemples (*training set*) contenant uniquement des informations valides, puis (2) la mise en évidence d'informations pathologiques, c'est à dire de données qui diffèrent de manière significative de la base d'exemple. Dans la littérature, on trouve différentes méthodes, comme la détection de valeurs statistiquement aberrantes (**statistical outlier detection**), la détection de nouveauté à partir de réseaux de neurones supervisés, les techniques basées sur les cartes auto-organisatrices (**self-organizing map**) ou encore les méthodes par dipôle fermé (**gated dipole methods**) :

L'approche standard pour la résolution d'un problème de détection de données aberrantes (*outlier detection*) [Hájek and Šidák, 1967; Devroye and Wise., 1980] consiste à estimer la distribution inconnue μ d'un ensemble de n variables aléatoires indépendantes pour ensuite détecter qu'une nouvelle entrée X n'appartient pas à cette distribution. De la même façon, la théorie des valeurs extrêmes (*extreme value theory*) [Gumbel, 1954] se concentre sur les distributions de données qui présentent des valeurs anormales dans les queues de la distribution qui génère les données.

Le tout premier filtre de détection de nouveauté adaptatif est certainement celui proposé par [Kohonen and Oja, 1976]. Ce filtre est en fait un algorithme de reconnaissance de formes où les nouvelles entrées sont comparées avec la forme apprise la plus proche, ce qui signifie que les signaux de sortie différents de 0 impliquent que l'entrée soit un stimulus nouveau.

Les réseaux auto-organisés fournissent également des solutions pour la détection de nouveauté en utilisant un apprentissage non-supervisé [Kohonen, 1993], en particulier la théorie de la résonance adaptative (*Adaptive Resonance Theory (ART)*) [Carpenter and Grossberg, 1988]. L'ART utilise un seuil de vigilance fixe comme élément de décision pour l'ajout ou non d'une nouvelle catégorie si aucune des catégories actuelles ne représentent correctement les données. Donc en un sens, un réseau ART propose une solution de détection de nouveauté fortement dépendante de son seuil de vigilance. De même, les réseaux de type NG (Neural Gas) [Martinetz and Schulten, 1991], GNG (Growing Neural Gas) [Fritzke, 1995] ou GWR (Grow When Required) [Marsland, 2002; Marsland et al., 2005; Neto, 2006] proposent des modèles de cartes auto-organisatrices capables d'ajouter de nouvelles catégories dynamiquement en fonction de la nouveauté présente en entrée. Le modèle permet à un robot patrouilleur de détecter les phénomènes anormaux présents dans son environnement à des fins de surveillance (et non d'auto-évaluation).

Les méthodes de réseaux de neurones supervisés proposent aussi une solution pour la détection de nouveauté en reconnaissant les entrées que le classifieur ne parvient pas à catégoriser correctement. De telles méthodes estiment la densité du noyau pour ensuite réaliser une détection de nouveauté selon un paradigme bayésien [Bishop, 1994; Roberts and Tarassenko, 1994].

Une autre solution est possible par le biais de champs de dipôles fermés (*gated dipole*), proposés par [Grossberg, 1972a,b], puis utilisés pour comparer des stimuli et modéliser l'attention que possède les animaux pour la nouveauté [Levine and Prueitt., 1992].

Les modèles neuronaux de la mémoire peuvent eux aussi être utilisés comme détecteur de nouveauté en apprenant une séquence d'états constituant un moyen simple de représenter un chemin au travers d'un environnement [Hasselmo and McClelland, 1999]. [Dollé et al., 2008] et [Girard et al., 2012] proposent des modèles de méta-contrôleurs pour la navigation spatiale capables de sélectionner à la volée la meilleure stratégie dans une situation donnée. Une compétition suivie d'un apprentissage par renforcement permet d'associer l'action qui convient le mieux à la situation. Le codage de contextes est donc nécessaire pour reconnaître la situation et se souvenir de l'action associée.

Ces différentes méthodes de détection de nouveauté se basent toutes sur l'analyse d'entrées sensorielles uniquement. Même si cette analyse s'avère suffisante dans le cadre de l'évaluation de problèmes impliquants une cause externe au robot (i.e. situation du patrouilleur en recherche d'un événement imprévu), elle reste incomplète dans le cadre l'auto-évaluation, c'est à dire l'évaluation de situations perceptives internes. Selon nous, le système doit apprendre à évaluer ses perceptions plutôt que ses seules informations sensorielles. Nous allons voir que bien souvent, le fait d'évaluer un comportement sensorimoteur revient à reconnaître une dynamique particulière entre sensation et action. Nous proposons donc d'ajouter la notion d'action comme entrée supplémentaire afin de permettre au système d'apprendre un modèle interne des dynamiques sensorimotrices de ses comportements.

6.2 Modèle des perceptions

A propos de la notion de perception

La perception est un phénomène actif, et l'on ne peut complètement séparer la sensation de l'action. Par exemple, lors de la préhension d'un objet, la trajectoire de la main est constamment réajustée en fonction de la perception d'abord visuelle puis tactile. On parle de boucle sensation-action.

L'importance de cette boucle peut être illustrée par l'expérience de [Held and Hein, 1963] : deux chats sont élevés dans le noir complet jusqu'à l'âge de dix semaines. L'un des chats est dit actif, l'autre passif. Ensuite, à raison de trois heures par jour, ils sont mis dans un dispositif tel que tout déplacement du chat actif provoque un déplacement identique du chat passif (le chat passif ne peut se déplacer seul). L'expérience montre que le développement du chat actif est normal, tandis que le chat passif n'a aucune perception de la profondeur. Heureusement, cette perception apparaît si le chat passif est rendu actif pendant 2 jours.

Dès 1945, les travaux de Merleau-Ponty sur la phénoménologie de la perception [Merleau-Ponty, 1945] récusent la notion de "*sensation pure*" qui ne correspond à aucune expérience vécue. Selon lui, les sensations sont avant tout relatives. Il s'accorde notamment avec la théorie de la Gestalt (la psychologie de la forme) [Smith, 1988a] selon laquelle les processus de la perception traitent les phénomènes comme des ensembles structurés (les formes) et non comme une simple addition ou juxtaposition d'éléments. L'approche dite "écologique" de Gibson [Gibson, 1979] s'accorde avec cette vision en proposant l'idée que l'être humain ne perçoit pas des unités d'information (image, son, pression, etc.) mais plutôt des flux sensoriels. C'est l'extraction des invariants de ces flux qui lui permet d'appréhender le monde qui l'entoure. La perception est

donc un phénomène fondamentalement actif qui consiste à agir de façon contrôlée (par exemple en déplaçant la tête), afin de générer des flux sensoriels et ainsi faciliter l'extraction de leurs invariants.

D'autres psychologues soulignent justement la place importante de l'action dans les processus de perception [Berthoz, 1997; Auvray et al., 2005;]. [Dewey, 1896] considère même l'action comme précédant la sensation : "Nous ne commençons pas à percevoir par le biais d'un stimulus sensoriel mais bien par une co-ordination sensorimotrice. En un sens, c'est le mouvement qui est premier et la sensation qui vient en second, le mouvement du corps, de la tête et des yeux déterminant la qualité de l'expérience. En d'autres termes, le véritable commencement est l'acte de voir ; le fait de regarder, et non la sensation lumineuse qui en résulte." (extrait et traduit de [Pfeifer and Verschure, 1994]).

Dès lors, l'évaluation d'un comportement sensorimoteur ne passe pas tant par la reconnaissance d'un pattern sensoriel particulier que par celle de la loi dynamique couplant sensation et action lorsque le système fait l'expérience de ce comportement. Les travaux de [Brooks, 1990] vont dans ce sens en proposant une modélisation de la perception par un ensemble de boucles sensori-motrices.

En suivant une approche similaire, la première brique de notre modèle encapsulera donc une représentation interne de ce couplage entre sensation et action. Le couplage définira un modèle interne minimaliste des interactions entre le robot et son environnement à partir duquel le système réalisera ses prédictions.

Modélisons les interactions dynamiques entre l'agent et l'environnement :

Apprendre à prédire les corrélations sensorimotrices d'une stratégie peut être vue comme le fait de trouver des invariants dans les perceptions du robot. Dans le domaine de la perception visuelle [Gibson, 1979], une affordance est définie comme la construction ou l'accès à un invariant caractérisant un comportement sensorimoteur particulier. Partant de ce prédicat, nous considérons la perception comme le résultat de l'apprentissage d'associations sensation/action laissant émerger un comportement global cohérent (voir [Gaussier et al., 2003, 2004; Maillard et al., 2005] pour une définition mathématique complète de la perception). Sur la base de cette hypothèse, nous définissons la perception au temps t_0 comme l'intégrale de la séquence d'affordances passées de l'agent. Une affordance correspond à un état sensation/action particulier.

$$Per(t_0) = \int_{-\infty}^{t_0} Sen^T(t).Ac(t).dt \quad (6.1)$$

Où Sen représente le vecteur des sensations (entrées sensorielles), et Ac un vecteur d'actions (donné par le retour proprioceptif de l'agent).

Notons y , comme Sen , un vecteur de n neurones y_i correspondants aux sensations de l'agent, $i \in N$. Cela peut être des cellules de lieu unimodales, multimodales ou des cellules de point de fuite (A) dans notre cas. y peut être vue comme un ensemble de variables aléatoires y_i . x est un vecteur de neurones x_i correspondant aux actions de l'agent (le neurone gagnant codant pour l'orientation du robot). Une matrice Per modélise la perception du robot en intégrant les sensations y et actions x au travers d'une fenêtre temporelle glissante définie par les liens récurrents de poids α . Per est en fait un produit tensoriel entre x et y sur lequel des liens récurrents sont

ajoutés. Cette matrice est une mémoire à court terme de la perception de l'agent. Le paramètre α détermine la taille de la fenêtre glissante et est défini arbitrairement en fonction de la rapidité de la boucle sensation/action de la stratégie considérée. $Per_{i,j}$ représente le n-uplet particulier des neurones x_i et y_j :

$$Per(t+1) = \alpha.Per(t) + (1-\alpha).Sen^T(t).Ac(t) \quad (6.2)$$

Le poids des liens récurrents α est inférieur à 1. En d'autres termes, cela signifie que les entrées récentes ont plus de poids dans le processus que les anciennes. Dans le cadre du traitement d'images, [Richefeu and Manzanera, 2004] propose un filtre différentiel de détection de mouvement utilisant un paramètre utilisé là aussi pour donner plus d'importance au passé proche qu'au passé lointain.

6.3 Modèle de détection de nouveauté

Dans la partie suivante nous allons détailler notre modèle générique de détection de nouveauté. Le modèle se base en partie sur les algorithmes de détection de nouveauté présentés précédemment. Comme la plupart de ces algorithmes, il consiste en 2 étapes : une phase de développement durant laquelle est réalisé l'apprentissage des corrélations sensorimotrices lorsque la stratégie est utilisée dans une situation usuelle (training set), puis une phase d'exploitation durant laquelle le système utilise sa capacité à détecter des motifs sensorimoteurs étrangers lors de situations inattendues.

Notre modèle interne de la dynamique du comportement est la base sur laquelle le système peut réaliser des prédictions et détecter de la nouveauté (voir fig. 6.1). On définit le vecteur \hat{y} , de taille identique à y comme estimateur de la moyenne. \hat{y} est l'espérance $E[y]$ des sensations y construite à partir de la matrice des perceptions Per , par le biais d'un algorithme en ligne de réduction des moindres carrés (*Least Mean Square* (LMS)[Widrow and Hoff, 1960]). A l'instar d'un conditionnement classique [Pavlov, 1927], le groupe de neurones \hat{y} va modifier à la volée le poids des connections provenant de la matrice de perception Per (stimulus inconditionnel (US)) dans le but d'estimer au mieux le vecteur de sensation y (stimulus conditionnel (CS)). Une différence absolue entre la sensation y et sa moyenne estimée \hat{y} définit un vecteur d'erreur instantanée $e = y - \hat{y}$. De même, un vecteur \hat{e} estime le moment d'ordre 1 ; $\mu_2 = E[e]$ de l'erreur courante e à partir de la matrice des perceptions Per par le biais d'un LMS réalisant son calcul en ligne. L'erreur d'ordre 2 est définie par $e_2 = e - \mu_2$. Le moment d'ordre 2 est également défini tel que $\mu_3 = E[|e - \mu_2|]$.

L'erreur d'ordre 3 est définie par $e_3 = e_2 - \mu_3$. La nouveauté η est définie comme la somme globale des neurones e_{3i} du vecteur d'erreurs d'ordre 3 : $\eta = |\sum_{i=1}^n e_{3i}|$. η peut finalement être résumée par :

$$\eta = E[|(|y - E[y]| - E[|y - E[y]|]) - E[|y - E[y]| - E[|y - E[y]|]]] \quad (6.3)$$

$$\text{avec } \|y\| = \sqrt{\sum_{i=0}^n y_i^2} \quad (\text{cas d'une norme } L^2) \quad \text{ou} \quad \|y\| = \sum_{i=0}^n |y_i| \quad (\text{cas d'une norme } L^1) \quad (6.4)$$

η représente l'erreur de prédiction du réseau, utilisée pour détecter des événements inattendus. On notera que la nouveauté η est ici définie comme une erreur d'ordre 3. Ce choix représente un bon compromis entre précision et latence. Nous discuterons de ce point en fin de chapitre (6.5). On notera également que les différents moments μ_2 , μ_3 and μ_4 représentent en fait respectivement une pseudo-variance, une pseudo-asymétrie (skewness) et un pseudo-kurtosis (coefficient d'aplatissement de Pearson) puisque leur mesure suit une norme L^1 au lieu de suivre une norme L^2 .

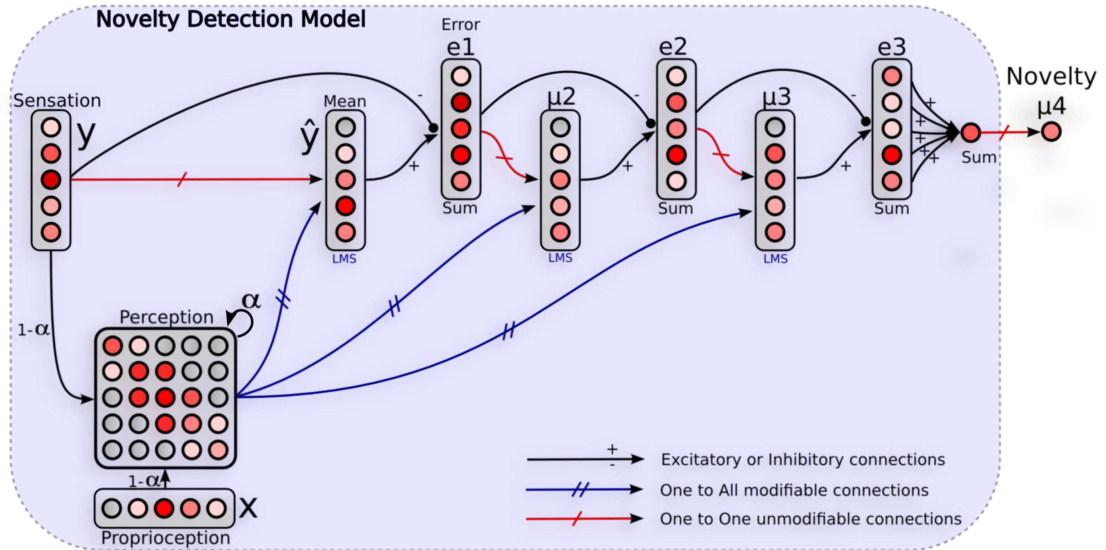


FIGURE 6.1 – Architecture neuronale de détection de nouveauté. Le réseau apprend les contingences sensorimotrices d'une stratégie donnée en apprenant à prédire la sensation courante à partir des perceptions passées. La perception est définie comme une mémoire à court terme des tuples sensation/action récents. Cette mémoire est utilisée pour approximer la moyenne, la variance et l'asymétrie statistique (skewness) du vecteur de sensations y . La nouveauté (novelty) est définie arbitrairement comme la somme des erreurs à l'ordre 3. Le niveau d'activité de cette erreur donne à l'agent un retour sur l'anormalité de la situation.

6.4 Expériences et résultats

Nous avons testé notre modèle dans différentes situations pour 2 stratégies différentes afin de valider la capacité du modèle à détecter la nouveauté. Ces 2 stratégies sont la navigation par associations Lieu/Action (voir 2.5.1) et une stratégie de suivi de route dont les détails sont donnés en annexes A. Les expériences suivantes ont été réalisées en simulation dans 2 environnements tridimensionnels différents. Nous avons utilisé une plate-forme robotique simulée de taille 40x40cm (une copie de la plate-forme Robulab de Robosoft) équipée de 2 roues motrices, de capteurs de proximité pour l'évitement d'obstacles, d'une caméra pan-tilt pour extraire les points d'intérêt dans la panorama visuel et enfin d'une caméra fixe pour la perception du point de fuite. (voir Fig.6.2 - A). Notre logiciel de simulation 3D (Webots de Cyberbotics) nous fournit un modèle physiquement réaliste du robot et des obstacles, mais ni le bruit, ni la simulation d'objets 3D à proximité des murs n'est prise en compte (des photographies 2D des murs de notre laboratoire servent simplement de textures aux murs virtuels).

Setup 1 : La stratégie Lieu/Action est activée (ON) tandis que la stratégie de suivi de route reste inactive (OFF). Le propos de l'expérience est de tester le mécanisme de détection de nouveauté sur le comportement engendré par la stratégie Lieu/Action. L'environnement est une salle virtuelle de 15x15m (voir Fig.6.2 - B) dont le sol est uniforme et les murs fournissent des points d'intérêt saillants. Le robot est entraîné par le professeur (apprentissage supervisé) à réaliser une trajectoire carrée par l'apprentissage d'associations Lieu/Action. Seulement 8 associations sont nécessaires ici pour réaliser la tâche de manière robuste. Une seconde salle, plus petite, communique avec la première. Cette seconde salle est inconnue du robot, puisqu'aucun apprentissage n'a été réalisé à l'intérieur de celle-ci. En conséquence, le fait de naviguer dans cette pièce résulte en des mouvements incohérents que le système devrait pouvoir détecter. Une fois la ronde apprise, le mécanisme d'auto-évaluation apprend à prédire les contingences sensori-motrices pendant que le robot accomplit sa ronde dans une situation normale. Cette phase est la phase d'entraînement (training set). Dans ce setup, le vecteur de sensation Sen n'est autre que le vecteur des 8 cellules de lieu apprises par le système. Le poids des liens récurrents est fixé empiriquement à $\alpha=0.95$, en se basant sur la fréquence des changements perçus dans les sensations. La boucle sensorimotrice de la stratégie Lieu/Action est assez lente puisque les états (cellules de lieu) ne se modifient que lorsque le robot navigue d'un lieu à un autre. Cet effet est lié en partie à la perte d'information due à l'utilisation d'une compétition stricte entre cellules de lieu et pourrait être diminué à l'avenir par l'usage d'une compétition plus souple. La fréquence de changement des états dépend principalement de la distance moyenne entre 2 lieux consécutifs ainsi que de la vitesse linéaire du robot. Pour notre problème, il faudra donc prendre un poids α proche de 1 (voir eq.6.2) pour bénéficier d'une grande fenêtre d'intégration temporelle. Trois tours ont été nécessaires au mécanisme d'évaluation pour prédire complètement les sensations à partir de l'ensemble des états sensorimoteurs perçus pendant la ronde. L'apprentissage est seulement terminé lorsque le degré de nouveauté atteint un minimum (typiquement un niveau inférieur à 0.4) et stagne dans chaque lieu. Ceci définit l'erreur de prédiction minimale que le système est capable d'obtenir sur cette tâche. Cette erreur minimale est directement liée au degré de profondeur (précision) du processus de prédiction (l'ordre n de l'erreur pour lequel nous définissons la nouveauté η). Comme nous l'avons précisé précédemment, le choix réalisé de définir la nouveauté comme l'erreur à l'ordre 3 implique que notre modèle n'est pas capable de caractériser des variations statistiques au delà d'une telle précision. Le fait de considérer les erreurs à l'ordre 4 et 5 dans le modèle devraient permettre ici d'apprendre respectivement la moyenne et la variance de la nouveauté à l'ordre 3. Le choix de définir la nouveauté comme l'erreur de prédiction à l'ordre 3 est un bon compromis entre latence et précision (voir discussion 6.5).

Setup 2 : La stratégie de suivi de route est activée (ON), tandis que la stratégie Lieu/Action est désactivée (OFF). Le but de l'expérience est de tester le mécanisme d'auto-évaluation sur la stratégie de suivi de route. L'environnement est un jardin virtuel de 40x40m (voir Fig.6.2 - C) avec une texture d'herbe au sol et d'arbre sur les murs. Le jardin est traversé par une route blanche dans sa diagonale. Le robot est capable de suivre correctement la route lorsque le point de fuite se situe dans son champ de vision. D'un autre côté, le robot est évidemment incapable de suivre quoi que ce soit de cohérent lorsqu'il perçoit uniquement de l'herbe ou des arbres. Dans ce cas, les mouvements proposés en sortie du réseau semblent aléatoires et le comporte-

ment n'est plus cohérent puisqu'il n'y a aucun point de fuite stable à suivre. Le mécanisme d'auto-évaluation apprend à prédire les contingences sensorimotrices pendant que le robot suit une route stable dans une situation normale. C'est la phase d'entraînement (training set). Le vecteur de sensation Sen est cette fois défini par 13 cellules de point de fuite. Le poids des liens récurrents est fixé empiriquement à $\alpha=0.7$, en se basant sur la fréquence des changements perçus dans les sensations. La boucle sensorimotrice de la stratégie de suivi de route est beaucoup plus rapide puisque les états (cellules de point de fuite) changent à une vitesse qui dépend directement de la vitesse de rotation du robot. Environ 200m (correspondant à 2 minutes d'entraînement) ont été nécessaires au mécanisme d'évaluation pour prédire complètement les sensations à partir de l'ensemble des états sensorimoteurs perçus lors d'un suivi de route. L'apprentissage est terminé lorsque le degrés de nouveauté atteint un minimum (un niveau inférieur à 0.4 comme précédemment) et stagne dans chaque lieu.

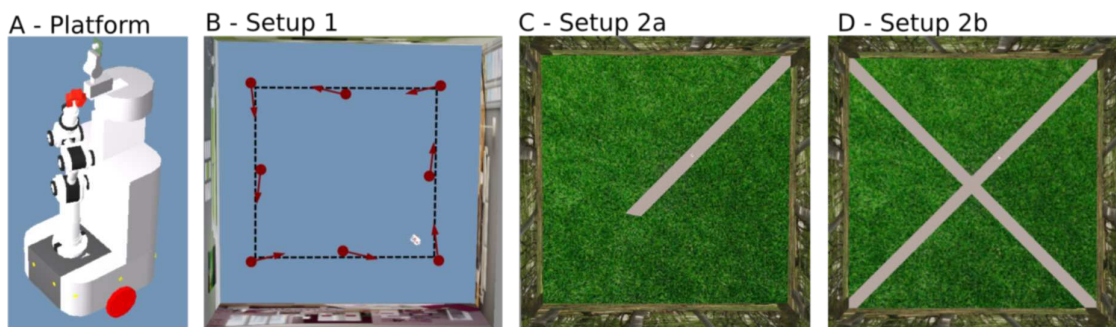


FIGURE 6.2 – Setup des expériences en simulation 3D. **A**– Plate-forme robotique utilisée. **B**– Setup 1 : test de la stratégie Lieu/Action. Le robot navigue dans une pièce de 15x15m. Il apprend 8 lieux associés à différentes actions (flèche rouge) pour reproduire une trajectoire carrée (trait en pointillés). **C**– Setup 2_a : test de la stratégie de suivi de route. Le robot navigue dans un jardin de 40x40m comprenant une route sur un sol d'herbe. **D**– Setup 2_b : Même environnement mais traversé cette fois de part en part par 2 routes selon ses diagonales.

Après que le système ait complètement appris à reproduire la trajectoire désirée (voir Setup 1, Fig.6.2) mais également à prédire les corrélations correspondantes, nous avons voulu tester les capacités du mécanisme de détection de nouveauté dans d'autres situations.

Dans une première expérience, nous avons testé la robustesse de la stratégie Lieu/Action dans une situation normale afin de déterminer le niveau minimum d'erreur que le système peut commettre (voir Fig.6.3 - A). Le robot effectue 12 tours standards sans perturbation. Les résultats montrent un comportement robuste et stable avec une trajectoire assez proche de la trajectoire désirée. Le niveau de nouveauté reste relativement bas puisqu'il ne dépasse jamais une valeur de 0.4, avec une moyenne de 0.2.

Dans une seconde expérience, nous introduisons volontairement un obstacle sur la trajectoire du robot de sorte qu'il soit obligé de l'éviter en utilisant sa stratégie d'évitement d'obstacles (voir Fig.6.3 - B). Grâce à une architecture de type *Subsumption*, une priorité réflexe est donnée à l'évitement d'obstacles. Le système contourne l'obstacle et converge de nouveau vers sa trajectoire grâce aux propriétés de généralisation des associations Lieu/Action apprises. Le niveau de nouveauté montre des pics d'activité lorsque le robot évite l'obstacle puisque l'orientation qu'il prend à ce moment précis ne correspond plus à celle apprise dans ce lieu. Ici, le modèle

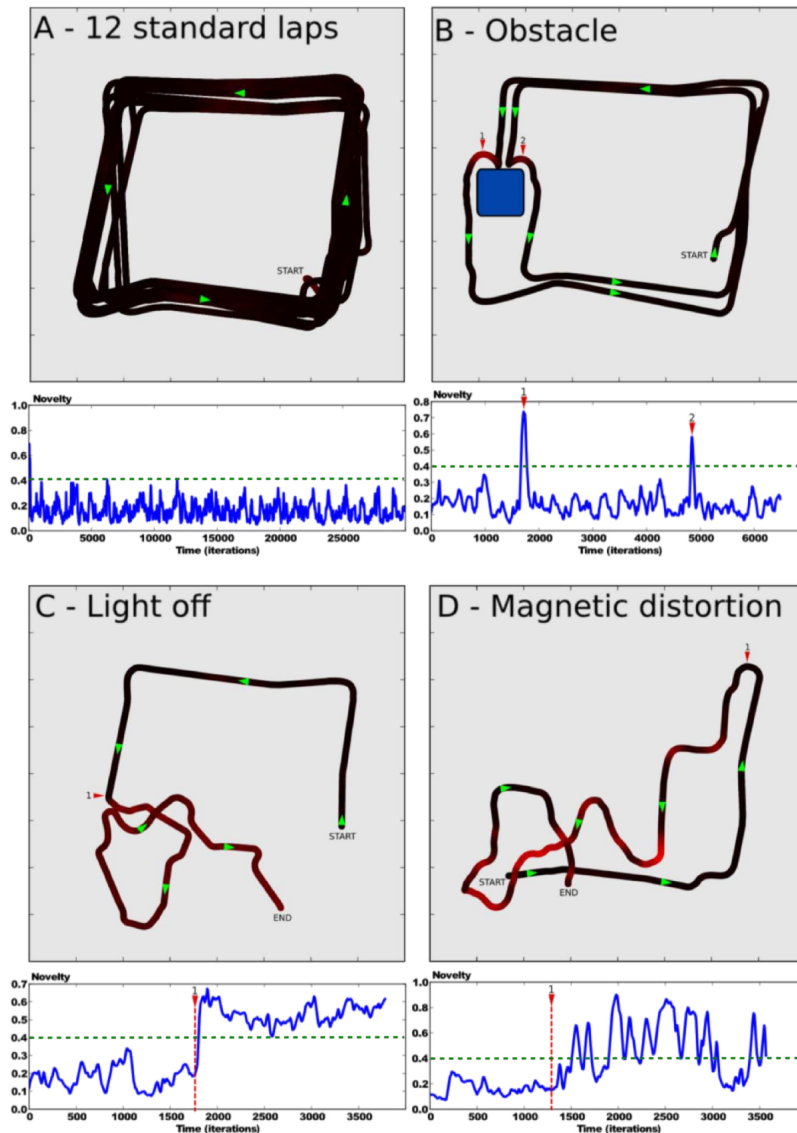


FIGURE 6.3 – Test de la détection de nouveauté sur la stratégie Lieux/Action (Setup 1 : voir Fig.6.2). **A – Haut** : Trajectoire du robot après 12 tours standards sans perturbation. **Bas** : Le niveau de nouveauté reste en dessous de 0.4 (trait en pointillés vert) avec une variance fixe puisque rien d’anormal ne survient. **B – Haut** : Un obstacle apparaît soudainement sur la trajectoire. La stratégie prioritaire d’évitement d’obstacle permet au robot de le contourner. Les propriétés de généralisation des cellules de lieux lui permettent de retourner sur sa trajectoire d’origine. **Bas** : La nouveauté montre un pic d’activité pour la première (1) et seconde (2) fois que le robot rencontre l’obstacle. **C – Haut** : Le robot réalise sa ronde correctement (de START à 1) quand la lumière s’éteint brusquement (1). Le comportement du robot devient aléatoire puisqu’aucun indice visuel n’est accessible. Il devient totalement perdu (de 1 à END). **Bas** : La moyenne de la nouveauté augmente de manière drastique. **D – Haut** : Le robot réalise sa ronde correctement (de START à 1) quand la direction du nord magnétique est pivotée soudainement de 90 degrés. **Bas** : La variance de la nouveauté augmente brusquement.

n’est pas capable de s’adapter à la nouveauté puisque les poids ne sont plus modifiés une fois la phase d’apprentissage terminée. Il sera intéressant à l’avenir d’améliorer le modèle pour lui permettre de s’adapter en permanence aux petites modifications de l’environnements.

Dans une troisième expérience, la lumière est soudainement éteinte (OFF) et le robot tente de continuer sa ronde dans le noir complet (voir Fig.6.3 - C). Le système visuel n'est alors pas capable de maintenir la cohérence de l'activité des cellules de lieu. Le robot devient totalement perdu. Ses mouvements sont globalement aléatoires. Le niveau moyen de nouveauté augmente sensiblement après que la lumière soit éteinte, mais sa variance reste plus ou moins identique. Cela peut s'expliquer par le fait que le système n'est plus capable de reconnaître aucun lieu, même si le système d'auto-évaluation tente de les prédire.

De la même manière, une quatrième expérience propose d'effectuer une rotation soudaine du nord magnétique de 90 degrés afin de simuler un sens de l'orientation défaillant (voir Fig.6.3 - D). Le robot effectue sa ronde lorsque la perturbation survient. Cette perturbation influe non seulement sur l'orientation suivi par le robot mais aussi sur la reconnaissance des lieux elle-même, puisque la position angulaire des amers visuels devient incohérente. Le comportement du robot est alors incohérent après quelques secondes. Le niveau de nouveauté montre alors de grandes variations d'amplitude puisque le robot suit des directions parfois inattendues et parfois prédites.

Finalement, une cinquième expérience propose de tester la capacité du système à détecter la nouveauté lorsque la stratégie de suivi de route est active (voir Setup 2, Fig.6.2). Dans une première phase (Setup 2_a), l'environnement contient une route qui s'arrête au milieu du jardin. Le robot commence son périple sur l'herbe dans le coin supérieur droit, orienté vers la route. Les résultats montrent que le robot converge vers la route et tend à s'aligner avec le point de fuite de celle-ci (voir fig. 6.4 - A). Il suit la route jusqu'à sa fin, après quoi il navigue dans l'herbe et son comportement devient alors aléatoire, puisqu'aucun point de fuite stable n'est accessible. Le niveau de nouveauté montre une décroissance progressive alors que le robot converge vers la route. Le niveau de nouveauté est minimum et assez stable lorsqu'il est aligné à la route. La nouveauté augmente progressivement lorsque le robot atteint la fin de la route. Elle reste grande pendant toute la phase de navigation incohérente sur l'herbe.

Nous avons également testé la robustesse de la détection pendant 1 heure de navigation simulée. L'environnement est une variante du précédent. Il est traversé de part en part par 2 routes dans ses diagonales (Setup 2_b). Les résultats montrent que le système est performant et peut distinguer les situations de suivi de route et de navigation dans l'herbe (voir fig. 6.4 - B).

6.5 Discussion

Dans les chapitres précédents, nous avons présenté comment notre modèle de cellules de lieu autorisait le robot à exhiber des comportements sensori-moteurs simples et robustes dans de petits environnements. Nous avons mis en évidence le besoin de trouver des solutions pour outre-passer les problèmes rencontrés lors de tentative de navigation dans des environnements plus complexes. Dans certaines situations, comme un milieu forestier par exemple, le nombre d'amers visuels disponibles et exploitables dans le panorama est très faible et le système visuel du robot traite alors des informations bruitées. De plus, le rapport signal à bruit tend à décroître avec le nombre de cellules de lieu recrutées. Dans ce genre d'environnements, un animal peut s'en remettre à d'autres stratégies pour naviguer, comme suivre un sentier ou naviguer à l'estime. Nous avons donc proposé de prendre en compte d'autres stratégies en étendant notre architec-

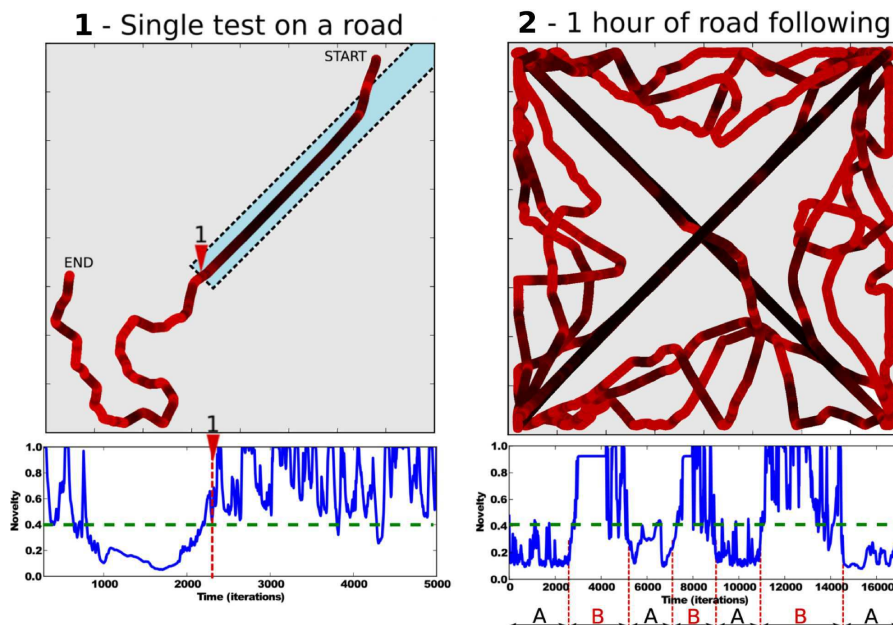


FIGURE 6.4 – Test de la détection de nouveauté sur la stratégie de suivi de route (Setup 2 : voir Fig.6.2). **Fig.1– Haut** : Trajectoire du robot sur un test unitaire (niveau de nouveauté : rouge=1, noir=0). Il n’y a qu’une seule route traversant le jardin et s’arrêtant au milieu de celui-ci. Le robot commence l’expérience dans l’herbe, dans le coin supérieur droit du jardin (START). Il converge vers la route puis la suit jusqu’à sa fin. Son comportement devient incohérent lorsqu’il quitte la route pour naviguer dans l’herbe, puisqu’aucun point de fuite stable n’est disponible. **Fig.1– Bas** : La nouveauté décroît progressivement au fur et à mesure que le robot converge vers la route. Elle est minimum lorsque ce dernier est aligné avec le point de fuite de la route. Enfin, elle augmente sensiblement lorsque le robot quitte la route. **Fig.2– Haut** : Trajectoire du robot après 1 heure de navigation autonome dans un jardin traversé par 2 routes. **Fig.2– Bas** : Évolution de la nouveauté pour une portion choisie de temps. L’activité de la nouveauté est en dessous de 0.4 (trait en pointillés vert) lorsque le robot suit une route (A) et réciproquement (B).

ture. L’ajout d’une stratégie de suivi de route robuste et biologiquement plausible permet au robot de converger naturellement vers un point de fuite et donc de suivre une route ou un chemin dans des situations où il n’est ni nécessaire, ni efficace d’utiliser un système de navigation à base d’associations Lieu/Action.

Ces 2 comportements fournissent une partie des compétences nécessaires au robot pour faire face à la plupart des situations rencontrées habituellement dans une tâche de navigation complexe. Mais ces comportements sont en compétition. Le robot a besoin d’un contrôleur au second-ordre qui sache donner la préséance à la meilleure stratégie en fonction de la situation. Un tel méta-contrôleur a besoin d’un mécanisme qui puisse évaluer ses comportements. Nous soutenons qu’afin d’évaluer les comportements, le système doit surveiller la nouveauté dans ses prédictions. Surveiller la nouveauté où les anomalies du comportement est donc une propriété clé pour mettre en place un système de meta-contrôle. Sur la base de cette hypothèse, nous avons proposé un modèle d’auto-évaluation basé sur la détection de nouveauté. Selon ce point de vue, le système doit, (1) apprendre à prédire ses sensations à partir de ses perceptions passées lors d’une phase d’entraînement, puis, (2) surveiller l’évolution de la nouveauté et répondre en fonction. En partant de théories en psychologie sur l’aspect actif de la perception, nous avons proposé un modèle interne des interactions sensorimotrices que le système entretient avec le monde extérieur. Ce modèle interne des perceptions offre une base générique pour réaliser des

prédictions sur les sensations de l'agent. Nous avons de bonnes raisons de croire que le modèle proposé est suffisamment générique pour pouvoir être adapté à l'évaluation de n'importe quel comportement sensorimoteur et ce pour un coût de calcul raisonnable, si l'on considère le fait que, la plupart du temps, sensations et actions sont corrélées (sauf cas pathologique). Cependant, comme la dimension du tenseur de perception peut être grande, il est important de définir des vecteurs d'entrée abstraits (par exemple quelques neurones d'amers plutôt qu'un flux visuel brut) afin d'empêcher une éventuelle explosion combinatoire.

Pour des raisons d'optimisation computationnelle, nous planifions de remplacer la matrice complète des perceptions par une matrice creuse (*sparse*) dans laquelle les produits seraient appris par des entités spécifiques (se référer à nos travaux sur la fusion des systèmes parahippocampal et perirhinal [Giovannangeli et al., 2006]). Pour aller un peu plus loin, le nombre d'états pourrait également être réduit si l'on abandonne la solution du codage par neurone "grand mère" (que ce soit au niveau des amers visuels ou des cellules de lieux) en la remplaçant par un codage épars. Cela permettrait d'utiliser certains aspects combinatoires à notre avantage. Une autre solution serait de représenter l'activité sensorielle par un code compressé à faible probabilité d'ambiguïté : une solution similaire à une fonction de hachage où une projection aléatoire de M vers N neurones avec $M \gg N$, à l'instar de la compression réalisée par notre modèle de cellules de grille (voir chap. 4).

Il faut aussi noter que le modèle ne peut pas, par lui-même, apprendre différents délais (timing) ou encore des phénomènes périodiques puisque le poids α des liens récurrents doit être paramétré empiriquement en fonction de la dynamique propre à chaque comportement. Mais le problème devrait être résolu si l'on considère une batterie de détecteurs de nouveauté, chacun avec une constante de temps différente.

Ensuite, nous avons proposé un modèle de prédiction des sensations basé sur l'estimation des différents moments par rapport à la moyenne. L'estimation est réalisée à l'aide d'un algorithme de réduction des moindres carrés (LMS). On notera que cet algorithme requiert des exemples aléatoires indépendants et identiquement distribués (*iid*) pour assurer sa convergence. En fait, même si de tels exemples sont inaccessibles dans nos expériences, la contrainte *iid* est négligeable ici. Grâce à la grande dimension de la matrice de perception, (1) le problème est linéaire (sauf cas pathologiques) et (2) le système n'a ni besoin d'un processus complexe de désapprentissage, ni d'une grande précision en sortie.

La nouveauté est ensuite mesurée en se basant sur la déviation de la perception actuelle par rapport à celle attendue. Elle est définie comme l'erreur de prédiction au niveau n . Dans nos expériences, nous avons définis la nouveauté comme l'erreur de prédiction à l'ordre 3. En réalité, le choix de l'ordre à partir duquel on parle de nouveauté est arbitraire. Il dépend principalement d'un compromis entre la latence du prédicteur et la précision désirée. Par exemple, définir la nouveauté comme l'erreur de prédiction à l'ordre 1 (une simple différence entre sensation actuelle et sensation moyenne apprise) résulte en une détection assez pauvre mais réduit le temps nécessaire au prédicteur pour apprendre à prédire la tâche. Avec un système aussi pauvre, des événements périodiques ou épisodiques seront toujours considérés comme nouveaux puisqu'ils diffèrent de la moyenne. Ils ne peuvent donc pas être prédit par le système puisque la variance n'est pas prise en compte.

Un prédicteur à l'ordre 2 permet de prendre en compte l'information de variance mais n'est souvent pas suffisant en pratique. Lors de la réalisation d'une ronde sensorimotrice, la variance d'une sensation peut être amenée à varier légèrement sans qu'il s'agisse d'un événement "nou-

veau" ou problématique pour le robot.

A l'inverse, définir la nouveauté comme l'erreur de prédiction à un ordre élevé donne au modèle une grande finesse de détection. Mais, dans ce cas, le prédicteur aura besoin de beaucoup de temps pour pouvoir complètement estimer chaque ordre (moyenne, variance, asymétrie, kurtosis, etc...). A cause de la contrainte d'un apprentissage temps-réel et de la mémoire limitée du modèle, l'estimation d'un moment particulier nécessite d'attendre la fin de l'estimation de chaque moment d'ordre inférieur. Ce point soulève plusieurs questions : Les animaux estiment-ils et prédisent-ils de la même manière ? Capturent-ils des aspects de nouveauté en caractérisant leur perception à l'aide de descripteurs statistiques similaires ? Possèdent-ils une quelconque latence d'estimation qui augmenterait avec le niveau de précision ?

Au delà de ces considérations, notre modèle présente 3 principaux défauts qu'il nous faudra résoudre dans un futur proche :

- Le poids α des liens récurrents, définissant une mémoire à court terme des perceptions de l'agent, doit être adapté d'une stratégie à l'autre. Ce paramètre dépend principalement de la dynamique propre de la stratégie impliquée. Ce paramètre pourrait être appris par le robot. Une autre solution serait de considérer une batterie de mémoires à court terme, chacune possédant une constante de temps propre.
- La taille du vecteur de sensation y a un impact direct sur la dynamique de l'activité de prédiction. L'usage d'une normalisation sur les entrées permet de fixer la borne maximale de nouveauté et ainsi éviter une saturation, cependant ce mécanisme n'est pas suffisant : l'impact de l'activité d'un neurone y_i sur le niveau de nouveauté est divisé par le nombre total de neurones de sensation y . Une solution serait d'utiliser une fonction d'activation non-linéaire (logarithmique ou sigmoïdale) au niveau de l'erreur de prédiction.
- Jusqu'ici, la phase d'apprentissage a toujours été séparée de la phase d'utilisation pour des raisons de simplicité. Cela nous mène au rôle premier des émotions qui est la régulation de l'apprentissage, par le biais de neuromodulations. Le système devrait donc pouvoir décider de lui-même quand apprendre une situation comme normale ou anormale. Ce problème s'avère en réalité très complexe et nous ne tenterons pas de le résoudre ici.

En résumé, nous avons présenté un modèle générique capable de détecter une potentielle anomalie dans son comportement. Le robot est désormais capable de se rendre compte que son comportement est inhabituel, cependant nous n'avons pas montré comment il pouvait faire usage de cette information pour adapter correctement son comportement. La détection d'anomalie n'a de sens que lorsqu'elle s'accompagne d'une régulation permettant de réduire la dissonance entre la situation perçue et la situation apprise.

Dans le chapitre suivant nous verrons comment un simple contrôleur émotionnel au second-order peut tirer parti de l'information de nouveauté afin de réguler le comportement ou encore de permettre au robot d'appeler à l'aide en cas d'échec critique.

Publications personnelles

- Jauffret, A., Grand, C., Cuperlier, N., Tarroux, P., and Gaussier, P. (2013d). How Can a Robot Evaluate its own Behavior ? A Neural Model for Self-Assessment. pages 1–8, International Joint Conference on Neural Networks (IJCNN)

"As-tu déjà entendu parler de politique chez l'insecte ? Moi non plus. L'insecte n'a pas de politique, il est très brutal, sans compassion ni compromis. On ne peut pas faire confiance à l'insecte. Je voudrais pouvoir devenir le premier insecte politicien. J'aimerais bien, tu sais. Mais ça me fait peur [...]. Je veux dire que je suis un insecte qui a rêvé qu'il était un homme, et qui aimait ça, mais le rêve est fini et l'insecte s'est réveillé [...]. Ce que je veux dire, c'est que je te ferai du mal si tu restes."

– D. Cronenberg "La Mouche"

CHAPITRE 7

De l'évaluation à la régulation

Dans le chapitre précédent, nous avons proposé un modèle générique d'auto-évaluation capable de donner au robot un premier retour sur la qualité de son comportement. Nous avons montré que ce modèle permettait de rendre compte de comportements anormaux lors de l'utilisation de deux stratégies de navigation de natures différentes. Cependant, le système ne faisait pas usage de cette information de "confiance" pour réguler son comportement. Dans ce chapitre nous proposons d'implémenter un contrôleur émotionnel dont le but sera d'interpréter l'évolution du niveau de nouveauté provenant du mécanisme d'auto-évaluation.

Nous proposons que le fait de considérer uniquement le niveau absolu de la nouveauté n'est pas suffisant pour prendre des décisions cohérentes et réguler correctement le comportement. Premièrement, de petites perturbations (petits obstacles, perturbation sur les capteurs, ambiguïté visuelle ou encore faux positifs) ne perturbent pas excessivement le comportement du robot. La plupart du temps, les propriétés de généralisation de la stratégie sensorimotrice permettent au robot de rester à l'intérieur du bassin d'attraction de la trajectoire apprise (voir Fig.2.12) et donc de réaliser la tâche correctement, même lorsque des événements inattendus surviennent. Il est donc plus intéressant de considérer l'évolution de la nouveauté plutôt que son niveau absolu. Le robot devra intégrer cette nouveauté au fil du temps et surveiller sa tendance pour être capable de juger de la pertinence de son comportement. Si l'erreur de prédiction reste élevée ou augmente, peu importe ce que le robot fait : le comportement devra être considéré comme inefficace. Et si cette inefficacité perdure, cela signifiera que le robot est coincé dans une impasse (*deadlock*). Une hypothèse similaire a été formulée par Schmidhuber [Schmidhuber, 1991] dans un système de contrôle basé sur la curiosité. Ce système propose une tentative de résolution du problème de la prise en compte des portions de l'environnement intrinsèquement imprévisibles. Il essaie de résoudre les tâches faciles avant de se tourner vers les tâches plus complexes. L'auteur propose d'apprendre à prédire les variations du cumul d'erreur plutôt que d'apprendre à prédire l'erreur elle-même.

Considérer l'évolution de l'erreur de prédiction dans le temps devrait permettre de fournir à notre système un retour plus précis sur la qualité de ses comportements. Dans ce chapitre, nous montrerons comment de simples inhibitions peuvent permettre au robot de réguler son comportement à l'instar d'un système homéostatique [Bernard, 1865]¹ poussant l'agent à entreprendre des actions dont le but est de diminuer les incohérences perceptives dont il fait l'expérience.

1. L'homéostasie est la capacité que peut avoir un système quelconque (ouvert ou fermé) à conserver son équilibre en dépit des contraintes extérieures.

[Ramachandran, 1998] montre que les douleurs "fantômes" résultant d'un membre amputé seraient essentiellement dues à l'incohérence constatée lorsque le sujet ordonne à sa main de bouger et qu'il constate qu'aucune réafférence ne vient confirmer ses attentes. Une expérience révélatrice propose d'intercaler un jeu de miroir devant le patient de manière à ce qu'il perçoive visuellement sa main fonctionnelle à la même place que sa main amputée. Lorsqu'on lui demande de bouger symétriquement les deux mains, la personne reçoit l'information visuelle attendue de l'action entreprise. Cette manœuvre permet de réduire les sensations de crampes insupportables provenant de muscles virtuels.

Dans ce cas de figure, la souffrance naîtrait d'une dissonance permanente (erreur de prédiction irréductible) entre les attentes du sujet et le retour perceptif de l'environnement. Nous sommes ici en présence d'une rupture perceptive, c'est à dire une incohérence forte et irréparable. Au lieu de créer une nouvelle cohérence comme dans l'illusion de Pinocchio vue précédemment, les échecs de création de sens engendrent des douleurs intenses entretenues par certaines régions du cortex pariétal [Flor et al., 2006].

La psychologue Françoise Schenk décrit la raison, non pas comme un système de logique froid, mais plutôt comme un processus d'entretien de cohérence profondément lié aux émotions [Schenk, 2009]. L'agent préfère ce qui est vrai et rejette ce qui est faux par une attitude rationnelle. Or ce processus même qui pousse à rejeter ce qui est incohérent s'accompagne de l'activation d'une région liée aux stimuli gustatif (insula antérieure). *"Tout se passe comme si la détection d'une affirmation incorrecte (2+2=5, par exemple) activait des circuits de rejets, tel le dégoût déclenché par une odeur désagréable"* [Harris et al., 2008].

Inversement, le constat de la "vérité" d'une affirmation déclenche des systèmes cérébraux situés dans le cortex préfrontal ventromédian, région qui serait associée à la gestion de récompenses intrinsèques.

Ces études tendent à montrer que le constat de la cohérence du monde prendrait une dimension fondamentalement émotionnelle. Schenk remarque alors une analogie intéressante avec la consommation d'aliments dans laquelle chaque aliment est d'autant plus valorisé qu'il comble un manque important. Le principe de cohérence élabore des hypothèses sur ce qu'il faut attendre de l'engagement de telle ou telle action, ce qui fait référence à la curiosité qui se distingue comme un besoin de compléter des informations pour rendre un ensemble plus cohérent. La consolidation des connaissances s'apparente à la consommation d'aliments, tandis que la curiosité serait un processus identique à l'appétit.

Partant de ces hypothèses, nous détaillerons en premier lieu notre contrôleur émotionnel capable d'analyser l'évolution de l'erreur de prédiction issue du mécanisme d'auto-évaluation pour réguler les différents comportements et déclencher un appel à l'aide lors de la détection d'un comportement inefficace chronique (deadlock). Nous présenterons ensuite les résultats expérimentaux obtenus en simulation grâce à ce contrôleur ainsi que nos premiers résultats issus d'expériences en environnement réel.

La modélisation des émotions semble indispensable à la fois au méta-contrôle de nos mécanismes cognitifs [Damasio, 1995; LeDoux, 1992; Panksepp, 1998] mais aussi pour communiquer de manière pré-verbale nos états internes.

De nombreux travaux en psychologie et neurobiologie montrent l'importance d'introduire des mécanismes émotionnels au coeur même des systèmes d'Intelligence Artificielle [Damasio, 1995; Picard, 1997; Canamero and Gaussier, 2004]. Ces processus émotionnels sont particulièrement importants pour la prise de décision puisqu'ils permettent de guider ou biaiser le

comportement plus rapidement que ne le permettent les processus rationnels, ou lorsque les inférences sont insuffisantes [Damasio, 2003]. Griffiths propose une taxonomie des émotions divisées en deux classes : les émotions primaires gérées par l'amygdale, et les émotions cognitives dont le siège serait plutôt situé dans le cortex préfrontal [Griffiths, 1997].

L'approche bio-inspirée basée sur des modèles émotionnels pour le méta-contrôle est encore peu explorée ([Balkenius et al., 2009; Canamero and Gaussier, 2004; Mannella et al., 2008; Hasson et al., 2011]).

Le rôle des émotions dans la communication est aussi très important et a été étudié dans le cadre des interactions entre enfants et adultes [Tronick, 1989]. Les enfants montrent, très jeunes, des capacités d'auto-évaluation (avant 2 ans) lorsqu'ils effectuent une tâche. Mais par contre, ils montrent peu d'intérêt pour l'approbation des parents et se tournent vers un autre but en cas d'échec. Ce n'est qu'à partir de l'âge de 2 ans que l'enfant montre des réactions claires (pleurer, s'accrocher aux parents) lorsqu'il est confronté à un échec déclenchant une auto-évaluation négative [Stipek et al., 1992; Kelley et al., 2000].

7.1 Un modèle de contrôleur émotionnel au second-ordre

Comme nous l'avons précisé plus haut, il semble plus intéressant de considérer les tendances du niveau de la nouveauté plutôt que son niveau absolu. Le système doit donc être capable de surveiller le progrès et la régression de l'erreur de prédiction au fil du temps. En se basant sur des travaux précédents [Hasson et al., 2011], nous proposons de calculer le progrès instantané $P(t) = \eta(t - \delta_t) - \eta(t)$ ainsi que la régression instantanée $R(t) = \eta(t) - \eta(t - \delta_t)$ comme les dérivées au premier ordre du signal de nouveauté $\eta(t)$. Un progrès désigne donc une hausse du niveau de nouveauté, et réciproquement pour une régression.

Définissons $F(t)$, un potentiel analogique de *frustration*, comme un neurone doté d'un lien récurrent. Ce neurone intègre les quantités de progrès $P(t)$ et de régression $R(t)$ instantané(e)s, ainsi que le niveau absolu de nouveauté $\eta(t)$ pour prendre en compte une éventuelle stagnation de ce niveau (voir fig.7.1). Ce potentiel de frustration sera appelé **niveau de stress** (*Distress*), dénoté $D(t)$. Les actions proposées par la stratégie correspondante sont inhibées proportionnellement à ce niveau de stress. La frustration est ensuite définie comme une décision binaire $F(t)$:

$$F(t) = \begin{cases} 1 & \text{if } D(t) > T \\ 0 & \text{otherwise} \end{cases} \quad (7.1)$$

avec T un seuil et $D(t)$ le niveau de stress défini tel que :

$$D(t) = \alpha_D D(t - \delta_t) + \beta_S S(t) + \beta_R R(t) - \beta_P P(t) + \beta_\eta \eta(t) \quad (7.2)$$

$S(t)$ est une récompense provenant du professeur et $\alpha_D, \beta_S, \beta_P, \beta_\eta$ sont les poids des différentes variables. Le neurone binaire $F(t)$, représentant la frustration, ne décharge que lorsque le niveau analogique de stress dépasse le seuil T (fixé arbitrairement à 0.9 dans les expériences suivantes). Cette frustration déclenche une procédure pré-codée (ad hoc) d'appel à l'aide : Le robot s'arrête et émet un signal sonore d'appel à l'aide afin de communiquer sa détresse. Le professeur pourra alors l'aider à se sortir de cette situation en lui montrant quoi faire. Ce système permet au robot d'améliorer son apprentissage dans les situations non-maîtrisées.

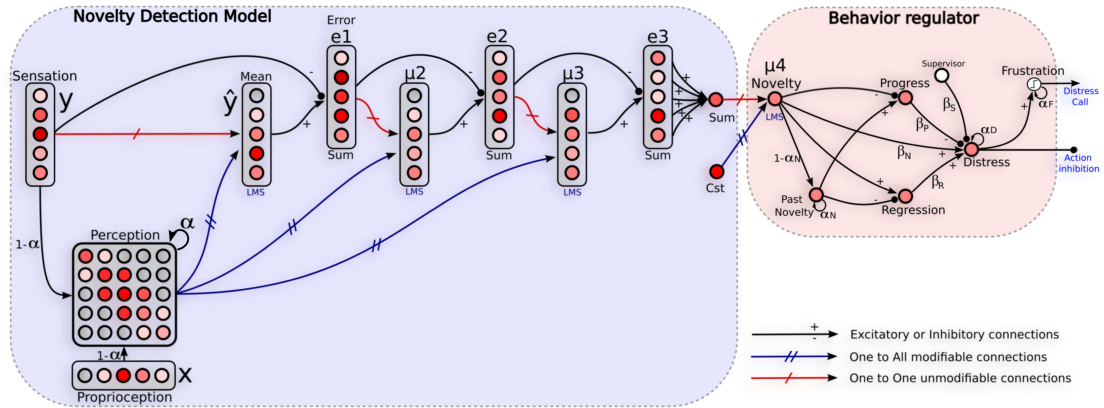


FIGURE 7.1 – Architecture neuronale d’auto-régulation. **Gauche** : Réseau utilisé pour la détection de nouveauté. **Droite** : Contrôleur émotionnel exploitant la nouveauté afin de réguler le comportement. Un neurone avec rétroaction, représentant le niveau de stress (Distress) du système, intègre les progrès et régressions instantané(e)s de l’erreur de prédiction. Le niveau absolu de nouveauté est également pris en compte pour détecter une éventuelle stagnation. Les actions proposées par la stratégie correspondante sont inhibées proportionnellement au niveau de stress. Un neurone (frustration) décharge uniquement lorsque le niveau de stress dépasse un certain seuil. Cette frustration permet au robot de déclencher un appel à l’aide. α_x et β_x sont les poids représentant les paramètres du régulateur.

7.2 Sélection et coopération de stratégie

Dans notre architecture, chaque stratégie et leur méta-contrôleur respectif fonctionnent en parallèle comme des canaux indépendants (voir Fig.7.2). Chaque stratégie propose une action (une orientation à suivre) dans un champ de 361 neurones. Chaque neurone du champ code pour une orientation particulière. Chaque champ d’action est modulé en amplitude (inhibé) proportionnellement au niveau de stress perçu pour la stratégie correspondante. Ainsi, l’amplitude de l’action d’une stratégie est liée au niveau de confiance instantanée attribuée à la stratégie en question. Les champs d’action des stratégies sont fusionnés dans un champ de neurone dynamique global (DNF). Comme expliqué au chapitre 1, le DNF fournit un compromis entre compétition (sélection) et coopération, plutôt qu’une compétition stricte comme le ferait un réseau WTA. Le DNF apporte en plus des propriétés de stabilité et de filtrage [Amari, 1977]. Les propriétés du champ ont été testées avec succès pour contrôler le mouvement d’un bras de robot par imitation, en utilisant un tracking visuel des mouvements [Gaussier et al., 1998; Andry, 2004], ou encore pour le contrôle de la navigation de robots mobiles [Schöner et al., 1995; Quoy et al., 2003].

De même, les ganglions de la base ont des propriétés de persistance permettant une économie d’énergie [Girard et al., 2003], propriétés inexistantes lors de l’utilisation d’un mécanisme de compétition simple (WTA). On peut illustrer l’intérêt de ce principe par le *paradoxe de l’âne de Buridan* [Gratet-Duplessis, 1851]. Selon ce paradoxe, si un agent ressent une faim légèrement plus forte que sa soif et qu’il commence à manger, sa faim diminue et l’agent se retrouve alors à ressentir une soif plus importante que sa faim. Il décide donc d’aller boire, et ce même s’il a encore faim. Le paradoxe amène à des alternances perpétuelles entre deux ressources situées à des lieux différents, ce qui induit un gaspillage d’énergie. Les ganglions de la base permettent de s’affranchir de ce problème par l’ajout d’une persistance de l’action en cours qui permet une économie de déplacements et donc d’énergie. Cet exemple illustre l’intérêt de l’utilisation d’un modèle de ganglions de la base pour la sélection de buts mais il reste tout aussi valable pour la

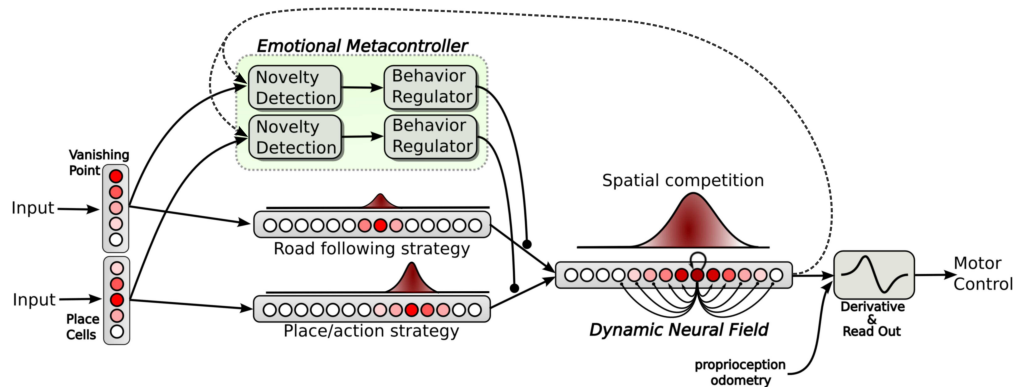


FIGURE 7.2 – Architecture neuronale de sélection de l’action. **Bas** : Chaque stratégie propose une action dans un champ de 361 neurones (chaque neurone codant pour une orientation particulière). Les champs d’action sont fusionnés au sein d’un champ de neurone dynamique. Ce champ offre une solution au compromis sélection/coopération pour la prise de décision, en plus d’être doté de propriété de stabilité et de filtrage temporel. **Haut** : Chaque contrôleur émotionnel apprend à prédire le comportement normal d’une stratégie particulière à partir des sensations et actions prodiguées par la stratégie. Un niveau de stress, dépendant de l’erreur de prédiction, est utilisé pour inhiber proportionnellement les actions.

sélection de stratégies dont il est question dans ce chapitre. On notera que ce paradoxe provient d’une légende et n’implique donc pas nécessairement une absence de ganglions de la base chez l’âne !

7.3 Expérience : La frustration et son rôle communicatif dans le développement de l’autonomie

7.3.1 Résultats en simulations

Dans cette expérience, nous mettons en évidence la nécessité d’un mécanisme de frustration pour demander de l’aide en situation de détresse. Le robot a appris à réaliser une ronde dans une pièce (voir Setup 1, Fig.7.3). Durant la première période, le robot réalise sa ronde dans une situation normale, sans perturbation (voir Fig.7.4 - A). Les résultats montrent un comportement robuste avec une trajectoire stable et proche de celle désirée. Le niveau de stress reste relativement bas (en dessous de 0.4), avec un niveau moyen en dessous de 0.3. Ce niveau représente l’erreur de prédiction minimale que le système est capable de réaliser pour cette tâche dans une situation normale.

Après un certain temps, le professeur interfère avec le robot pour dévier sa trajectoire vers la seconde pièce qui lui est inconnue (voir Fig.7.4 - B). Le robot essaie de réaliser sa ronde en prenant l’orientation associée avec la cellule de lieu gagnante. Le niveau de stress augmente progressivement puisqu’il n’y a plus de cohérence comportementale : la séquence sensori-motrice perçue est très différente de celle apprise.

Après peu de temps, le niveau de stress atteint un seuil de frustration (0.9 dans cette expérience) qui déclenche une procédure ad-hoc d’appel à l’aide. Le robot s’arrête et émet un son caractéristique pour demander de l’aide. Le professeur répond à l’appel et vient aider le robot à se sortir de cette situation en le tirant par la laisse dans la direction requise pour sortir de cette pièce

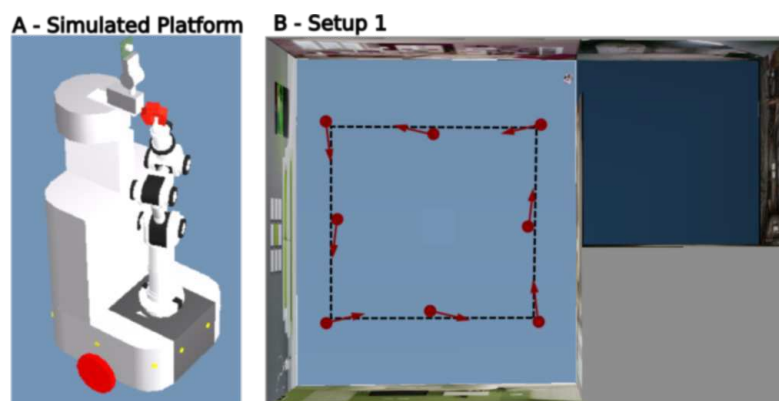


FIGURE 7.3 – Setup de l'expérience de test du mécanisme interne de frustration. **A**– Plate-forme robotique utilisée. **B**– La ronde est identique à celle utilisée dans le setup 1. Une seconde pièce, plus petite, est ajoutée à droite de la première. Cette pièce est inconnue du robot (aucun lieu n'y a été appris). Le but de l'expérience est de dévier intentionnellement le robot pendant qu'il réalise sa ronde afin de l'amener dans la pièce inconnue. Le robot devrait alors évaluer que son comportement est imprévisible, conclure qu'il est perdu et enfin appeler à l'aide.

(voir Fig.7.4 - C). La correction du professeur est perçue comme un signal poussant le robot à apprendre une nouvelle association Lieu/Action. Le mécanisme de prédiction apprend alors cette nouvelle situation comme normale. Le robot parvient dès lors à sortir de la seconde pièce et à retourner dans la première. Le niveau de stress décroît rapidement puisque l'interaction avec le professeur agit comme un signal inhibiteur dans notre modèle (voir fig. 7.1, lien "*supervisor*"). Une fois retourné dans la première pièce, le robot converge à nouveau vers l'attracteur stable qu'est sa trajectoire apprise.

Dans l'expérience suivante, le robot est placé dans la seconde pièce, dans un endroit différent de celui appris (voir Fig.7.4 - D). Comme le robot a déjà rencontré cette situation par le passé, il sait quelle orientation emprunter cette fois pour s'en échapper. Grâce aux propriétés de généralisation des cellules de lieu et à la stratégie d'évitement d'obstacle, il parvient à retourner dans la première pièce. Le niveau de stress reste bas puisque la situation est désormais considérée comme normale (apprise et prédite). Néanmoins, lorsque le robot atteint la frontière entre les deux pièces, son comportement reste hésitant. Ceci est en fait dû à des ambiguïtés visuelles et à un apprentissage trop pauvre à cet endroit. Les deux lieux les plus proches sont associés à deux actions contradictoires (aller à droite, aller à gauche). Le pavage des cellules de lieu pourrait être raffiné à cet endroit mais il n'est pas nécessaire finalement pour réaliser la tâche, puisque le robot ne se frustre pas et parvient à retourner sur son attracteur. La phase d'hésitation ne dure pas assez longtemps pour que le niveau de stress atteigne le seuil de frustration. Le robot a donc bien pu tirer parti de l'interaction qu'il a lui même initié avec le professeur, puisque son apprentissage lui permet désormais de sortir de la seconde pièce de manière autonome, sans intervention d'un tiers.

7.3.2 Expérience sur robot réel

L'expérience suivante se déroule dans un environnement intérieur réel (une partie de notre laboratoire). Nous avons utilisé une plate-forme robotique similaire à la version simulée (voir Fig.7.5 - A). Le but ici est de tester les capacités du mécanisme d'auto-évaluation dans une sit-

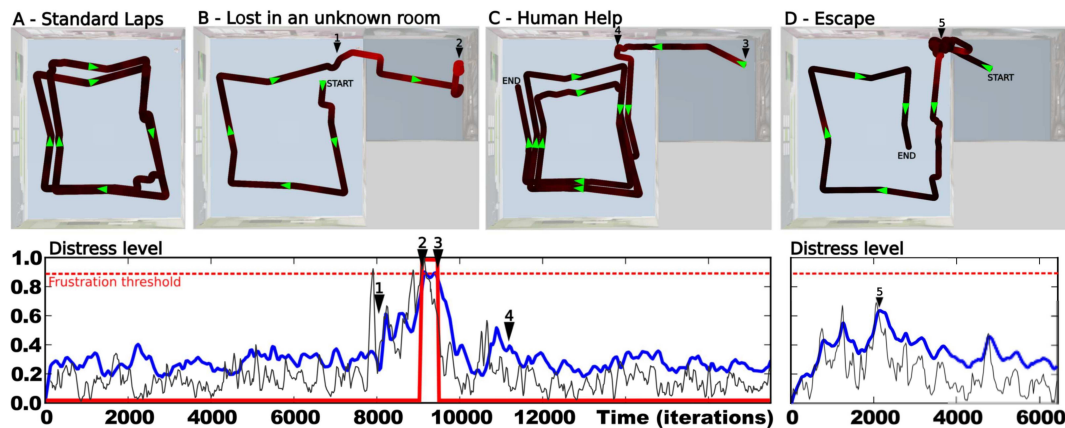


FIGURE 7.4 – Simulation démontrant l'intérêt d'un mécanisme interne de frustration (voir Setup 1, Fig.7.3). **Haut** : **A** – Trajectoire du robot pendant la réalisation de sa ronde dans une situation normale (sans perturbation). **B** – Le professeur perturbe intentionnellement le robot (1) en le déviant dans la seconde pièce (inconnue du robot). Le robot tente de réaliser sa ronde sans succès. Le niveau de stress augmente progressivement au vue de l'incohérence de la nouvelle séquence sensorimotrice perçue. Finalement, le stress atteint un seuil de frustration et le robot s'arrête et appelle à l'aide (2). **C** – Le professeur vient aider le robot en le tirant par la laisse dans la direction permettant de sortir de cette pièce (3). Le robot apprend cette nouvelle association Lieu/Action. Une fois retourné dans la première pièce, il converge naturellement vers son bassin d'attraction stable et continue sa ronde. **D** – Le robot est désormais capable de tirer parti de la démonstration du professeur pour quitter la seconde pièce de manière autonome. **Bas** : Évolution du niveau de stress (courbe bleue) et de frustration (courbe rouge) dans le temps. Le stress augmente lorsque le robot entre initialement dans la pièce inconnue (1). Il atteint un seuil de frustration (pointillés rouges) (2), puis revient à un niveau normal après l'intervention du professeur (4). Le stress augmente lorsque le robot se trouve entre les 2 pièces (5). Cette augmentation indique qu'un raffinement d'apprentissage est possible ici, bien que non-nécessaire, puisque le robot n'est pas frustré et parvient après quelques essais à rejoindre sa ronde d'origine.

uation réelle avec toutes les contraintes que cela implique (lumière ambiante changeante, mouvement de personnes ou d'objets, bruit de mesure, etc...) L'environnement est composé de deux pièces connexes de tailles assez similaires, reliées l'une à l'autre par une porte et un couloir (voir Fig.7.5 - C,D,E). Nous testerons ici le mécanisme de frustration pour le raffinement de l'apprentissage et non la régulation d'un panel de comportements. Pour cela, seule la stratégie Lieu/Action est active. La tâche du robot est de réaliser la ronde complète en passant par la première pièce, puis par le couloir, puis par la seconde pièce, pour enfin revenir dans la première. La ronde est donc une boucle. 14 associations Lieu/Action ont été nécessaires pour obtenir un comportement correct et réaliser la ronde de manière autonome (voir Fig.7.5 - F). A l'instar d'une interaction homme/chien stéréotypée [Giovannangeli and Gaussier, 2010], le professeur utilise une laisse pour guider le robot dans la direction voulue (voir Fig.7.5 - B). Le robot apprend un nouvel état Lieu/Action lorsqu'il détecte une différence entre les ordres du professeur et sa commande propre. Cette détection module le niveau de vigilance du robot. Elle autorise le recrutement d'une nouvelle cellule de lieu et l'apprentissage d'association entre lieu et direction à suivre. De cette manière, le robot est capable d'apprendre la trajectoire que le professeur lui montre. Cet apprentissage proscriptionnel (corriger les erreurs du robot plutôt que de prescrire le chemin) est nécessaire pour obtenir un bassin d'attraction stable et robuste.

Afin d'éviter que le robot ne soit frustré inutilement dès le départ, le mécanisme d'auto-évaluation est désactivé lors du premier tour (voir Fig.7.7). Nous avons choisi, pour réduire le temps d'apprentissage, d'obtenir un comportement relativement stable sur l'ensemble de la

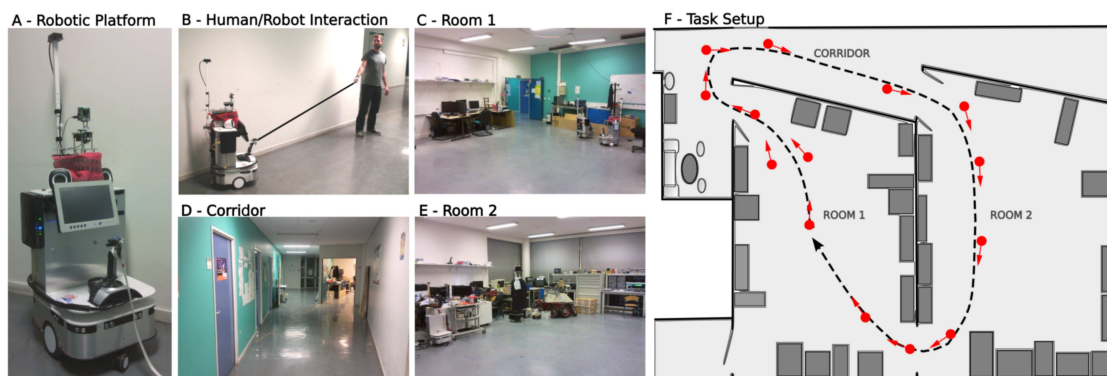


FIGURE 7.5 – Setup de l'expérience de frustration sur robot réel. **A** – Plate-forme robotique réelle utilisée lors de l'expérience (RoboLab de Robosoft). **B** – Apprentissage supervisé : Le professeur utilise une laisse pour guider le robot dans la direction désirée. Le robot obéit au professeur et apprend la trajectoire de façon autonome. **C, D, E** – L'environnement comprend 2 pièces de taille identique ainsi qu'un couloir. **F** – Trajectoire apprise par le robot. Celui-ci commence sa ronde dans la première pièce, passe par le couloir, traverse la seconde pièce, puis retourne dans la première. 14 Associations Lieu/Action (flèches rouges) ont été nécessaires pour reproduire de façon robuste la ronde désirée (ligne en pointillés noirs).

ronde avant d'autoriser le système d'auto-évaluation à apprendre.

Le second tour correspond à une phase de méta-apprentissage intense, puisque chaque lieu est nouveau pour le prédicteur (voir Fig.7.6 - A). Le niveau de stress montre un pic pour chaque lieu. Le stress décroît à partir du 3ème tour puisque les états Lieu/Action sont désormais prédit. Après cette période d'entraînement, nous laissons le robot réaliser sa ronde quelques temps, en corrigeant sa trajectoire seulement si nécessaire (voir Fig.7.6 - B). Sans perturbation particulière, le niveau de stress reste bas (en dessous de 0.5), sauf dans une petite portion de l'environnement où il demeure important. Ce lieu singulier correspond à une zone où le robot passe proche d'une fenêtre et se voit éblouit par le couché de soleil au moment de l'expérience. Le niveau de stress plus important dans cette zone signifie que le robot possède une confiance médiocre en son comportement lorsqu'il est ébloui. Quoiqu'il en soit, ce manque de confiance n'est pas suffisant pour déclencher une frustration, d'autant plus que le comportement du robot reste cohérent durant cette phase.

Quelques temps plus tard, le professeur intervient pour dévier le robot dans une pièce jusqu'alors inconnue de ce dernier (voir Fig.7.6 - C). De même que pour l'expérience de frustration en simulation, le niveau de stress augmente progressivement et atteint finalement le seuil de frustration. Le robot s'arrête et appelle à l'aide. Le professeur répond à sa requête en lui montrant le chemin à suivre pour sortir de cette pièce. Le robot apprend ce nouvel état Lieu/Action et retourne sur sa ronde stable.

On notera aussi qu'à la fin de l'expérience, la zone dans laquelle le robot avait le moins confiance est maintenant cohérente avec les prédictions du système car le soleil venait de se coucher et ne perturbait plus la caméra.

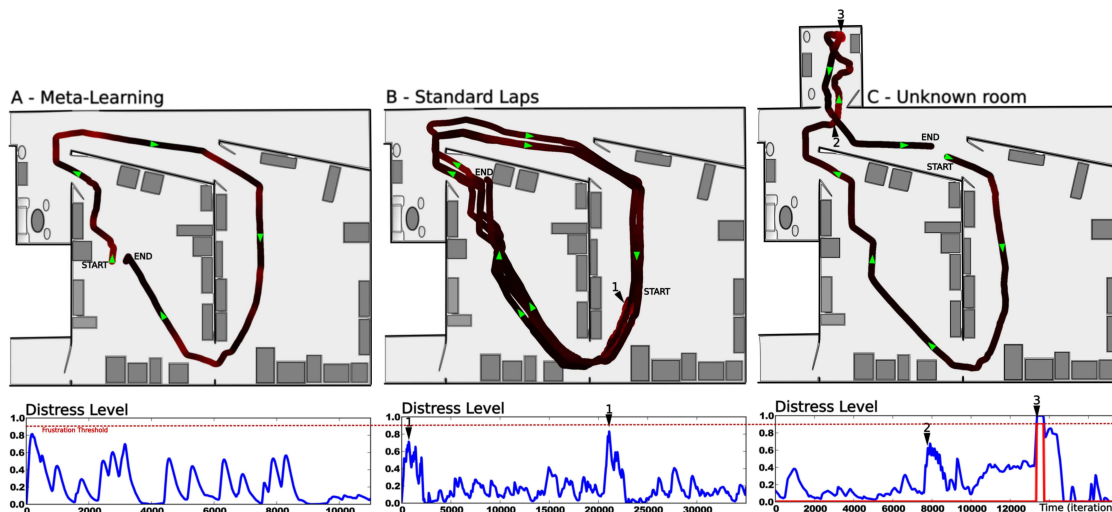


FIGURE 7.6 – Expérience de frustration sur robot réel (voir Fig.7.5). **A**– Trajectoire du second tour, durant la phase de méta-apprentissage. Le robot a déjà appris la ronde et apprend maintenant à la prédire. On note des pics de stress réguliers, chaque fois que le robot quitte un lieu prédit pour entrer dans un lieu inconnu. Le niveau de stress décroît au fur et à mesure qu’il apprend à les prédire. Globalement, chaque état sensorimoteur est prédit après un tour. **B**– Après cette phase d’apprentissage, on laisse le robot réaliser sa ronde de manière autonome, sans perturbation. Le niveau de stress reste bas sur l’ensemble de la trajectoire excepté dans une zone particulière (1) dans laquelle le robot est ébloui par le couché de soleil. Le robot est moins confiant dans cette zone, sans pour autant être frustré. **C**– Plus tard, le professeur dévie le robot dans une pièce inconnue (2). Ce dernier tente de réaliser sa ronde sans succès et finit par être frustré (3). Le professeur l’aide alors en lui montrant la voie vers la sortie, après quoi le robot est capable d’exploiter cette démonstration pour s’échapper de la pièce sans aide extérieure.

7.4 Expérience : Le stress et son rôle actif dans la régulation des comportements

7.4.1 Résultats en simulations

Cette fois, les deux stratégies (Lieu/Action et Suivi de route) sont actives et fonctionnent en parallèle. Le propos de l’expérience est de tester la coopération des stratégies dans un environnement complexe. Naviguer dans un environnement large et complexe requiert un méta-contrôleur pour faire coopérer les différentes stratégies d’une manière cohérente. L’environnement est un mélange entre le Setup 1 et 2 (voir Fig.7.8). Un jardin est maintenant connexe aux deux pièces et une route permet de les relier en passant par ce jardin. Le robot est entraîné à réaliser le tour complet : passer par les deux pièces puis le jardin. Notre modèle permet au robot de réaliser la ronde par l’apprentissage de 7 associations Lieu/Action seulement. En fait, le système n’a pas besoin d’apprendre de lieu dans le jardin puisque suivre la route est une stratégie suffisante ici pour réaliser la tâche désirée par le professeur. Ainsi, le professeur n’a pas besoin de corriger le robot dans cette zone puisque son comportement est déjà celui désiré.

Chaque stratégie propose une action (une orientation) dont l’amplitude est modulée en temps réel par l’évaluation que le système fait de cette stratégie. Les actions proposées par chaque stratégie sont ensuite fusionnées au sein d’un champ de neurone dynamique qui contrôle les mouvements du robot (voir 7.2). Il est important de noter que la procédure d’appel à l’aide suite à une frustration est désactivée dans les expériences qui suivent. Puisque nous désirons tester les

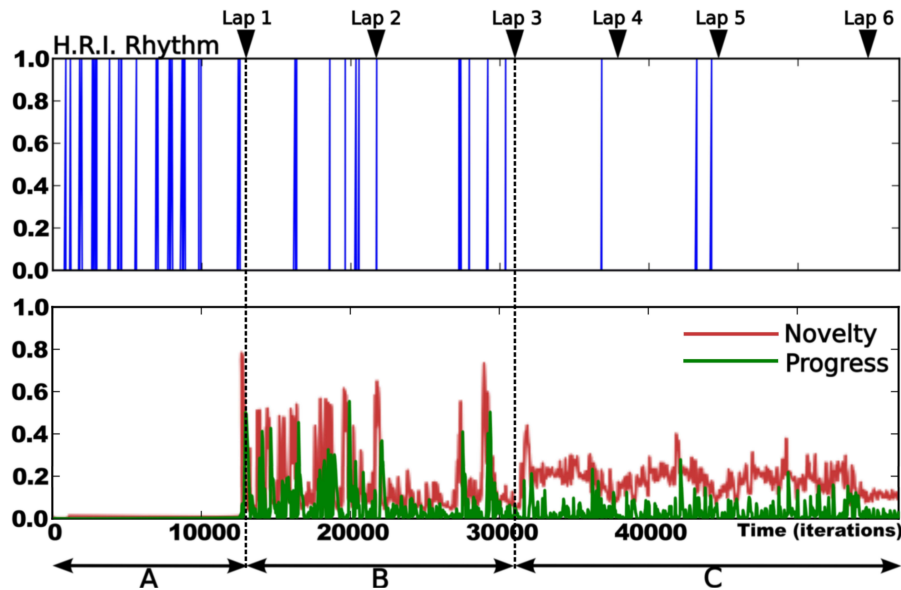


FIGURE 7.7 – Détails de la phase d'apprentissage sur les 6 premiers tours. **Haut**– Rythme des interactions homme/machine (Human Robot Interaction : HRI). Les impulsions de type Dirac correspondent aux instants de guidage. La fréquence des interactions décroît au fil du temps puisque le robot apprend à être autonome. Cette fréquence donne une mesure directe de l'autonomie du robot (l'autonomie est inversement proportionnelle à la fréquence des interactions). **Bas**– Niveau de Nouveauté (en rouge) et Progrès (en vert). On observe 3 phases : **A**– correspond à la session d'apprentissage de la ronde (premier tour). La fréquence élevée des impulsions indique que le professeur est prescriptif/directif puisque le robot ne connaît rien de la tâche à accomplir. L'apprentissage du prédicteur est désactivé durant cette phase. **B**– correspond à la phase de prédiction (2ème et 3ème tours). Le professeur évalue le comportement du robot et le corrige seulement si nécessaire. Le méta-contrôleur est activé et apprend à prédire la séquence sensorimotrice. Par conséquent, les niveaux de nouveauté et de progrès sont hauts durant cette phase. **C**– correspond à la phase finale de l'apprentissage. Le robot est presque autonome mais de rares corrections sont encore nécessaires à certains endroits pour raffiner la trajectoire ou élargir le bassin d'attraction.

possibilités de compétition et coopération des stratégies, le robot ne doit pas déclencher d'appel à l'aide si l'une ou l'autre de celles-ci engendre un signal de frustration.

Dans une première simulation, nous avons voulu tester chaque stratégie séparément pour s'assurer de l'incapacité du système à réaliser la tâche avec une seule d'entre elles. Lors du test de la stratégie Lieu/Action (voir Fig.7.9 - A), les résultats montrent que le robot réalise sa ronde correctement dans les deux pièces, mais tombe dans une situation de deadlock lorsqu'il navigue dans le jardin, ce qui l'amène à être frustré. De la même manière, lors du test de la stratégie de suivi de route (voir Fig.7.9 - B), le robot parvient à suivre la route correctement dans le jardin mais tombe dans une situation de deadlock lorsqu'il entre dans la première pièce, à la suite de quoi il se retrouve frustré.

Dans une seconde simulation, le robot réalise un tour de manière autonome (voir Fig.7.10 - A). Il commence sur la route, au milieu du jardin. Les niveaux de stress de chaque stratégie reflètent à quel point leur actions sont inhibées. Ainsi la stratégie Lieu/Action est fortement inhibée lorsque le robot navigue dans le jardin et est très activée en intérieur. Réciproquement, la stratégie de suivi de route est fortement activée lorsque le robot fait face à la route mais inhibée

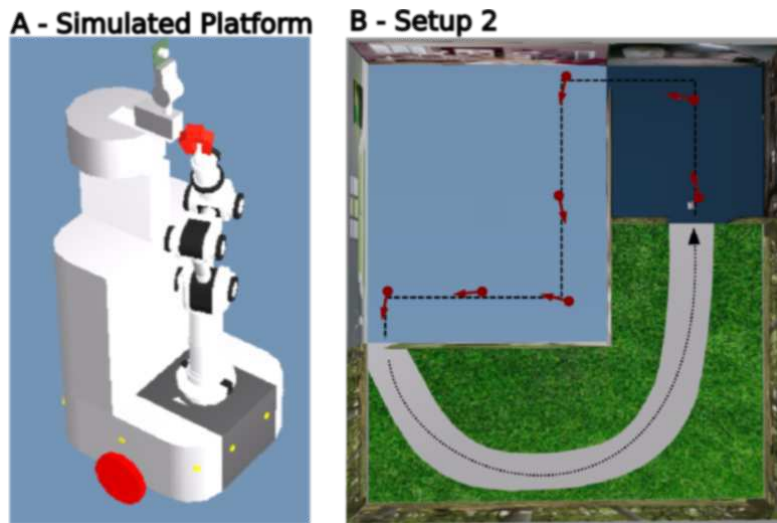


FIGURE 7.8 – Setup de test de la régulation des comportements en simulation. **A** – Plate-forme robotique utilisée. **B** – Ronde de test de la sélection et coopération de stratégie. Un jardin est ajouté, reliant les 2 pièces par une route. Seulement 7 associations Lieu/Action sont nécessaires pour réaliser le tour complet (trait en pointillés noirs). Le robot n’a pas besoin d’apprendre quoique ce soit dans le jardin puisque le suivi de route est suffisant pour réaliser la ronde.

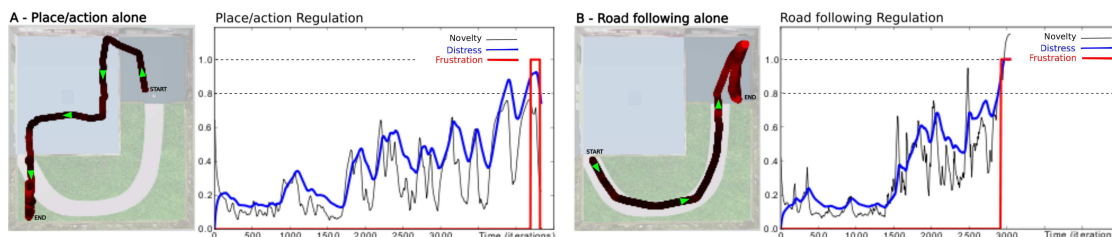


FIGURE 7.9 – Test unitaire des stratégies. (Setup 2, Fig.7.3). Une seule stratégie à la fois est testée afin de s’assurer que la tâche ne peut pas être résolue sans la coopération des 2. **A** – Test de la stratégie Lieu/Action seule. Le robot commence dans la petite pièce. Le comportement est cohérent jusqu’au moment où le robot entre dans le jardin, puisqu’aucun lieu n’a été appris dans cette partie de l’environnement. Le niveau de stress atteint le seuil de frustration. **B** – Test de la stratégie de suivi de route seule. Le robot commence au début de la route. Son comportement de suivi de route est cohérent jusqu’au moment où il entre dans la petite pièce, puisqu’aucun point de fuite cohérent n’est alors perçu. Le niveau de stress atteint là aussi le seuil de frustration.

la plupart du temps dans les deux pièces.

Le robot réalise le premier tour correctement mais échoue à la toute fin. Il tombe dans un dead-lock lorsqu’il entre à nouveau dans le jardin. Ceci est dû à l’orientation que le robot a appris à suivre en sortant de la pièce. Si cette orientation est trop différente de celle de la route, le système n’est tout simplement pas capable d’évaluer le suivi de route comme étant pertinent. Ce fait soulève un problème fondamental de la régulation des comportements par auto-évaluation, à savoir que l’évaluation nécessite un certain temps afin de pouvoir rendre un jugement fiable sur la stratégie évaluée. La stratégie, quant à elle, requiert la confiance du mécanisme d’évaluation pour pouvoir s’exprimer.

Dans une 3ème simulation, le robot commence sa ronde dans la première pièce. En sortant dans le jardin, il parvient à converger vers la route après quelques hésitations, mais ne parvient

pas à entrer dans la petite pièce (voir Fig.7.10 - B). Le robot parvient à converger vers la route principalement par chance. Il échoue cependant lorsqu'il doit entrer dans la petite pièce puisque les propriétés de généralisation des cellules de lieu permettent à celui-ci de se croire dans la pièce avant même d'y être entré (il tourne trop tôt et tombe dans un minimum local). Ce problème est lié à un apprentissage trop pauvre à cet endroit et peut être résolu en apprenant un nouvel état Lieu/Action à la fin de la route pour assurer une convergence vers l'entrée de la pièce.

Dans une 4ème simulation, le robot parvient à réaliser 3 tours de manière autonome (voir Fig.7.10 - C). Les résultats montrent que, la plupart du temps, les évolutions des niveaux de stress des deux stratégies sont en opposition de phase. Le niveau de stress de la stratégie Lieu/Action reste bas en intérieur puisque les cellules de lieux perçues sont bien prédites. Le stress est important en extérieur puisqu'aucun point d'intérêt visuel n'est disponible et que de toute façon, aucun lieu n'a été appris dans cet environnement. Réciproquement, le niveau de stress de la stratégie de suivi de route est bas lorsque le robot suit la route mais reste relativement haut en intérieur puisqu'il est alors difficile de percevoir un point de fuite cohérent. Les niveaux de stress sont directement proportionnels à l'inhibition des stratégies correspondantes. Le comportement est donc principalement dirigé par le suivi de route en extérieur et par la stratégie Lieu/Action en intérieur, même si des coopérations peuvent apparaître à certains moments.

En allant justement plus loin que ces résultats prévisibles, on peut dire que ces expériences montrent de bonnes propriétés émergentes de la synergie des deux stratégies. A vrai dire, nous avons rencontré beaucoup de situations pour lesquelles la coopération améliore la performance obtenue avec une seule stratégie. Cela se produit généralement lorsque le robot doit passer une porte. Dans la plupart des cas, le contraste créé par les contours d'une porte ouverte sur une autre pièce est perçue comme un point de fuite cohérent par le robot. Celui-ci montre donc une tendance naturelle à converger vers ce genre de portes, ce qui lui permet de passer d'une pièce à l'autre sans besoin de recruter plusieurs états Lieu/Action. Il faut noter qu'avant la mise en place de ce système, le passage de porte était une tâche difficile pour nos plate-formes robotiques puisque cette tâche demandait d'apprendre un bassin de plusieurs lieux convergeants vers la porte ainsi qu'un évitement d'obstacle très précis. Pour finir, ces résultats soulignent aussi quelques problèmes rencontrés durant les phases de transitions entre stratégies. Le professeur doit faire attention à bien guider le robot dans la direction requise pour percevoir le point de fuite de la route, sans quoi le suivi de route ne peut pas être évalué comme correct et le robot finira par être frustré. Ceci est dû principalement au délais requis par le méta-contrôleur pour évaluer une stratégie. Il en résulte un comportement inefficace où le robot tente de passer d'une stratégie à l'autre sans succès.

7.5 Discussion

Dans ce chapitre, nous avons abordé deux rôles différents que peut jouer notre mécanisme d'auto-évaluation dans le cadre d'une navigation robotique complexe sur de grandes distances. Nous avons présenté son rôle de régulateur dans la gestion des comportements en fonction d'une situation, et son rôle social dans la communication de sa propre frustration pour sortir d'une impasse et raffiner son apprentissage.

Les résultats montrent que notre modèle de détection de nouveauté présente certaines pro-

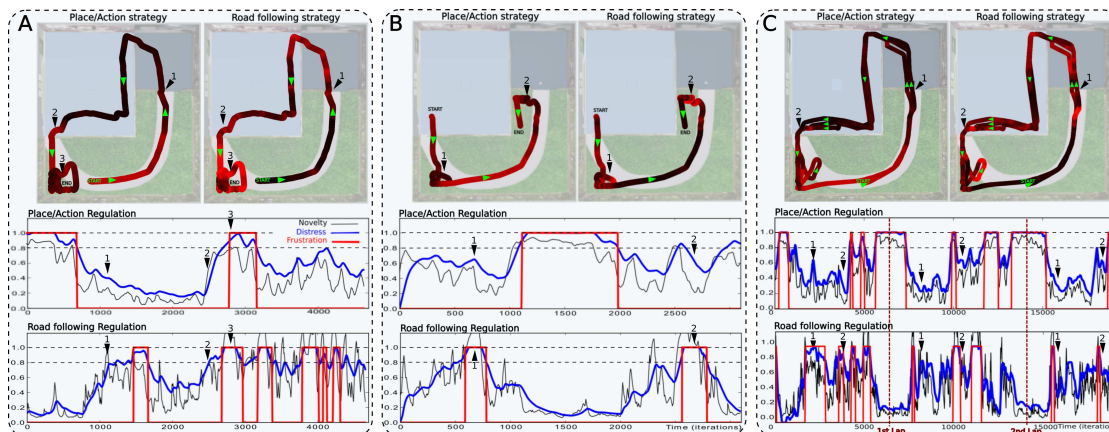


FIGURE 7.10 – Simulations soulignant l'intérêt d'une coopération entre stratégies. (voir Setup 2, Fig.7.3). **A**– Le robot commence sur la route dans le jardin et réalise un tour de manière autonome. Il échoue à la fin lors de sa sortie dans le jardin et tombe dans une situation de deadlock (3). Le niveau de stress est proportionnel au niveau d'inhibition de la stratégie correspondante. La stratégie Lieu/Action est fortement inhibée lorsque le robot navigue en extérieur (de START à 1) et fortement active à l'intérieur (de 1 à 2), et réciproquement pour la stratégie de suivi de route. A la fin du tour, le robot entre dans le jardin mais ne parvient pas à converger vers la route à cause de son orientation (de 2 à 3). **B**– Dans cet autre exemple, le robot parvient à converger vers la route par chance après quelques hésitations (1). Il la suit correctement mais tourne trop tôt et ne parvient pas à entrer dans la petite pièce (2). **C**– Le robot parvient cette fois à réaliser 3 tours en autonomie. (1) et (2) indique respectivement les moments où il entre et sort. Les niveaux de stress des 2 stratégies sont, la plupart du temps, en opposition de phase. La fin de chaque tour est indiqué par un trait en pointillés rouges. Le comportement est principalement dirigé par la stratégie Lieu/Action en intérieur et par le suivi de route en extérieur.

priétés de généralisation, puisque la même architecture fonctionne pour au moins deux stratégies sensorimotrices de natures différentes. Cependant, nous avons montré que le fait de ne considérer que le niveau absolu de nouveauté n'était pas suffisant pour prendre une décision correcte sur l'efficacité d'un comportement. Par exemple, de petites perturbations n'influent en réalité pas significativement le comportement. La plupart du temps, le robot reste à l'intérieur de son bassin d'attraction et continue sa ronde correctement, même lorsque de petits événements inattendus surviennent. La raison est que la nouveauté en elle-même ne correspond pas à une récompense positive ou négative, mais plutôt à un signal émotionnel neutre. Nous avons montré qu'il était donc plus intéressant de considérer l'évolution du niveau de nouveauté plutôt que son niveau absolu. Surveiller les tendances de la nouveauté, en intégrant son activité dans le temps, fournit au système une bonne solution pour juger ses propres comportements. Si l'erreur de prédiction stagne à un niveau élevé ou bien si elle s'accroît, peu importe les actions que le robot tente, le comportement doit être considéré comme inefficace. Et si cette inefficacité perdure, cela signifie que le robot est coincé dans une situation de deadlock.

En suivant ces hypothèses, nous avons proposé un modèle de stress capable de moduler chaque stratégie en fonction des progrès du prédicteur et donc d'adapter le comportement à la situation. Nous avons présenté un ensemble d'expériences qui mettent en avant le besoin qu'a le robot d'utiliser un tel contrôleur afin de sélectionner la stratégie adéquate.

Par ailleurs, nous avons souligné le rôle des émotions dans la communication en ajoutant au contrôleur une simple procédure d'appel à l'aide déclenchée par la frustration du robot. Cette procédure permet au robot de communiquer son incapacité à réaliser sa tâche en appelant à l'aide

si aucune stratégie n'est évaluée comme efficace (si changer de stratégie n'induit aucun progrès de prédiction). Même si la procédure proposée reste ad-hoc en se basant sur un mécanisme d'appel à l'aide pré-câblé, elle est déclenchée par un signal porteur de sens. Ce signal de frustration pointe du doigt les situations pour lesquelles un affinement d'apprentissage est possible. Pourtant, ce contrôleur émotionnel n'est pas suffisant à lui seul pour atteindre une autonomie complète. Contrairement à la plupart des systèmes de motivations intrinsèques ([Schmidhuber, 1991; Kaplan and Oudeyer, 2007; Baranes, 2011; Barto et al., 2004; Santucci et al., 2012; Schembri et al., 2007]), notre système a toujours besoin d'un professeur pour apprendre par démonstration, même si cette démonstration n'est pas prescriptive. Dans nos travaux, la frustration est vue comme une motivation intrinsèque très utile à l'agent pour accroître graduellement son autonomie sur le long terme mais d'une manière supervisée. De futurs travaux se focaliseront sur la manière dont pourra être utilisé ce signal interne pour améliorer l'apprentissage d'une façon totalement autonome, c'est à dire sans besoin de la supervision d'un professeur.

Les travaux futurs consisteront à tester les performances du modèle sur de longues distances en environnements extérieurs (une expérience de navigation sur plusieurs kilomètres). Il faudra également se poser la question de la manière dont une simple boucle de rétroaction pourrait permettre au système de lever, de manière active, les ambiguïtés éventuelles perçues. A cause de l'ambiguïté de la perception, le robot a parfois besoin de variations dans ses sensations pour être capable de mesurer correctement la qualité d'une stratégie sensorimotrice. Il sera intéressant d'étudier comment utiliser correctement l'erreur de prédiction comme un signal de rétroaction modulant les actions. Par exemple, une erreur importante devrait déclencher un bruit important sur les actions du robot, impliquant des changements dans ses sensations. Ces variations ne feront décroître l'erreur que si les contingences sensori-motrices deviennent prévisibles. Dans le cas contraire, des variations incohérentes la feront augmenter. L'altération du comportement serait alors proportionnelle à l'erreur de prédiction. Nous prédisons qu'un tel mécanisme homéostatique permettra au robot d'auto-réguler plus finement son comportement. En mettant à jour ses connaissances par des modifications actives de son comportement, il devrait ainsi pouvoir vérifier l'exactitude de ses attentes.

Publications personnelles

- Jauffret, A., Cuperlier, N., Gaussier, P., and Tarroux, P. (2013c). From self-assessment to frustration, a small step towards autonomy in robotic navigation. *Frontiers in Neuro-robotics*, 7(16)
- Jauffret, A., Cuperlier, N., Belkaid, M., Gaussier, P., and Tarroux, P. (2013b). Frustration as a way toward autonomy and self-improvement in robotic navigation. In *The Third Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics. ICDL-EpiRob*, pages 1–7, Japon
- Jauffret, A., Belkaid, M., Cuperlier, N., Tarroux, P., and Gaussier, P. (2013a). Frustration: a generic mechanism to improve autonomy in robotics. In *third symposium on biology of decision making*, France

Conclusion

Cette thèse visait à concevoir une architecture de contrôle bio-inspirée permettant à un robot mobile de naviguer de manière autonome sur de grandes distances dans des environnements dynamiques, inconnus et complexes.

Dans cette optique, nous avons fait le choix d'aborder le problème sous l'angle des processus développementaux, c'est à dire l'apprentissage, par le robot, de tâches de navigation en interaction avec un professeur naïf. Afin d'accroître l'autonomie comportementale du robot, nous avons mis en évidence la nécessité d'une capacité d'auto-évaluation des perceptions pour pouvoir prendre des décisions. Nous avons également mis en évidence le rôle des émotions dans la régulation des comportements et son rôle social pour l'apprentissage interactif.

L'objectif de cette thèse était de montrer que l'auto-évaluation d'une perception peut être abordée comme la reconnaissance d'une dynamique sensorimotrice particulière et non pas celle d'une simple forme sensorielle. J'ai proposé un modèle générique capable d'apprendre à reconnaître la dynamique d'un comportement par l'extraction d'invariants statistiques du couplage existant entre sensations et actions. Ce modèle s'est construit progressivement par l'intégration de briques élémentaires au cours des différents chapitres.

Résumons

En introduction de ce manuscrit, nous avons constaté les problèmes auxquels fait face l'approche symbolique de l'intelligence artificielle. La cognition basée sur une représentation symbolique prédéfinie du monde se limite à la résolution de problèmes dont l'espace des états est borné et ne peut s'appliquer directement à un environnement réel, continu, complexe et dynamique. L'approche éactive propose une alternative élégante à ce problème en prônant l'aspect primordial du corps physique "situé" dans son environnement. Cette incarnation est nécessaire pour que le système puisse comprendre le monde dans lequel il évolue sur la base des interactions qu'il entretient avec ce dernier. Dans ce cadre, l'approche bio-inspirée de la robotique tente de tirer parti de solutions efficaces utilisées par les animaux pour s'affranchir de l'utilisation de symboles ou d'un modèle interne détaillé de l'environnement.

Dans le chapitre 2, nous avons évoqué rapidement l'état de l'art des modèles permettant la navigation autonome d'un robot mobile. Après un rappel des méthodes historiques utilisées pour la navigation maritimes, nous nous sommes concentré sur les méthodes de navigation actuelles dites "classiques" en abordant la problématique du SLAM (Simultaneous Localisation And Mapping). Ces méthodes se basent sur la construction d'une carte détaillée de l'environnement pour permettre au système de naviguer de manière précise dans celui-ci. Le coût mémoire important

ainsi que l'aspect statique de la carte montrent les limites de cette approche face à de grands environnements et dont la nature est dynamique.

Nous avons pu constater que les animaux utilisaient des solutions simples et robustes pour se déplacer en abordant le problème sous un autre angle. Chez les insectes, la réactivité et l'extraction d'invariants utiles pour la tâche suffisent à faire émerger un comportement de navigation robuste et adapté au monde réel sans besoin d'un système nerveux complexe. Chez les mammifères, l'apprentissage de cellules de lieu permet la construction d'une représentation mentale topologique de l'environnement tout en restant minimaliste et bien adaptée aux tâches que l'animal doit réaliser. Partant de ce constat, le champ de la robotique bio-inspirée s'est développé, conjointement aux approches classiques, en proposant des solutions élégantes et robustes au problème de la navigation autonome. En suivant cette approche, j'ai présenté les modèles de couplage sensorimoteur (PerAc) et de cellules de lieu visuelles sur lesquels s'appuient mes travaux. Ces modèles permettent l'émergence de comportements complexes via l'apprentissage d'un nombre restreint d'associations sensorimotrices.

Les bonnes capacités de généralisation du modèle ainsi que l'utilisation d'un mécanisme de compétition entre lieux permettent une reconnaissance fiable lorsque les amers appris sont globalement accessibles depuis n'importe quelle position de l'espace. Cependant, nous avons pu constater au chapitre 3 que le passage à l'échelle de ce modèle à des environnements plus grands et plus complexes pouvait être problématique. Si ce modèle permet l'apprentissage de tâches simples et robustes comme un comportement de ronde ou de retour au nid, ces tâches sont limitées à de petits environnements et de courtes périodes. Nous avons identifié trois causes majeures expliquant ces limites : la nature ambiguë des informations visuelles qui amène à des interprétations multiples des vues locales extraites, le traitement identique des amers avec l'incapacité du modèle à procéder à un tri et enfin l'aspect dynamique de l'environnement qui induit une dérive dans la reconnaissance des lieux.

Pour outrepasser ces limitations, nous avons proposé d'adopter un point de vue probabiliste en analysant le modèle initial sous le prisme de l'approche bayésienne. Cette comparaison nous a permis de constater que les notions manipulées par l'approche neuronale et celles couramment utilisées par les théories bayésiennes ne sont pas si éloignées. La principale différence provient de l'usage d'un terme de normalisation présent dans le calcul d'une inférence bayésienne. L'usage d'un mécanisme de compétition dans le modèle neuronal permet de s'affranchir de cette normalisation.

Nous avons ensuite montré qu'un simple apprentissage par conditionnement pouvait être utilisé à différents niveaux pour réaliser des estimations de densité de probabilité et ainsi agir comme un filtre attentionnel ne sélectionnant que l'information pertinente. Ce modèle permet de filtrer les amers redondants ou ambigus et de sélectionner ceux qui se trouvent être pertinents pour la tâche. Le système est alors capable d'apprendre à inférer la reconnaissance d'un amer visuel à partir de la perception de son azimut et d'autre part, d'apprendre des lois de probabilité (distributions) qui étaient autrefois définies de manière a priori dans le réseau. De la même manière, nous avons montré comment le robot peut construire des cellules de lieu "probabilistes" en apprenant à inférer la reconnaissance d'un lieu à partir de la perception d'une constellation d'amers-azimuts donnée. Ce modèle de prédiction uni-modale constitue la première brique d'auto-évaluation du système, l'évaluation de la pertinence des informations sensorielles (cohérence intra-modale).

Si cette analyse nous a permis d'identifier la manière dont peuvent être améliorées les informations sensorielles perçues par le système, elle n'en reste pas moins une première étape

vers l'autonomie puisqu'un modèle uniquement basé sur des informations unimodales montre ses limites lorsque la modalité en question est mise en défaut. Améliorer la robustesse de la reconnaissance passe aussi par l'aspect multimodal des informations traitées. L'utilisation d'informations visuelles implique généralement des processus coûteux et n'est parfois pas suffisant pour assurer une reconnaissance de lieux robuste dans des environnements où la modalité visuelle ne s'y prête pas. A titre d'exemple, un couloir, une forêt ou simplement un environnement peu éclairé impliquent de recourir à des modalités de natures différentes, comme l'usage d'informations idiothétiques. Dans le chapitre 4, nous avons donc proposé une stratégie de navigation supplémentaire basée sur une représentation proprioceptive des lieux (construite à partir d'informations internes au système). En partant d'un mécanisme simple d'intégration de chemin (chapitre 2), nous avons introduit un modèle de cellules de grille, inspiré de résultats neurobiologiques obtenus chez le rat. L'activité des cellules de grille obtenue est le résultat d'une simple projection (compression avec pertes des informations idiothétiques) puis d'une fusion de l'information d'intégration de chemin sur des neurones du cortex entorhinal. Cette projection agit comme une opération de modulo appliquée à l'intégration de chemin. Le modèle contraste avec la plupart des autres modèles de cellules de grille puisqu'il ne fait pas l'hypothèse d'une intégration de chemin réalisée par les cellules elles-mêmes mais, à l'inverse, fait l'hypothèse que l'intégration de chemin sur de grandes distances pourrait être stockée dans d'autres aires cérébrales.

Nous avons démontré la nécessité d'un mécanisme de calibration, contrôlé par un besoin périodique de retour au nid, pour maintenir une activité de grille précise et cohérente dans le temps. Nous avons également montré comment un ensemble de cellules de grille pouvait être fusionnées pour générer des cellules de lieu idiothétiques par un simple apprentissage hebbien. Ce modèle ne nécessite pas de construire une représentation cartésienne de l'environnement et permet de couvrir un espace large en utilisant un petit nombre de résolutions de grille premières entre elles (seulement 3 modulo). La nature binaire de l'activité des cellules de grille ne permet cependant que d'obtenir des champs de lieu étroits sans aucune capacité de généralisation. Nous avons émis l'hypothèse que ces capacités de généralisation pourraient émerger chez l'animal grâce au nombre important de cellules de grille présentes dans les différentes couches de dMEC. La simulation d'un grand nombre de cellules permet en effet d'obtenir une taille de champ plus large.

L'ajout de cette seconde modalité pose le problème de l'intégration des informations allothétiques et idiothétiques au sein d'un même espace cohérent. Nous nous sommes donc intéressés, dans le chapitre 5, à la manière dont peuvent être fusionnées les représentations issues de ces deux modalités. Le modèle de fusion proposé repose sur un simple conditionnement entre les deux voies et permet d'obtenir des cellules de lieu multimodales robustes aux perturbations (ambiguïtés visuelles, kidnapping et fermeture de boucle). Par ailleurs, j'ai montré que la synergie émergente de cette fusion pouvait, dans la plupart des cas, améliorer la reconnaissance en donnant plus de poids aux votes unanimes. Ce modèle de fusion va plus loin qu'une simple compétition puisque la reconnaissance multimodale obtenue est globalement plus robuste que celle de la modalité la plus fiable.

Pour finir, j'ai ouvert la discussion sur la fusion de modalités en proposant un modèle générique pour l'intégration d'un plus grand nombre de modalités. Après apprentissage des corrélations intermodales, ce système bas-niveau devrait permettre d'augmenter la robustesse au bruit, de lever certaines ambiguïtés et de compenser les insuffisances sensorielles afin de les rendre in-

visibles, autant que possible, pour les modules de plus haut niveau. Ce modèle de prédictions intermodales constitue la seconde brique d'auto-évaluation du système, l'évaluation de la pertinence des informations multimodales (cohérence inter-modale).

Cependant, quelle que soit la robustesse des informations que traite le système, j'ai souligné le fait qu'on ne peut éviter les cas de comportements pathologiques. Il apparaît inévitable qu'à certains moments le robot échoue dans la réalisation de sa tâche à cause d'un apprentissage trop superficiel ou lorsqu'il se retrouve en dehors du bassin d'attraction appris. Pour des tâches compliquées, être autonome signifie également être capable de se rendre compte d'un comportement défaillant afin de prendre des mesures pour y remédier. De plus, au même titre qu'une seule modalité sensorielle n'est pas suffisante pour obtenir une perception robuste, nous avons pu constater que le système pouvait avoir besoin de recourir à d'autres stratégies lorsque la situation l'impose.

Nous avons donc proposé de prendre en compte d'autres stratégies en étendant notre architecture. L'ajout d'une stratégie de suivi de route robuste et biologiquement plausible permet au robot de converger naturellement vers un point de fuite et donc de suivre une route ou un chemin dans des situations où il n'est ni nécessaire, ni efficace d'utiliser un système de navigation à base d'associations Lieu/Action.

Mais, que ce soit pour la détection d'un apprentissage incomplet ou pour contrôler les différentes stratégies dont peut disposer le robot, celui-ci nécessite un contrôleur au second-ordre capable d'évaluer ses perceptions pour prendre les bonnes décisions. Dans le chapitre 6, j'ai donc proposé un modèle capable d'apprendre à reconnaître la dynamique résultante du couplage entre sensations et actions. L'apprentissage de cette dynamique, par l'extraction de moments statistiques, forme un modèle interne des interactions du robot avec son environnement, un modèle de ses perceptions. En se basant sur les solutions proposées précédemment (chapitres 3 et 5), l'estimation des moments est réalisée à l'aide de simples conditionnements.

Le système est ainsi au fait d'un éventuel comportement anormal par le biais d'une surprise (erreur de prédiction) engendrée par une dynamique sensori-motrice inhabituelle. Sa généralité nous permet d'aller vers un modèle des émotions plus complet que celui proposé par [Hasson et al., 2011]. Ce modèle de détection de nouveauté constitue la troisième brique d'auto-évaluation du système, l'évaluation de la pertinence des comportements (cohérence sensori-motrice).

Nous avons montré que notre modèle est suffisamment générique pour pouvoir être adapté à deux comportements sensorimoteurs de natures différentes. Cela dit au chapitre 7, nous avons constaté que la surprise liée à la détection d'un comportement inhabituel n'est pas suffisante, à elle seule, pour renseigner le système sur l'efficacité de son comportement. Celui-ci doit également surveiller les tendances de cette erreur de prédiction, c'est à dire l'évolution de l'anomalie dans le temps afin de pouvoir juger de la criticité de celle-ci. J'ai alors montré comment un contrôleur émotionnel au second ordre pouvait permettre au système de réguler son comportement à partir du signal de surprise provenant du méta-prédicteur introduit au chapitre 6. Ce méta-contrôleur est capable d'évaluer en temps-réel la dynamique des deux stratégies testées afin de leur attribuer un niveau de confiance (stress) et pondérer en fonction leur participation dans le comportement final. De même que pour la fusion de modalités sensorielles (chapitre 5), nous avons constaté qu'une coopération entre stratégies permet de laisser émerger des propriétés d'enrichissement mutuelle, qui n'auraient pu exister avec l'usage d'une compétition stricte.

Par ailleurs, j'ai souligné le rôle social des émotions pour l'apprentissage à long terme. L'ajout d'un simple mécanisme permettait au robot de communiquer son niveau de frustration lorsqu'il se rend compte de son incapacité à réaliser la tâche. A l'instar de certains processus développementaux, ce mécanisme autorise le système à initier une interaction pour pointer du doigt les situations qu'il ne maîtrise pas et pour lesquelles un raffinement de l'apprentissage est nécessaire. La frustration est alors un mécanisme émotionnel très utile à l'agent pour accroître graduellement son autonomie sous la supervision partielle d'un autre agent.

Discussion et perspectives

Dans cette thèse, nous avons proposé des solutions au problème de l'autonomie en robotique mobile. Pour le robot, atteindre progressivement une autonomie suffisante passe nécessairement par l'ajout de capacités d'auto-évaluation. Cette auto-évaluation doit être présente à tous les niveaux dans l'architecture et se décline selon trois facettes :

1. l'évaluation intra-modale par l'extraction d'invariants présents au sein d'une sensation donnée.
2. l'évaluation inter-modale par l'apprentissage des corrélations existantes entre les différentes sensations.
3. l'évaluation sensori-motrice permettant au système d'apprendre à reconnaître la dynamique de ses comportements.

Les modèles que j'ai proposé, en réponse à ces trois aspects, sont relativement génériques tout en se basant sur de simples conditionnements.

Concernant le système visuel, plusieurs améliorations sont envisageables :

1. Si nous avons proposé des solutions pour permettre la pondération des amers dans la reconnaissance des lieux, nous n'avons pas traité du problème du recrutement de ces amers. Dans le modèle actuel, la pauvreté des mécanismes de vision induit une faible capacité de discrimination puisque la dynamique de reconnaissance des amers est entièrement comprise dans une plage d'activité comprise entre 0.8 et 1. Par voie de conséquence, le système recrute systématiquement chaque amer perçu, qu'il ait été vu auparavant ou non. Ce problème peut être repoussé en améliorant par exemple la résolution des imagerie extraites, ce qui permettrait une meilleure discrimination visuelle, ou bien en considérant une manière plus intelligente et économique d'encoder les lieux. Selon une approche plus développementale, l'apprentissage pourrait être réalisé de manière progressive et hiérarchique, tout comme le bébé dans son berceau apprend en premier lieu à reconnaître un ensemble d'objets, chaque objet étant un assemblage particulier de primitives visuelles et d'affordances simples. Ce n'est qu'en grandissant, que le fait de se déplacer implique le besoin d'apprendre à reconnaître des lieux, chaque lieu pouvant être vu comme un assemblage particulier d'objets visuels. De la même façon, une base d'amers pourrait être apprise par le robot dans un premier temps avant d'envisager le recrutement de cellules de lieu, puis de cellules de contexte (un contexte regroupant un ensemble de lieux aux caractéristiques communes), etc...
2. Dans le chapitre 4, nous avons montré que l'activité périodique des cellules de grille décelées dans le cortex entorhinal du rat pouvait s'expliquer par la simple projection d'un

ensemble de neurones sur un autre de taille plus réduite. Nous avons émis l'hypothèse que le but d'une telle projection serait avant tout de réaliser une compression avec perte des informations d'intégration de chemin. La compression avec perte permettrait ainsi de générer un espace de travail restreint sur lequel l'animal peut effectuer des opérations mentales de manière plus aisée. L'un des attraits certains de cette compression est que la taille de l'espace de travail ainsi généré augmente de façon linéaire avec la surface métrique à coder (contre une augmentation exponentielle pour des données non-compressées). Ainsi, un petit nombre de cellules, dont les modulo sont premiers entre eux, permettent d'encoder un grand espace. Le modèle visuel (PrPh), quant à lui, ne réalise pas une telle compression et montre vite ses limites pour de grands environnements ou l'explosion du nombre d'amers recrutés devient problématique. Pourtant, une projection similaire avec faible probabilité d'ambiguïté pourrait être réalisée au niveau de la fusion entre amers et azimuts.

Au niveau de la fusion de modalité, il est également possible d'aller plus loin que nous l'avons fait : dans le chapitre 5, nous n'avons traité que du problème de l'intégration bi-modale, faisant intervenir vision et proprioception. En partant du modèle générique de fusion multimodale proposé, on peut imaginer le fait de rendre le codage des lieux plus robuste encore par l'ajout d'autres modalités comme des informations auditives, olfactives, un système d'odométrie visuelle ou encore l'activité de capteurs de proximité. Dans le chapitre 4, nous avons montré que l'apprentissage d'un profil particulier au niveau des capteurs ultrasons pouvait permettre de reconnaître un coin (cellules de bords) et ainsi renforcer la robustesse de la reconnaissance en ce lieu singulier. De la même façon, l'apprentissage de profils d'activité spécifiques pourrait, par exemple, permettre de discriminer un environnement ouvert d'un couloir étroit et ainsi réduire les possibilités d'ambiguïté. De plus, la reconnaissance du passage d'une porte pourrait permettre de définir un lieu saillant précis, particulièrement adapté pour autoriser la recalibration du champ d'intégration de chemin.

Concernant l'auto-évaluation, le modèle proposé pourrait également être amélioré. Pour une analyse complète du problème, la capacité d'auto-évaluation se scinde en réalité en deux étapes :

1. La détection des conséquences observables d'un phénomène anormal.
2. Le fait d'inférer la cause du phénomène à partir de l'observation de ses conséquences.

Ce constat pose la question de la manière dont un système peut remonter à la source d'un problème en observant ses conséquences. En effet, du point de vue du robot, il n'est pas trivial de comprendre si le problème en question vient de lui (capteur en panne, mauvaise reconnaissance, comportement inefficace ou apprentissage incomplet) ou bien de l'extérieur (éblouissement visuel, mouvement d'objets, changements physiques dans l'environnement). A titre d'exemple, le système actuel n'est pas capable de faire de distinction entre une anomalie liée à une sortie de son domaine d'apprentissage ou bien liée à un obstacle présent en travers de la route apprise (e.g. une porte fermée). Une partie de la solution pourrait consister à rendre le système pro-actif en lui permettant de déclencher un comportement générant des variations sensorielles à la suite d'une surprise. Ces variations amélioreraient sa capacité à mesurer la dynamique sensorimotrice du comportement évalué. On peut illustrer ce principe par l'exemple suivant : afin de savoir si un objet est plat ou bien en relief, il est souvent nécessaire de tourner autour, ou de tourner l'objet avec sa main et ainsi produire des variations dans la forme perçue. Ces variations spécifiques,

dont nous sommes capables de reconnaître la dynamique, nous permettent alors de statuer parmi plusieurs hypothèses, quant à la géométrie de l'objet.

Dans le chapitre 7, nous avons ouvert la discussion sur ce point en proposant d'ajouter une simple boucle de rétroaction pour permettre au système de lever les ambiguïtés plus rapidement. L'erreur de prédiction peut être utilisée comme signal de rétroaction modulant ses actions. Une surprise conséquente déclencherait un bruit important sur les actions du robot afin d'induire des changements dans ses sensations. Ces variations ne feront décroître l'erreur que si les invariants sensorimoteurs deviennent prévisibles. Dans le cas contraire, des variations incohérentes la feront augmenter. Certains résultats préliminaires à propos de la stratégie de suivi de route (non publiés ici) me poussent à croire que ce mécanisme homéostatique permettrait au robot de s'auto-réguler plus finement, de façon active. En effet, lorsque le point de fuite est aligné avec le robot, ce dernier n'effectue plus de rotation. Le méta-contrôleur n'est alors plus capable de reconnaître la dynamique de convergence si celui-ci n'induit aucune variation comportementale. Le méta-contrôleur a besoin de percevoir le comportement convergent du système vers le point de fuite pour pouvoir correctement reconnaître cette dynamique de convergence et ainsi juger de la qualité du suivi.

D'autre part, lors d'une frustration, plutôt que de faire appel à l'humain, le système pourrait déclencher un comportement de retour à un point connu par intégration de chemin ou bien plus simplement une exploration aléatoire de la zone (mouvement brownien). Le robot pourrait se voir attribuer une collection de stratégies, innées ou apprises, où l'appel à l'aide serait vu comme la stratégie utilisée en dernier recours.

Pour aller plus loin, il sera nécessaire à terme que le système puisse réaliser des inférences plus complexes : Le modèle d'auto-évaluation que j'ai présenté permet surtout d'induire, mais pas de déduire ni d'inférer la réelle cause d'un phénomène. Pourtant, le jeune enfant est non seulement capable de détecter une anomalie, mais également d'en inférer la cause. Par exemple, lorsqu'un jouet ne fonctionne plus, l'enfant détecte rapidement l'anomalie puis cherche à comprendre ce qui a causé cette anomalie. Il est alors face à 2 choix : l'objet est-il en panne ? ou est-ce moi qui m'y prend mal ?

L'enfant obtient la réponse en formulant des inférences de haut niveau à partir d'informations extraites de l'environnement présent et de connaissances a priori (extraites d'expériences passées) [Gweon and Schulz, 2011] :

- *"si j'ai vu un autre agent utiliser l'objet sans problème, alors c'est moi qui ne sait pas m'en servir"*. L'enfant demande alors de l'aide à un autre agent.
- *"si j'ai l'habitude d'utiliser cet objet moi même, alors c'est certainement l'objet qui est cassé"*. Dans ce cas, l'enfant reporte son attention sur un autre objet.

Enfin, dans ces travaux, la régulation comportementale est réalisée de telle sorte que le système choisisse simplement l'affordance sensorimotrice la plus proche/attractive. Cet homéostat lui permet de s'assurer qu'il effectue correctement la ronde apprise. Ceci étant dit, pour donner du sens à l'auto-évaluation, il manque une récompense (valence positive ou négative) explicite associée aux actions entreprises par le système :

- Cette récompense peut être externe : il pourrait s'agir d'un bouton physique présent sur le robot, dont l'appui simulerait une approbation ou une réfutation de la part de l'humain.
- La récompense peut également être issu d'un mécanisme interne : l'attrait pour la nouveauté en est un bon exemple.

En effet, dans la nature, un animal a tout intérêt à rester suffisamment curieux pour continuer

à explorer son environnement de temps à autre (mais pas trop), et ce même si ce dernier lui est familier. La curiosité est un moteur pour accélérer l'apprentissage et élargir le champ de connaissance de l'agent. Une attirance pour des situations nouvelles pousserait ainsi celui-ci à prendre certains risques calculés. Il pourrait, par exemple, décider de désobéir volontairement pour se mettre dans des situations non-maîtrisées, pour lesquelles un progrès en apprentissage est possible.

De manière générale, le système a besoin de récompenses à différents niveaux et dont l'organisation pourrait être hiérarchique. A bas niveau, il s'agirait de "drives", c'est à dire de satisfactions physiologiques associées à des besoins primaires. A plus haut niveau, il pourrait s'agir de "motivations", des états émotionnels plus complexes capables d'inhiber temporairement certains besoins primaires en vu de l'obtention d'une plus grande récompense à moyen ou long terme. Ce système de récompense sera sûrement indispensable dans l'optique d'une navigation extérieure sur plusieurs kilomètres. Ainsi, pour atteindre un but lointain, nécessitant plusieurs stratégies et beaucoup de contextes différents, il pourrait être nécessaire de scinder le parcours en introduisant des sous-buts (checkpoints) pour permettre au robot de valider sa ronde étape par étape. Le fait d'associer des états émotionnels à des séquences sensorimotrices, portions du comportement global, devrait autoriser une sélection de l'action qui ne soit plus purement réactive et une autonomie décisionnelle accrue. Nous pourrions alors observer l'émergence d'une gamme de comportements plus riches et d'une personnalité propre à chaque agent puisque dépendante de ses expériences uniques.

Bibliographie

- Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology : CB*, 14(3) :257–262. 107
- Amaral, D. G. and Witter, M. P. (1989). The Three-Dimensional Organization of the Hippocampal Formation : a Review of Anatomical Data. *Neuroscience*, 31(3) :571–591. 88
- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics*, 27(2) :77–87. 34, 138
- Anderson, C. W., Sutton, R. S., and Barto, A. G. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(5) :834–846. 29
- Andreasson, H., Treptow, A., and Duckett, T. (2005). Localization for Mobile Robots using Panoramic Vision, Local Features and Particle Filter. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. 45
- Andry, P. (2004). Learning Invariant Sensorimotor Behaviors : A Developmental Approach to Imitation Mechanisms. *Adaptive Behavior*, 12(2) :117–140. 120, 138
- Andry, P., Gaussier, P., Moga, S., Banquet, J. P., and Nadel, J. (2001). The Dynamics of Imitation Processes : From Temporal Sequence Learning to Implicit Reward Communication. *IEEE Trans. On Man, Systems and Cybernetics Part A : Systems and Humans*, 31(5) :431–442. 19
- Apek, K. and Wyllie, D. (2010). *R.U.R (Rossum's Universal Robots)*. Echo Library. 13
- Arkin, R. (1998). *Behavior-based Robotics*. Bradford book. MIT Press. 15
- Arleo, A. and Gerstner, W. (2000). Spatial cognition and neuro-mimetic navigation : a model of hippocampal place cell activity. *Biological cybernetics*, 83(3) :287–299. 55, 93, 109
- Arleo, A. and Rondi-Reig, L. (2007). Multimodal sensory integration and concurrent navigation strategies for spatial cognition in real and artificial organisms. *Journal of integrative neuroscience*, 6(3) :327–366. 55
- Arleo, A., Smeraldi, F., Hug, S., and Gerstner, W. (2001). Place cells and spatial navigation based on vision, path integration, and reinforcement learning. In *Path Integration, and Reinforcement Learning*. *Advances in Neural Information Processing Systems*, pages 89–95. 109

- Asimov, I. (2004). *I, Robot*. The Robot Series. Random House Publishing Group. 13
- Auvray, M., Lenay, C., O'regan, K. J., and Lefèvre, J. (2005). Suppléance perceptive, immersion et informations proprioceptives. *Arob@se*, 1 :94–113. 124
- Bach-y Rita, P., Collins, C. C., Saunders, F. A., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, 221(5184) :963–964. 107
- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., and Frith, C. D. (2010). Optimally interacting minds. *Science*, 329(5995) :1081–1085. 116
- Balkenius, C., Morén, J., and Winberg, S. (2009). Interactions between Motivation, Emotion and Attention : From Biology to Robotics. In Cañamero, L., Oudeyer, P. Y., and Balkenius, C., editors, *Proceedings of the 9th International Conference on Epigenetic Robotics*, pages 25–32. 19, 137
- Ballard, D., Rao, R., and Zhang, Z. (1999). A Model of Predictive Coding Based on Spike Timing. *National Resource Lab*, pages 1–23. 81
- Banquet, J. P., Gaussier, P., Quoy, M., Revel, A., and Burnod, Y. (2005). A hierarchy of associations in hippocampo-cortical systems : cognitive maps and navigation strategies. *Neural computation*, 17(6) :1339–1384. 56
- Baranes, A. (2011). Motivations Intrinsèques et Contraintes Maturationnelles pour l'Apprentissage Sensorimoteur. *PhD Thesis. University of Bordeaux 1*. 148
- Barto, A. G., Singh, S., and Chentanez, N. (2004). Intrinsically Motivated Learning of Hierarchical Collections of Skills. In *3rd International Conference on Development Learning (ICDL)*, pages 112–119. 120, 148
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, 53 :370–418. 67
- Bernard, C. (1865). *Introduction à l'étude de la médecine expérimentale*. J. B. Baillière et Fils. 135
- Berthoz, A. (1997). *Sens du mouvement (Le)*. Sciences Humaines. Editions Odile Jacob. 124
- Berthoz, A. and Petit, J. (2006). *Phénoménologie et physiologie de l'action*. Sciences Humaines. Editions Odile Jacob. 117
- Bertozzi, M. and Broggi, A. (1998). Gold : a parallel real-time stereo vision system for generic obstacle and lane detection. *Image Processing, IEEE Transactions on*, 7(1) :62–81. 181
- Beurle, R. L. (1956). Properties of a Mass of Cells Capable of Regenerating Pulses. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 240(669) :55–94. 34
- Bierre, P. (1985). The professor's challenge. *AI Magazine*, 5(4) :60–70. 14
- Bird, C. M. and Burgess, N. (2008). The hippocampus and memory : insights from spatial processing. *Nature reviews. Neuroscience*, 9(3) :182–194. 51

- Bishop, C. (1994). Novelty detection and neural network validation. *IEE Proceedings - Vision, Image, and Signal Processing*, 141(4) :217. 122
- Braitenberg, V. (1984). *Vehicles : Experiments in Synthetic Psychology*. Bradford Books, Cambridge. 15, 54, 183
- Broca, P. (1871). *Mémoires d'anthropologie*. Number 1 in Mémoires d'anthropologie. Reinwald. 50
- Brooks, R. A. (1986). A Robust Layered Control System For a Mobile Robot. *IEEE Journal of Robotics and Automation*, 2(1) :14–23. 15
- Brooks, R. A. (1990). Elephants Don't Play Chess. *Robotics and Autonomous System*, 6 :3–15. 124
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1-3) :139–159. 54
- Brun, V. H., Otnass, M. K., Molden, S., Steffenach, H.-A., Witter, M. P., Moser, M.-B., and Moser, E. I. (2002). Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science (New York, N.Y.)*, 296(5576) :2243–2246. 88
- Burgess, N., O'Keefe, J., and Recce, M. (1994). A model of hippocampal function. *Neural Networks*, 7(6-7) :1065–1081. 55
- Burgess N. Barry C., O. J. (2007). An oscillatory interference model of grid cell firing. *Hippocampus*, 17. 90
- Burwell, R. D. and Hafeman, D. M. (2003). Positional firing properties of postrhinal cortex neurons. *Neuroscience*, 119(2) :577–588. 57
- Canamero, L. and Gaussier, P. (2004). Emotion understanding : robots as tools and models. *in Emotional Development*. 19, 136, 137
- Carpenter, G. and Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, 21(3) :77–88. 30, 122
- Cartwright, B. A. and Collett, T. S. (1983). Landmark learning in bees. *Journal of Comparative Physiology A : Sensory, Neural, and Behavioral Physiology*, 151(4) :521–543. 54
- Changeux, J.-P., Wacongne, C., and Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, 32(11) :3665–3678. 81
- Chatila, R. and Laumond, J. (1985). Position referencing and consistent world modeling for mobile robots. *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, 2. 44
- Cheung, A., Ball, D., Milford, M., Wyeth, G., and Wiles, J. (2012). Maintaining a Cognitive Map in Darkness : The Need to Fuse Boundary Knowledge with Path Integration. *PLoS Comput Biol*, 8(8) :e1002651. 102

- Chomsky, N. (1961). Formal discussion : the development of grammar in child language. In Brown, R., editor, *The acquisition of language*. Purdue University. 66
- Clarke, A. and Kubrick, S. (1999). *2001 : A Space Odyssey*. Turtleback Books. 13
- Crevier, D. (1993). *AI : the tumultuous history of the search for artificial intelligence*. Basic Books. 13
- Cuperlier, N., Gaussier, P., and Quoy, M. (2008). Interest of Spatial Context for a Place Cell Based Navigation Model. In Asada, M., Hallam, J., Meyer, J.-A., and Tani, J., editors, *From Animals to Animats 10*, volume 5040 of *Lecture Notes in Computer Science*, pages 169–178. Springer Berlin Heidelberg. 17, 65
- Cuperlier, N., Quoy, M., and Gaussier, P. (2006). Navigation and Planning in an Unknown Environment Using Vision and a Cognitive Map. In Christensen, H. I., editor, *EUROS*, volume 22 of *Springer Tracts in Advanced Robotics*, pages 129–142. Springer. 38
- Cuperlier, N., Quoy, M., and Gaussier, P. (2007). Neurobiologically inspired mobile robot navigation and planning. *Frontiers in neurorobotics*, 1 :3. 120
- D. Lack, R. M. L. (1938). Stockholm Bird Observatory homing experiments. *Puffins, storm petrels and Manx shearwaters*. *Brit Birds*, 31 :242–248. 47
- Damasio, A. (1999). *The feeling of what happens : body and emotion in the making of consciousness*. A Harvest book. Harcourt Inc. 106
- Damasio, A. R. (1995). *Descartes' Error : Emotion, Reason, and the Human Brain*. Harper Perennial, 1 edition. 19, 136
- Damasio, A. R. (2003). *Looking for Spinoza : Joy, Sorrow, and the Feeling Brain*, volume 1st. Harcourt. 137
- Darwin, C. (1873). Origin of Certain Instincts. *Nature* 7, 179 :417–418. 48
- Day, B. L. and Fitzpatrick, R. C. (2004). The vestibular system. *Journal of Applied Physiology*, 15(15) :R583–R586. 108
- Deci, E. (1975). *Intrinsic Motivation*. Bibliographie pp.295. Plenum Publishing Company Limited. 119
- Deci, E. and Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Perspectives in Social Psychology. Springer. 119
- Deci, E. and Ryan, R. (2002). *Handbook of Self-determination Research*. University of Rochester Press. 119
- Delorme, A. and Flückiger, M. (2003). *Perception et réalité : Introduction à la psychologie des perceptions*. Neurosciences & cognition. De Boeck Supérieur. 107
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1) :1–38. 44

- Deneve S, P. A. (2004). Bayesian multisensory integration and cross-modal spatial links. *J. Physiol. Paris* 98 (1-3), pages 249–58. 16, 63, 90
- Devroye, L. and Wise., G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. In *SIAM J. Appl. Math.* 38, pages 80–488. 122
- Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, 3(4) :357–370. 124
- Dingle Hugh and Drake, V. A. (2007). What Is Migration ? *BioScience*, 57(2) :113–121. 47
- Dollé, L., Khamassi, M., Girard, B., Guillot, A., and Chavarriaga, R. (2008). Analyzing interactions between navigation strategies using a computational model of action selection. *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, pages 71–86. 55, 123
- Dolorfo, C. and D.G., A. (1998). Entorhinal cortex of the rat : topographic organization of the cells of origin of the perforant path projection to the dentate gyrus. *Comp. Neurol.*, 398 :25–48. 88
- Doya, K. (2007). Reinforcement learning : Computational theory and biological mechanisms. *HFSP journal*. 67
- Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification*. Pattern Classification and Scene Analysis : Pattern Classification. Wiley. 69
- Ebbinghaus, H., Ruger, H., and Bussenius, C. (1913). *Memory : a contribution to experimental psychology*. Educational reprints. Teachers College, Columbia University. 81
- Ehrsson, H. H., Holmes, N. P., and Passingham, R. E. (2005). Touching a rubber hand : feeling of body ownership is associated with activity in multisensory brain areas. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(45) :10564–10573. 108
- Ekman, P. (1984). Expression and the nature of emotion. In *K. Scherer and P. Ekman (Eds.), Approaches to emotion*, 24 :319–343. 19
- Etienne, A. S. (1998). Mammalian Navigation, Neural Models and Biorobotics. *Connection Science*, 10(3-4) :271–289. 91
- Etienne, A. S. and Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus*, 14(2) :180–192. 48, 50, 101
- Feldman, J. A. and Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, 6 :205–254. 30
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press. 108
- Fiete, I. R., Burak, Y., and Brookings, T. (2008). What Grid Cells Convey about Rat Location. *The Journal of Neuroscience*, 28(27) :6858–6871. 90

- Filliat, D. and Meyer, J.-A. (2002). Global localization and topological map-learning for robot navigation. In *Proceedings of the seventh international conference on simulation of adaptive behavior on From animals to animats*, ICSAB, pages 131–140, Cambridge, MA, USA. MIT Press. 68
- Flor, H., Nikolajsen, L., and Staehelin Jensen, T. (2006). Phantom limb pain : a case of maladaptive CNS plasticity ? *Nature reviews. Neuroscience*, 7(11) :873–881. 136
- Floreano, D., Beyeler, A., and Zufferey, J.-C. (2009). Vision-based control of near-obstacle flight. *Autonomous Robots*, 27(3) :201–219. 54
- Foka, A. F. and Trahanias, P. E. (2010). Probabilistic Autonomous Robot Navigation in Dynamic Environments with Human Motion Prediction. *International Journal of Social Robotics*, 2(1) :79–94. 46
- Foster, D. J., Morris, R. G., and Dayan, P. (2000). A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus*, 10(1) :1–16. 55
- Foster, T. C., Castro, C. A., and McNaughton, B. L. (1989). Spatial selectivity of rat hippocampal neurons : dependence on preparedness for movement. *Science (New York, N.Y.)*, 244(4912) :1580–1582. 52
- Freeman, W. and Adelson, E. (1991). The design and use of steerable filters. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(9) :891–906. 181
- Freides, D. (1974). Human Information Processing and Sensory Modality :. *Psychological Bulletin*, 81(5) :284–310. 107
- Fritzke, B. (1995). A growing neural gas network learns topologies. In *Advances in Neural Information Processing Systems 7*, pages 625–632. MIT Press. 122
- Fuhs, M. C. and Touretzky, D. S. (2006). A Spin Glass Model of Path Integration in Rat Medial Entorhinal Cortex. *The Journal of Neuroscience*, 26(16) :4266–4276. 90, 93
- Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., and Moser, M.-B. (2004). Spatial representation in the entorhinal cortex. *Science*, 305 :1258–1264. 88
- Gallistel, C. R. and Gelman, R. (1990). The what and how of counting. *Cognition*, 34(2) :197–199. 42
- Gaussier, P., Baccon, J.-C., Prepin, K., Nadel, J., and Hafemeister, L. (2004). Formalization of recognition, affordances and learning in isolated or interacting animats. *Adaptive Behavior SAB 04*. 124
- Gaussier, P., Banquet, J. P., Sargolini, F., Giovannangeli, C., Save, E., and Poucet, B. (2007). A model of grid cells involving extra hippocampal path integration, and the hippocampal loop. *Journal of Integrative Neuroscience*, 6(3) :447–476. 56, 61, 62, 90, 91, 92, 93, 114
- Gaussier, P., Cuperlier, N., Laroque, P., Gaussier, N., and Quoy, M. (2010). Cognitive map plasticity and imitation strategies to improve individual and social behaviors of autonomous agents. *Paladyn*, 1(1) :25–36. 120

- Gaussier, P., Joulain, C., Banquet, J.-P., Lepretre, S., and Revel, A. (2000). The visual homing problem : an example of robotics/biology cross fertilization. *Robotics and Autonomous Systems*, 30(1) :155–180. 60, 61
- Gaussier, P., Joulain, C., Revel, A., Zrehen, S., and Banquet, J. (1997). Autonomous robot learning : what can we take for free ? *ISIE '97 Proceeding of the IEEE International Symposium on Industrial Electronics*, 1. 61
- Gaussier, P., Moga, S., Quoy, M., and Banquet, J.-P. (1998). From perception-action loops to imitation processes : A bottom-up approach of learning by imitation. *Applied Artificial Intelligence*, 12(7-8) :701–727. 138
- Gaussier, P., Prepin, K., and Nadel, J. (2003). Toward a cognitive system algebra : Application to facial expression learning and imitation. In Pierre, S., Barbeau, M., and Kranakis, E., editors, *Embodied Artificial Intelligence*, volume 2865 of *Lecture Notes in Computer Science*, pages 243–258. Springer. 124
- Gaussier, P., Revel, A., Banquet, J. P., and Babeau, V. (2002). From view cells and place cells to cognitive map learning : processing stages of the hippocampal system. *Biological cybernetics*, 86(1) :15–28. 56, 120
- Gaussier, P. and Zrehen, S. (1995). PerAc : A neural architecture to control artificial animals. *Robotics and Autonomous Systems*, 16(2-4) :291–320. 47, 48, 55, 56, 60
- Gentaz, E., Hatwell, Y., and Streri, A. (2001). Constructivist and ecological approaches in tactual perception. *Behavioral and Brain Sciences*, 25(01). 106
- Gibson, J. (1978). The Ecological Approach to the Visual Perception of Pictures. *Leonardo*, 11(3) :227–235. 54
- Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, USA. 123, 124
- Giovannangeli, C. (2007). *Navigation autonome bio-inspirée en environnement intérieur et extérieur : Apprentissages sensori-moteurs et planification dans un cadre interactif*. PhD thesis, Université de Cergy-Pontoise. 59, 64, 65
- Giovannangeli, C. and Gaussier, P. (2007). Orientation System in Robots : Merging Allothetic and Idiothetic Estimations. pages 349–354. 57, 102
- Giovannangeli, C. and Gaussier, P. (2010). Interactive Teaching for Vision-Based Mobile Robots : A Sensory-Motor Approach. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 40(1). 16, 56, 60, 141
- Giovannangeli, C., Gaussier, P., and Banquet, J. P. (2006). Robustness of Visual Place Cells in Dynamic Indoor and Outdoor Environment. *International Journal of Advanced Robotic Systems*, 3(2) :115–124. 120, 132
- Girard, B., Cuzin, V., Guillot, A., Gurney, K. N., and Prescott, T. J. (2003). A Basal ganglia inspired model of action selection evaluated in a robotic survival task. *Journal of Integrative Neuroscience*, 2(3) :179–200. 138

- Girard, B., Dollé, L., Staffa, M., Caluwaerts, K., Khamassi, M., Grand, C., Favre-Félix, A., and N'Guyen, S. (2012). A biologically inspired meta-control navigation system for the Psikharpax rat robot. *Bioinspiration & Biomimetics*, 7(2) :025009. 123
- Glanzer, M. and Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 5(4) :351 – 360. 81
- Gleitman, L., Liberman, M., and Osherson, D. (1995). *Language*, volume 1. MIT Press. 66
- Godefroid, J. (2001). *Psychologie : science humaine et science cognitive*. Ouvertures psychologiques. De Boeck Université. 106
- Gopnik, A. (2001). PSYCHOLOGY : The Meaning of Make-Believe. *Science*, 292(5514) :57a–57. 67
- Gopnik, A. and Rosati, A. (2001). Duck or rabbit ? Reversing ambiguous figures and understanding ambiguous representations. *Developmental Science*, 4(2) :175–183. 67
- Gopnik, A., Sobel, D. M., Schulz, L. E., and Glymour, C. (2001). Causal learning mechanisms in very young children : two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental psychology*, 37(5) :620–629. 67
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2) :107–113. 45
- Gothard, K. M., Hoffman, K. L., Battaglia, F. P., and McNaughton, B. L. (2001). Dentate Gyrus and CA1 Ensemble Activity during Spatial Reference Frame Shifts in the Presence and Absence of Visual Input. *Journal of Neuroscience*, 21 :7284–7292. 52
- Gould, J. L. (1986). The biology of learning. *Annual review of psychology*, 37 :163–192. 53
- Gourichon, S., Meyer, J. A., and Pirim, P. (2002). Using Coloured Snapshots For Short-Range Guidance In Mobile Robots. *International Journal of Robotics and Automation*, 17 :12–16. 54
- Grah, G., Wehner, R., and Ronacher, B. (2005). Path integration in a three-dimensional maze : ground distance estimation keeps desert ants *Cataglyphis fortis* on course. *The Journal of experimental biology*, 208(Pt 21) :4005–4011. 48, 49
- Grandjean, D. and Peters, C. (2011). Novelty Processing and Emotion : Conceptual Developments, Empirical Findings and Virtual Environments. In Cowie, R., Pelachaud, C., and Petta, P., editors, *Emotion-Oriented Systems*, Cognitive Technologies, pages 441–458. Springer Berlin Heidelberg. 121
- Gratet-Duplessis, P. (1851). *La fleur des proverbes français*. Petite encyclopédie récréative. Passard. 138
- Griffith, J. S. (1963). On the stability of brain-like structures. *Biophysical journal*, 3 :299–308. 34

- Griffiths, P. E. (1997). *What Emotions Really Are : The Problem of Psychological Categories*. Science and Its Conceptual Foundations S. University of Chicago Press. 137
- Grossberg, S. (1972a). A neural theory of punishment and avoidance, I : Qualitative theory. *Mathematical Biosciences*, 15(1-2) :39–67. 29, 122
- Grossberg, S. (1972b). A neural theory of punishment and avoidance, II : quantitative theory. *Mathematical Biosciences*, 15(3–4) :253–285. 122
- Grossberg, S. (1976). Adaptive pattern classification and universal recording : II. Feedback, expectation, olfaction, illusions. *Biological Cybernetics*, 23 :187–202. 30
- Grossberg, S. (1988). Nonlinear neural networks : Principles, mechanisms, and architectures. *Neural Networks*, 1(1) :17–61. 30
- Grossberg, S. E. and Morahan, P. S. (1971). Repression of interferon action : induced dedifferentiation of embryonic cells. *Science (New York, N.Y.)*, 171(966) :77–79. 29
- Gumbel, E. J. (1954). *Statistical theory of extreme values and some practical applications : a series of lectures*. Applied mathematics series. U. S. Govt. Print. Office. 122
- Gweon, H. and Schulz, L. (2011). 16-Month-Olds Rationally Infer Causes of Failed Actions. *Science*, 332(6037) :1524. 155
- Gweon, H., Tenenbaum, J. B., and Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20) :9066–9071. 81
- Haasbroek, N. D. (1968). *Gemma Frisius, Tycho Brahe and Snellius and their triangulations*. Publication of the Netherlands Geodetic Commission. Rijkscommissie voor geodesie. 42
- Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052) :801–6. 22, 52, 88, 89, 94, 102, 103
- Hájek, J. and Šidák, Z. (1967). *Theory of rank tests*. Czechoslovak Academy of science. Academic Press. 122
- Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42 :335–346. 14
- Harris, S., Sheth, S. A., and Cohen, M. S. (2008). Functional neuroimaging of belief, disbelief, and uncertainty. *Annals of neurology*, 63(2) :141–147. 136
- Hart, P., Nilsson, N., and Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on*, 4(2) :100–107. 13
- Hartmann, G. and Wehner, R. (1995). The ant's path integration system : a neural architecture. *Biological Cybernetics*, 73(6) :483–497. 87
- Hasselmo, M. E. and Fehrlau, B. P. (2001). Differences in time course of ACh and GABA modulation of excitatory synaptic potentials in slices of rat hippocampus. *Journal of neurophysiology*, 86(4) :1792–1802. 51

- Hasselmo, M. E. and McClelland, J. L. (1999). Neural models of memory. *Current opinion in neurobiology*, 9(2) :184–188. 123
- Hasselmo, M. E. and Schnell, E. (1994). Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1 : computational modeling and brain slice physiology. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 14(6) :3898–3914. 51
- Hasselmo, M. E., Schnell, E., and Barkai, E. (1995). Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region ca3. *J Neurosci*, 15(7 Pt 2) :5249–5262. 122
- Hasson, C., Gaussier, P., and Boucenna, S. (2011). Emotions as a dynamical system : the interplay between the meta-control and communication function of emotions. *Paladyn*, 2(3) :111–125. 120, 137, 152
- Hatwell, Y. (1987). Motor and Cognitive Functions of the Hand in Infancy and Childhood. *International Journal of Behavioral Development*, 10(4) :509–526. 107
- Hatwell, Y., Steri, A., and Gentaz, E. (2003). *Touching for Knowing*. John Benjamins Publishing Company. 107
- Haykin, S. S. (2002). *Adaptive filter theory*. Prentice-Hall information and system sciences series. Prentice Hall. 109
- Hebb, D., Allport, G., and Kandel, E. (1949). Hebbian theory. *Current Opinion in Neurobiology*, pages 1–4. 29
- Heisenberg, W. (1927). Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Zeitschrift für Physik*, 43(3-4) :172–198. 79
- Held, R. and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *J Comp Physiol Psychol*, pages 872–876. 123
- Herrmann, J. M., Pawelzik, K., and Geisel, T. (1999). Self-Localization of Autonomous Robots by Hidden Representations. *Autonomous Robots*, 7 :31–40. 93
- Hirel, J., Gaussier, P., Quoy, M., Banquet, J. P., Save, E., and Poucet, B. (2013). The hippocampo-cortical loop : spatio-temporal learning and goal-oriented planning in navigation. *Neural networks : the official journal of the International Neural Network Society*, 43 :8–21. 120
- Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4) :500–544. 26
- Issa, J. B. and Zhang, K. (2012). Universal conditions for exact path integration in neural systems. *Proceedings of the National Academy of Sciences*, 109(17) :6716–6720. 87
- J. Haugeland, A. Descartes, D. D. D. D. (1995). Mind embodied and embedded. *Acta Philosophica Fennica*, 58 :233–267. 15

- James, W. (1950). *The Principles of Psychology*. Number vol.~1 in Dover Books on Western Philosophy. Dover Publications. 66
- Jauffret, A., Belkaid, M., Cuperlier, N., Tarroux, P., and Gaussier, P. (2013a). Frustration : a generic mechanism to improve autonomy in robotics. In *third symposium on biology of decision making*, France.
- Jauffret, A., Cuperlier, N., Belkaid, M., Gaussier, P., and Tarroux, P. (2013b). Frustration as a way toward autonomy and self-improvement in robotic navigation. In *The Third Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics. ICDL-EpiRob*, pages 1–7, Japon.
- Jauffret, A., Cuperlier, N., Gaussier, P., and Tarroux, P. (2013c). From self-assessment to frustration, a small step towards autonomy in robotic navigation. *Frontiers in Neurobotics*, 7(16).
- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2011). Toward long range robot navigation. In *The International Workshop on Bio-Inspired Robots.*, Nantes, France.
- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2012a). Multimodal integration of visual place cells and grid cells for navigation tasks of a real robot. In *From Animals to Animats 12*, number 7426 in LNAI, pages 136–145, Odense, Danemark. Springer.
- Jauffret, A., Cuperlier, N., Tarroux, P., and Gaussier, P. (2012b). Multimodal integration of visual place cells and grid cells for robots navigation. In *Second Symposium on Biology of Decision Making*, Paris, France.
- Jauffret, A., Grand, C., Cuperlier, N., Tarroux, P., and Gaussier, P. (2013d). How Can a Robot Evaluate its own Behavior ? A Neural Model for Self-Assessment. pages 1–8, International Joint Conference on Neural Networks (IJCNN).
- Jaulin, L. (2009). A Nonlinear Set Membership Approach for the Localization and Map Building of Underwater Robots. *IEEE Transactions on Robotics*, 25(1). 46
- Jaulin, L. (2011). Range-Only SLAM With Occupancy Maps : A Set-Membership Approach. *IEEE Transactions on Robotics*, 27(5) :1004–1010. 46
- Jean, K. (1990). L'instrument méridien. In *Astrométrie moderne*, volume 358 of *Lecture Notes in Physics*, pages 151–172. Springer Berlin Heidelberg. 44
- Jeffery, K. J. (1998). Learning of landmark stability and instability by hippocampal place cells. *Neuropharmacology*, 37(4-5) :677–687. 49, 89
- Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3) :401–422. 45
- Kagan, J. (1972). Motives and development. *Journal of personality and social psychology*, 22(1) :51–66. 108
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems 1. *Journal Of Basic Engineering*, 82(Series D) :35–45. 44

- Kaplan, F. and Oudeyer, P.-y. (2007). Un robot motivé pour apprendre : le rôle des motivations intrinsèques dans le développement sensorimoteur. *Enfance*, 1 :46–58. 119, 148
- Keeton, W. T. (1971). Magnets interfere with pigeon homing. *Proceedings of the National Academy of Sciences of the United States of America*, 68(1) :102–106. 47
- Kelemen, E. and Fenton, A. A. (2010). Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames. *PLoS biology*, 8(6) :e1000403. 52
- Kelley, S. A., Brownell, C. A., and Campbell, S. B. (2000). Mastery motivation and self-evaluative affect in toddlers : Longitudinal relations with maternal behavior. *Child development*, 71(4) :1061–1071. 137
- Kellman, P. J. and Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive psychology*, 15(4) :483–524. 66
- Kersten, D., Yuille, A., and Mamassian, P. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55 :271–304. 67
- Khatib, O. (1986). Real-Time Obstacle Avoidance For Manipulators and Mobile Robots. *Int. Journ. of Rob. Res.*, 5(1) :90–98. 68
- Klopf, A. and Laboratory, A. F. C. R. L. U. D. S. (1972). *Brain Function and Adaptive Systems—a Heterostatic Theory*. Special reports. Data Sciences Laboratory, Air Force Cambridge Research Laboratories, Air Force Systems Command, United States Air Force. 29
- Klopf, A. H. (1975). A comparison of natural and artificial intelligence. *SIGART Bull.*, (52) :11–13. 29
- Knierim, J. J., Kudrimoti, H. S., and McNaughton, B. L. (1998). Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *Journal of neurophysiology*, 80(1) :425–446. 52
- Knight, R. (1996). Contribution of human hippocampal region to novelty detection. *Nature*, 383(6597) :256–259. 122
- Kohonen, T. (1984). *Self-Organization and Associative Memory*. Springer-Verlag, New York. 30, 32
- Kohonen, T. (1993). Physiological Interpretation of the Self-Organizing Map Algorithm. *Neural Networks*, 6 :895–905. 122
- Kohonen, T. (2001). *Self-Organizing Maps*. Physics and astronomy online library. Springer. 73
- Kohonen, T. and Oja, E. (1976). Fast adaptive formation of orthogonalizing filters and associative memory in recurrent networks of neuron-like elements. *Biological cybernetics*, 21(2) :85–95. 122
- Kolb, B. and Tees., R. (1990). The Cerebral Cortex of the Rat. In *MIT Press*. 57

- Komarova, N. L., Niyogi, P., and Nowak, M. A. (2001). The evolutionary dynamics of grammar acquisition. *Journal of theoretical biology*, 209(1) :43–59. 66
- Kong, H., Audibert, J.-Y., and Ponce, J. (2009). Vanishing point detection for road detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 96–103. 181
- Körding, K. P., Ku, S.-p., and Wolpert, D. M. (2004). Bayesian integration in force estimation. *Journal of neurophysiology*, 92(5) :3161–3165. 67
- Kuipers, B. and Byun, Y.-T. (1991). A Robot Exploration and Mapping Strategy Based on a Semantic Hierarchy of Spatial Representations. *Journal of Robotics and Autonomous Systems*, 8 :47–63. 44
- Lackner, J. R. (1988). Some proprioceptive influences on the perceptual representation of body shape and orientation. *Brain : a journal of neurology*, 111 (Pt 2 :281–297. 108
- Lagarde, M., Andry, P., and Gaussier, P. (2008). Distributed real time neural networks in interactive complex systems. In Chbeir, R., Badr, Y., Abraham, A., Laurent, D., Köppen, M., Ferri, F., Zadeh, L. A., and Ohsawa, Y., editors, *CSTST*, pages 95–100. ACM. 185
- Landau, B., Gleitman, H., and Spelke, E. (1981). Spatial knowledge and geometric representation in a child blind from birth. *Science (New York, N.Y.)*, 213(4513) :1275–1278. 66
- Langston, R. F., Ainge, J. a., Couey, J. J., Canto, C. B., Bjerknes, T. L., Witter, M. P., Moser, E. I., and Moser, M. B. (2010). Development of the Spatial Representation System in the Rat. *Science*, 328(5985) :1576–1580. 101
- Lazarus, R. S. and Kemper, T. D. (1992). Emotion and Adaptation. *Contemporary Sociology*, 21(4) :522. 121
- Lebeltel, O., Bessiere, P., Diard, J., and Mazer, E. (2004). Bayesian Robot Programming. *Autonomous Robots*, 16(1) :49–79.
- LeDoux, J. E. (1992). Brain mechanisms of emotion and emotional learning. *Current Opinion in Neurobiology*, 2(2) :191 – 197. 19, 136
- Lee, D. K., Itti, L., Koch, C., and Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature neuroscience*, 2(4) :375–381. 30
- Lee, J., Kim, Y., Park, C., Park, C., and Paik, J. (2006). Robust Feature Detection Using 2D Wavelet Transform Under Low Light Environment. In Huang, De-ShuangLi, K. and Irwin, G. W., editors, *Intelligent Computing in Signal Processing and Pattern Recognition International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006*, pages 1042–1050. Springer Berlin / Heidelberg. 46
- Leonard, J. and Durrant-Whyte, H. (1991). Simultaneous map building and localization for an autonomous mobile robot. In *Intelligent Robots and Systems '91. 'Intelligence for Mechanical Systems, Proceedings IROS '91. IEEE/RSJ International Workshop on*, pages 1442–1447 vol.3. 41

- Leonard, J. and Newman, P. (2003). Consistent, convergent, and constant-time SLAM. In *International Joint Conference on Artificial Intelligence*, volume 18, pages 1143–1150. Lawrence Erlbaum Associates LTD. 44, 45
- Leprêtre, S., Gaussier, P., and pierre Cocquerez, J. (2000). From navigation to active object recognition. *SAB2000, Zurich*. 57
- Levine, D. S. and Prueitt., P. S. (1992). Simulations of conditioned perseveration and novelty preference from frontal lobe damage. In *Michael L. Commons, Stephen Grossberg, and John E.R. Staddon, editors, Neural Network Models of Conditioning and Action*, chapter 5 :123–147. 122
- Levitt, T. S. and Lawton, D. T. (1990). Qualitative Navigation for Mobile Robots. *Artif. Intell.*, 44(3) :305–360. 42
- Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *The Behavioral and brain sciences*, 28(2) :169–194 ; discussion 194–245. 121
- Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*. 30
- Lisman, J. E. and Otmakhova, N. A. (2001). Storage, recall, and novelty detection of sequences by the hippocampus : elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus*, 11(5) :551–568. 122
- Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., and Fry, P. A. (1993). Nonvisual navigation by blind and sighted : assessment of path integration ability. *Journal of experimental psychology. General*, 122(1) :73–91. 49
- Luyat, M. and Gentaz, E. (2002). Body tilt effect on the reproduction of orientations : studies on the visual oblique effect and subjective orientations. *Journal of experimental psychology. Human perception and performance*, 28(4) :1002–1011. 108
- M. Mirolli, G. B. (2013). Functions and mechanisms of intrinsic motivations : the knowledge versus competence distinction. *Springer Verlag, Berlin*, pages 49–72. 120
- Ma, B., Lakshmanan, S., and Hero, A. O. (2000). Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion. *IEEE Transactions on Intelligent Transportation Systems*. 181
- Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat Neurosci*, 9(11) :1432–1438. 68
- Maass, W. (2000). On the computational power of winner-take-all. *Neural Computation*, 12(11) :2519–2535. 31
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Cam, L. M. L. and Neyman, J., editors, Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press. 31

- Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., and Okeefe, J. (1998). Knowing where and getting there. *Science*, 280 :921–924. 49
- Maillard, M., Gapenne, O., Gaussier, P., and Hafemeister, L. (2005). Perception as a dynamical sensori-motor attraction basin. In Et al., C., editor, *Advances in Artificial Life (8th European Conference, ECAL)*, volume 3630 of *Lecture Note in Artificial Intelligence*, pages 37–46. Springer. 124
- Malika Auvray, J. K. O. (2003). Voir avec les oreilles. enjeux de la substitution sensorielle. *Pour la science*. 107
- Mallot, H. A. and Franz, M. O. (2000). Biomimetic robot navigation. *Robotics and Autonomous Systems*, 30(1-2) :133–153. 53, 54
- Mamassian, P., Landy, M. S., and Maloney, L. T. (2002). Bayesian modelling of visual perception. In Rao, R., Lewicki, M., and Olshausen, B., editors, *Probabilistic Models of the Brain ; Perception and Neural Function.*, pages 13–36. Cambridge, MA : MIT Press. 67
- Mannella, F., Zappacosta, S., Mirolli, M., and Baldassarre, G. (2008). A Computational Model of the Amygdala Nuclei's Role in Second Order Conditioning. In Asada, M., Hallam, J., Meyer, J.-A., and Tani, J., editors, *From Animals to Animats : Proceedings of the International Conference on the Simulation of Adaptive Behaviour (SAB)*, volume 5040 of *Lecture Notes in Computer Science*, pages 321–330. Springer-Verlag. 19, 137
- Marieb, E. (1993). *Anatomie et physiologie humaines*. De Boeck-Université. De Boeck université. 106
- Marsland, S. (2002). Novelty Detection in Learning Systems. *Computing Surveys*, 3 :1–39. 121, 122
- Marsland, S., Duckett, T., and Nehmzow, U. (2001). Learning to select distinctive landmarks for mobile robot navigation. *Robotics and Autonomous Systems*, 37(4) :241–260. 64, 65
- Marsland, S., Nehmzow, U., and Shapiro, J. (2005). On-line novelty detection for autonomous mobile robots. *Robotics and Autonomous Systems*, 51(2-3) :191–206. 122
- Martinetz, T. and Schulten, K. (1991). A "Neural-Gas" Network Learns Topologies. *Artificial Neural Networks*, I :397–402. 122
- Mataric, M. J. (1990). A Distributed Model for Mobile Robot Environment-Learning and Navigation. *Security*. 93
- Mataric, M. J. (1992). Integration of representation into goal-driven behavior-based robots. *Robotics and Automation, IEEE Transactions on*, 8(3) :304–312. 44
- Mataruna, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition : the Realization of the Living*. Reidel, Dordrecht. 15, 16, 91
- McCorduck, P. (2004). *Machines who think : a personal inquiry into the history and prospects of artificial intelligence*. Ak Peters Series. A.K. Peters. 13

- McCulloch, W. S., Landahl, H. D., and Pitts, W. (1943). A statistical consequence of the logical calculus of nervous nets. *The Bulletin of Mathematical Biophysics*, 5(4) :135–137. 26
- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264 :746–748. 107
- McNaughton, B., Battaglia, F., Jensen, O., Moser, E., and Moser, M. (2006). Path integration and the neural basis of the cognitive map. *Nature Reviews Neurosci.*, 7 :663–678. 89, 90, 113
- Mehler, J. and Bever, T. G. (1967). Cognitive Capacity of Very Young Children. *Science*, 158(3797) :141–142. 66
- Melchor J. Antunano, M. D. (2005). Medical Facts for Pilots. *Civil Aerospace Medical Institute*, Publication : AM-400-03/1. 108
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Bibliothèque des idées. Gallimard. 123
- Milford, M. J. and Wyeth, G. F. (2008). Biologically Inspired SLAM System. *October*, 24(5) :1038–1053. 55, 90
- Minsky, M. (1967). *Computation : finite and infinite machines*. Prentice-Hall series in automatic computation. Prentice-Hall. 13
- Miroslav Kubat, Joao Gama, P. U. (2004). *Incremental learning and concept drift : Editor's introduction*, volume 8. IOS Press. 65
- Mittelstaedt, H. (2000). Triple-loop model of path control by head direction and place cells. *Biological cybernetics*, 83(3) :261–270. 61, 87
- Moravec, H. (1988). *Mind Children : The Future of Robot and Human Intelligence*. Harvard University Press. 13
- Moravec, H. and Elfes, A. (1985). High resolution maps from wide angular sensors. In *Proceedings of the IEEE International Conference On Robotics and Automation ICRA85*, volume 2, pages 116–121. IEEE Press. 44
- Müller, M. and Wehner, R. (1988). Path integration in desert ants, *Cataglyphis fortis*. *Proceedings of the National Academy of Sciences of the United States of America*, 85(14) :5287–5290. 48, 49
- Muller, R. U. and Kubie, J. L. (1987). The Effects of Changes in the Environment Hippocampal Cells on the Spatial Firing of. *New York*, 7(7) :1951–1968. 51
- Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological cybernetics*, 66(3) :241–251. 81
- Neto, H. (2006). *Visual Novelty Detection for Autonomous Inspection Robots*. PhD Thesis. University of Essex. 122

- Newborn, M. and Newborn, M. (1997). *Kasparov Vs. Deep Blue : Computer Chess Comes of Age*. Springer-Verlag New York, Inc., Secaucus, NJ, USA. 13
- Newman, P. and Cummins, M. (2008). FAB-MAP : Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, 27(6) :647–665. 56
- Nguyen, V., Harati, A., Martinelli, A., Siegwart, R., and Tomatis, N. (2006). Orthogonal SLAM : a Step toward Lightweight Indoor Autonomous Navigation. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 46
- Nilsson, N. (1998). *Artificial Intelligence : A New Synthesis*. The Morgan Kaufmann Series in Artificial Intelligence Series. Morgan Kaufmann Publishers. 13
- Nilsson, N. J. (1984). Shakey the Robot. *SRI AI Center Technical Note*. 44
- Nourbakhsh, I. and Fong, T. (2003). Socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4) :139–141. 120
- O’Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 34(1) :171–175. 50
- O’Keefe, J. and Nadel, L. (1978). *The hippocampus as a cognitive map*. Clarendon Press, Oxford, United Kingdom. 52, 88
- Oudeyer, P.-Y., Kaplan, F., and Hafner, V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, 11(2). 19
- Oudeyer, P.-Y., Kaplan, F., Hafner, V. V., and Whyte, A. (2005). The playground experiment : Task-independent development of a curious robot. In Bank, D. and Meeden, L., editors, *Proceedings of the AAAI Spring Symposium on Developmental Robotics, 2005*, pages 42–47, Stanford, California. 67
- P. Andry, P. G. e. J. N. (2002). From Visuo-Motor Development to Low-level Imitation. In *Proceedings of the second workshop on Epigenetic Robotics (Lund University Cognitive Studies, 94)*, pages 2–15. 19, 120
- P. Gaussier, C. Joulain, S. Z. J. B. A. R. (1997). Visual navigation in an open environment without map. In *IROS’97*, pages 545–550. 16
- Panksepp, J. (1998). *Affective Neuroscience : The Foundations of Human and Animal Emotions : The Foundations of Human and Animal Emotions*. Series in Affective Science. Oxford University Press, USA. 19, 136
- Pavlov, I. (1927). Conditioned reflexes : An investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences*, 17(3). 26, 125
- Pfeifer, R. and Verschure, P. (1994). The Artificial Life Route to Artificial Intelligence, chapter The challenge of autonomous systems : Pitfalls and how to avoid them. *MIT Press, Cambridge, MA*. 124

- Pfister, J.-P., Barber, D., and Gerstner, W. (2003). Optimal hebbian learning : A probabilistic point of view. In Kaynak, O., Alpaydin, E., Oja, E., and Xu, L., editors, *ICANN*, volume 2714 of *Lecture Notes in Computer Science*, pages 92–98. Springer. 68
- Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*. Delachaux & Niestlé. 47
- Piaget, J. (1967). *Logique et connaissance scientifique : sous la direction de Jean Piaget*. Encyclopédie de la Pléiade, v. 22. Editions Gallimard. 66
- Picard, R. W. (1997). *Affective Computing*. MIT Press, Cambridge, MA. 19, 136
- Pick, A. D. (1965). Improvement of visual and tactual form discrimination. *Journal of experimental psychology*, 69 :331–339. 107
- Poole, D., Mackworth, A., and Goebel, R. (1998). *Computational Intelligence : A Logical Approach*. Oxford University Press. 13
- Postman, L. and Phillips, L. (1964). *Short-term Temporal Changes in Free Recall*. Defense Technical Information Center. 81
- Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). Probabilistic brains : knowns and unknowns. *Nat Neurosci*, 16(9) :1170–1178. 67
- Pouget, A., Deneve, S., and Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3 :741–747. 109
- Power, R. P. and Graham, A. (1976). Dominance of touch by vision : generalization of the hypothesis to a tactually experienced population. *Perception*, 5(2) :161–166. 107
- Quirk, G. J., Muller, R. U., and Kubie, J. L. (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 10(6) :2008–2017. 52
- Quirk, G. J., Muller, R. U., Kubie, J. L., and Ranck, J. B. (1992). The positional firing properties of medial entorhinal neurons : description and comparison with hippocampal place cells. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 12(5) :1945–1963. 52, 60, 88
- Quoy, M., Moga, S., and Gaussier, P. (2003). Dynamical neural networks for planning and low-level robot control. *IEEE Transactions on Systems, Man, and Cybernetics - Part A : Systems and Humans*, 33(4). 138
- Quoy, M., Moga, S., Gaussier, P., and Revel, A. (2000). Parallelization of neural networks using pvm. In Dongarra, J., Kacsuk, P., and Podhorszki, N., editors, *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, volume 1908 of *Lecture Notes in Computer Science*, pages 289–296. Springer Berlin Heidelberg. 185
- Ramachandran, V. S. (1998). *Phantoms in the Brain : Probing the Mysteries of the Human Mind*. HarperCollins. 106, 136

- Rao, R., Olshausen, B., and Lewicki, M. (2002). *Probabilistic Models of the Brain : Perception and Neural Function*. Bradford book. MIT Press. 68
- Rédei, G. (2008). *Encyclopedia of Genetics, Genomics, Proteomics, and Informatics*. Number vol. 1 in *Encyclopedia of Genetics, Genomics, Proteomics, and Informatics*. Springer. 51
- Redish, A. D. and Touretzky, D. S. (1997). Cognitive maps beyond the hippocampus. *Hippocampus*, 7(1) :15–35. 55
- Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological review*, 74(1) :71–80. 29
- Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *Am Psychol*, 43(3) :151–160. 63
- Revel, A. (1997). *Contrôle d'un robot autonome par approche neuro-mimétique*. PhD Thesis. University of Cergy-Pontoise. 181
- Richefeu, J. and Manzanera, A. (2004). A New Hybrid Differential Filter For Motion Detection. In *Proceedings of Computer Vision and Graphics (ICCVG)*. 125
- Roberts, S. and Tarassenko, L. (1994). A Probabilistic Resource Allocating Network for Novelty Detection. *Neural Computation*, 6(2) :270–284. 122
- Rosenblatt, F. (1958). The perceptron : A probabilistic model for information storage and organization in the brain. *Psychological Review*. 26
- Ruffier, F. and Franceschini, N. H. (2005). Optic flow regulation : the key to aircraft automatic guidance. *Robotics and Autonomous Systems*, 50(4) :177–194. 54
- Rumelhart, D. E. and Zipser, D. (1986). *Feature discovery by competitive learning*, volume 1, chapter 5, pages 151–193. MIT Press, Cambridge, MA, USA. 30
- Russell, J. and Cohn, R. (2012). *Moravec's Paradox*. Book on Demand. 15
- S. Leprêtre, P. Gaussier, J. C. (1999). Un modèle neuronal pour la reconnaissance d'objets. In *GRETSI 99*, pages 435–438. 17
- Santucci, V. G., Baldassarre, G., and Mirolli, M. (2012). Intrinsic motivation mechanisms for competence acquisition. In *ICDL-EPIROB*, pages 1–6. IEEE. 120, 148
- Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B., Moser, M., and Moser, E. (2006). Conjunctive representation - supporting material. *Science*, 312 :758–762. 52
- Save, E., Cressant, A., Thinus-Blanc, C., and Poucet, B. (1998). Spatial firing of hippocampal place cells in blind rats. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18(5) :1818–1826. 52
- Scheier, C. and Pfeifer, R. (1997). Information theoretic implications of embodiment for neural network learning. In *ICANN*, pages 691–696. 15

- Schembri, M., Mirolli, M., and Baldassarre, G. (2007). Evolving internal reinforcers for an intrinsically motivated reinforcement-learning robot. *2007 IEEE 6th International Conference on Development and Learning*. 120, 148
- Schenk, F. (2009). Les émotions de la raison. *Revue européenne des sciences sociales XLVII-144*. 136
- Scherer, K. (1994). Toward a concept of “modal emotions”. *The nature of emotion : Fundamental questions*, pages 25–31. 121
- Schmidhuber, J. (1991). Curious model-building control systems. *[Proceedings] 1991 IEEE International Joint Conference on Neural Networks*. 119, 135, 148
- Schöner, G., Dose, M., Engels, C., Robotics, E., and Systems, A. (1995). Autonomous Systems Dynamics of behavior : theory and applications for autonomous robot architectures. *Robotics and Autonomous Systems*, 16(2-4) :213–245. 35, 36, 138
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in cognitive sciences*, 12(3) :106–113. 107
- Schulz, L. E. and Gopnik, A. (2004). Causal learning across domains. Technical Report 2, University of California, Berkeley, CA, USA. laurasch@socrates.berkeley.edu. 66
- Scoville, W. B. and Milner, B. (2000). Loss of recent memory after bilateral hippocampal lesions. 1957. *The Journal of neuropsychiatry and clinical neurosciences*, 12(1) :103–113. 50
- Searle, J. R. (1999). Chinese room argument. In Wilson, R. A. and Keil, F. C., editors, *The MIT Encyclopedia of the Cognitive Sciences*, pages 115–116. MIT Press, Cambridge, MA. 14
- Seguinot, V., Cattet, J., and Benhamou, S. (1998). Path integration in dogs. *Animal behaviour*, 55(4) :787–97. 49
- Seymour, B., O’Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., and Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature neuroscience*, 8(9) :1234–1240. 81
- Sharp, P. (1999a). Complementary Roles For Hippocampal Versus Subicular/enthorinal Place Cells in Coding Place, Context, and Events. *Hippocampus*, 4 :432–443. 88
- Sharp, P. E. (1999b). Complimentary roles for hippocampal versus subicular/entorhinal place cells in coding place, context, and events. *Hippocampus*, 9(4) :432–443. 52, 60
- Shelley, M. (1823). *Frankenstein : or, The modern Prometheus*. Lackington, Allen Co. 13
- Simon, H. (1965). *The shape of automation for men and management*. Harper & Row. 13
- Skinner, B. F. (1938). *The behavior of organisms : and experimental analysis / by B. F. Skinner*. Appleton-Century-Crofts New York. 27
- Smith, B. (1988a). *Foundations of Gestalt theory*. Philosophia Verlag. 123

- Smith, B. (1988b). Knowing versus Knowing That. *Practical Knowledge*. 45
- Smith, R., Self, M., and Cheeseman, P. (1987). Estimating uncertain spatial relationships in robotics. *Proceedings. 1987 IEEE International Conference on Robotics and Automation*, 4, 41, 44
- So, R. and Lo, W. T. (1999). Cybersickness : an experimental study to isolate the effects of rotational scene oscillations. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 237–241. 108
- Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., and Moser, E. I. (2008). Representation of Geometric Borders in the Entorhinal Cortex. *Science*, 322(5909) :1865–1868. 102
- Sparbert, J., Dietmayer, K., and Streller, D. (2001). Lane detection and street type classification using laser range images. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, pages 454–459. 181
- Spencer, H. (1870). *The Principles of Psychology*. Number vol. 1 in A system of synthetic philosophy ; vol. 4-5. Williams and Norgate. 20
- Sreenivasan, S. and Fiete, I. (2011). Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat Neurosci*, 14(10) :1330–1337. 90
- Srinivasan, M. and Zhang, S. W. (2003). Small brains, smart minds : Vision, perception and 'cognition' in honeybees. *Iete Journal of Research*, 49(2-3) :127–134. 47, 49
- Stein, B. and Meredith, A. (1993). *The Merging of the Senses*. Bradford books. Mit Press. 107
- Steux, B. and Hamzaoui, O. E. (2010). tinySLAM : A SLAM algorithm in less than 200 lines C-language program. *2010 11th International Conference on Control Automation Robotics & Vision*, pages 1975–1979. 46
- Stipek, D., Recchia, S., and McClintic, S. (1992). Self-evaluation in young children. *Mono-graphs of the Society for Research in Child Development*, 57(1) :1–98. 137
- Strösslin, T., Sheynikhovich, D., Chavarriga, R., and Gerstner, W. (2005). Robust self-localisation and navigation based on hippocampal place cells. *Neural networks : the official journal of the International Neural Network Society*, 18(9) :1125–1140. 93, 110
- Sun, T.-Y., Tsai, S.-J., and Chan, V. (2006). Hsi color model based lane-marking detection. In *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, pages 1168–1172. 181
- Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. pages 9–44. Kluwer Academic Publishers. 29
- Suzuki, W. A., Miller, E. K., and Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *Journal of neurophysiology*, 78(2) :1062–1081. 57
- Taylor, S. E., Wayment, H. A., and Neter, E. (1995). Self-Evaluation Processes. *Personality and Social Psychology Bulletin*, 21(12) :1278–1287. 121

- Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., and Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science (New York, N.Y.)*, 332(6033) :1054–1059. 121
- Thompson, L. T. and Best, P. J. (1990). Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats. *Brain research*, 509(2) :299–308. 51
- Thrun, S. (1995). An Approach to Learning Mobile Robot Navigation. *Robotics and Autonomous Systems*, 15 :301–319.
- Thrun, S. (2002a). Probabilistic robotics. *Communications of the ACM*, 45(3). 45, 68
- Thrun, S. (2002b). Robotic Mapping : A Survey. *Science*, 298(February) :1–35. 44, 45, 46, 67
- Thyryon, C. and Roll, J. (2009). Perceptual integration of illusory and imagined kinesthetic images. *Journal of Neuroscience*, 29(26) :8483–92. 108
- Tinbergen, N. (1969). *The Study of Instinct*. Clarendon Press. 47, 48
- Tolman, E. C. and Honzik, C. H. (1930). Introduction and removal of reward and maze performance in rats. *University of California Publications in Psychology*, 4 :257–275. 52
- Torralba, A. (1999). *Analogue Architectures For Vision : Cellular Neural Networks and Neuro-morphic Circuits*. PhD thesis, TIRF - Institut Polytechnique de Grenoble, Grenoble, France. 68
- Treisman, M. (1977). Motion sickness : an evolutionary hypothesis. *Science*, 29 ;197(4302) :493–5. 108
- Tronick, E. Z. (1989). Emotions and emotional communication in infants. *The American psychologist*, 44(2) :112–119. 137
- Trullier, O. and Meyer, J. A. (1997). Biomimetic navigation models and strategies in animats. *AI communications*, 10(2) :79–92. 53
- Tsymbal, A. (2004). The problem of concept drift : definitions and related work. Technical Report TCD-CS-2004-15, The University of Dublin, Trinity College, Department of Computer Science, Dublin, Ireland. 65
- Ullah, M. M., Pronobis, A., Caputo, B., Luo, J., Jensfelt, P., and Christensen, H. I. (2008). Towards robust place recognition for robot localization. In *ICRA*, pages 530–537. IEEE. 68
- Varela, F. (1996). *Invitation aux sciences cognitives*. Collection Points. Série Sciences. Éd. du Seuil. 14
- Varela, F., Thompson, E., and Rosch, E. (1993). *The Embodied Mind*. MIT Press. 16, 91
- Von Frisch, K. (1955). *The Dancing Bees : An Account of the Life and Senses of the Honey Bee*. A Harvest book. Harcourt, Brace & World. 47

- Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les paralléloèdres primitifs. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908(134) :198–287. 32
- Wacongnea, C., Labyta, E., Van Wassenhovea, V., Bekinschteind, T., Naccachee, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceeding of the national academy of sciences*, 108(51) :20754–20759. 121
- Walsh, J. K. (1973). Effect of visual and tactual stimulation on learning abstract forms : A replication. *Bulletin of Psychonomic Science*, 2 :357–359. 107
- Walter, W. G. (1953). *The Living Brain*. Norton. 15, 41, 54
- Wang, Y., Teoh, E. K., and Shen, D. (2004). Lane detection and tracking using b-snake. *Image and Vision Computing*, 22(4) :269 – 280. 181
- Watkins, C. J. C. H. and Dayan, P. (1992). Technical note q-learning. *Machine Learning*, 8 :279–292. 29
- Wehner (1994). The polarization-vision project : championing organismic biology. In *Neural basis of behavioural adaptations. Fortschritte der Zoologie 39* (ed. K. Schildberger and N. Elsner), pages 103–143. 49
- Wehner, R. and Räber, F. (1979). Visual spatial memory in desert ants, *Cataglyphis bicolor* (Hymenoptera : Formicidae). *Cellular and Molecular Life Sciences*, 35(12) :1569–1571. 47, 48
- Wehner, R. and Srinivasan, M. (1981). Searching behaviour of desert ants, genus *Cataglyphis* (Formicidae, Hymenoptera). *Journal of comparative physiology*, 142(3) :315–338. 49
- Widmer, G. and Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1) :69–101. 65
- Widrow, B. and Hoff, M. E. (1960). Adaptive Switching Circuits. In *1960 {IRE} {WESCON} Convention Record, Part 4*, pages 96–104, New York. IRE. 27, 125
- Wiener, N. (1948). *Cybernetics ; or, Control and communication in the animal and the machine*. Actualités scientifiques et industrielles. J. Wiley. 15
- Wiering, M. and van Otterlo, M. (2012). *Reinforcement Learning : State-of-the-Art*. Adaptation, Learning, and Optimization. Springer. 29
- Williams, B., Cummins, M., Neira, J., Newman, P., Reid, I., and Tardós, J. (2009). A comparison of loop closing techniques in monocular SLAM. *Robotics and Autonomous Systems*. 45
- Wilson, H. R. and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2) :55–80. 34
- Wilson, M. A. and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science (New York, N.Y.)*, 261(5124) :1055–1058. 51

- Wilson, S. W. (1991). The animat path to AI. In *Proceedings of the first international conference on simulation of adaptative behavior : From animals to animats*, pages 15–20. 15
- Witten, I. and Frank, E. (2005). *Data Mining : Practical Machine Learning Tools and Techniques, Second Edition*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science. 65
- Wittlinger, M., Wehner, R., and Wolf, H. (2006). The ant odometer : stepping on stilts and stumps. *Science (New York, N.Y.)*, 312(5782) :1965–1967. 48, 49
- Wittmann, T. and Schwegler, H. (1995). Path integration - A network model. *Biological Cybernetics*, 73(6) :569–575. 87, 91
- Wolpert, L. (2002). Love is a many-moleculed thing. *The Observer*. 47
- Wundt, W. and Judd, C. (1897). *Outlines of psychology*. W. Engelmann. 20
- Wystrach, A., Mangan, M., Philippides, A., and Graham, P. (2013). Snapshots in ants ? New interpretations of paradigmatic experiments. *The Journal of experimental biology*, 216(Pt 10) :1766–70. 47
- Xu, F. and Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, 105(13) :5012–5015. 67
- Yu, B. and Jain, A. (1997). Lane boundary detection using a multiresolution hough transform. In *Image Processing, 1997. Proceedings., International Conference on*, volume 2, pages 748–751 vol.2. 181
- Zhang, S. and Manahan-Vaughan, D. (2013). Spatial Olfactory Learning Contributes to Place Field Formation in the Hippocampus. *Cerebral Cortex*. 52
- Zillich, M., Prankl, J., Morwald, T., and Vincze, M. (2011). Knowing your limits - self-evaluation and prediction in object recognition. *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 813–820. 121

Strategie de suivi de route

A.1 Un algorithme minimaliste pour réaliser un comportement de suivi de route

Dans cette section, nous allons présenter une stratégie rapide, robuste et biologiquement plausible de suivi de route. Actuellement, la plupart des méthodes tentant de résoudre ce problème se basent sur l'extraction des bords de la route ou du chemin à suivre. Pour détecter ces bords, de nombreuses techniques ont été proposées comme celles basées sur des lasers [Sparbert et al., 2001], sur des radars [Ma et al., 2000], sur la vision stéréoscopique [Bertozzi and Broggi, 1998], sur des différences de couleurs [Sun et al., 2006], à base d'une transformé de Hough [Yu and Jain, 1997], d'un panel de filtres orientables [Freeman and Adelson, 1991] et de modèles de splines [Wang et al., 2004]. Toutes ces méthodes ne sont guère adaptées lorsque la route à suivre ne présente pas de bordures significatives. [Kong et al., 2009] propose un autre algorithme qui permet de trouver des points de fuite dans l'image et de leur accorder des niveaux de confiances en utilisant des techniques de segmentation de routes. Mais cette méthode requiert l'utilisation de filtres de Gabor pour extraire les informations de textures et de bordures de la route. De tels filtres impliquent un coût de calcul élevé, ce qui se révèle problématique lorsque l'on considère un robot qui doit être capable d'exécuter un grand nombre d'autres fonctionnalités en temps réel.

Ici, nous proposons une méthode plus simple mais assez efficace et qui présente l'intérêt d'être biologiquement plausible. Ce modèle est une amélioration du modèle neuronal proposé à la base par [Revel, 1997]. L'algorithme consiste à trouver, parmi N (nombre fixe) points de fuite potentiels, le meilleur candidat (voir fig. A.1). Par exemple, si l'on considère 5 points de fuite potentiels régulièrement espacés sur une ligne d'horizon, le modèle neuronale comprend alors 5 neurones dont chacun d'eux intègre l'activité des contours convergents vers le point de fuite considéré. Chaque neurone répond donc en fonction de la reconnaissance d'un point de fuite dans la zone de l'image qu'il observe. Nous appellerons ces neurones des cellules de point de fuite. La cellule la plus active est celle qui reconnaît le mieux la forme d'un point de fuite (c'est à dire les contours convergents dans la zone qu'elle observe). Un simple mécanisme de compétition (WTA) permet de mettre en avant la cellule gagnante.

Pour calculer l'activité d'une cellule de point de fuite, notre méthode consiste à faire la somme des contours alignés avec 4 directions privilégiés pour cette cellule. L'algorithme est le suivant :

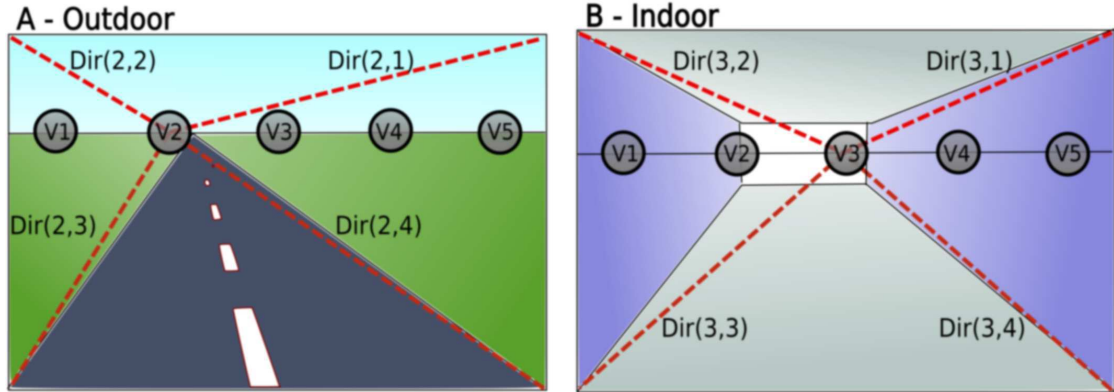


FIGURE A.1 – Illustration de principe pour 5 cellules de point de fuite (respectivement V1, V2, V3, V4 et V5). Les cellules répondent pour des points de fuite potentiels régulièrement espacés sur une ligne d’horizon. L’activité d’une cellule dépend de la quantité de gradient dans l’image qui soit aligné avec les 4 directions privilégiées (pointillés rouges) du point de fuite correspondant. Le robot se dirige vers le point de fuite dont la cellule est la plus active. **A** - Exemple en extérieur. La cellule 2 est gagnante puisque les contours de la route sont alignés avec le point de fuite V2 correspondant. **B** - Exemple en intérieur. Les contours d’un couloir peuvent avoir du sens même au dessus de l’horizon lorsqu’il s’agit d’identifier la position du point de fuite. La cellule V3 est la plus active.

- Acquisition d’une image en basse résolution (160*120 pixels) par une caméra grand angle (Champ de vision à 90 degrés).
- Extraction des contours par calcul du gradient de l’image.
- Sélection des points de fuite potentiels : division de l’image en N parties selon le plan horizontal. Chaque point de fuite potentiel est placé sur une ligne d’horizon, au centre de sa partie. Un paramètre fixé empiriquement détermine l’élévation de la ligne d’horizon, puisqu’il n’y a pas de détection automatique de l’horizon pour le moment. L’élévation est fixé à 20% du haut de l’image dans nos expériences.
- 4 directions privilégiés sont déterminées pour chaque point de fuite considéré (voir Fig. A.1).
- Chaque point de fuite V_n est ensuite représenté par un neurone $ACT(n)$ dont l’activité est donnée par l’équation suivante :

$$ACT(n) = \text{dir}(n, 1) \times \text{dir}(n, 2) + \text{dir}(n, 3) \times \text{dir}(n, 4) \quad (\text{A.1})$$

$$\text{dir}(n, i) = \sum_{l=1}^N f(\cos(\theta_{vp}(i, n, l)) \times (1 - \|\cos(\theta_g(n, l))\|)) \quad (\text{A.2})$$

avec :

$$1. f(x) = \begin{cases} x & \text{for } 0.9 < x < 1 \\ 0 & \text{for } x < 0.9 \\ 1 & \text{for } x > 1 \end{cases}$$

La valeur de 0.9 à été choisie pour permettre une généralisation ou erreur de 25° pour chaque directions privilégiée.

2. $\theta_{vp}(i, n, l)$ correspond à l'angle entre la direction privilégiée et le vecteur { pixel ; point de fuite }
3. $\theta_g(n, l)$ correspond à l'angle entre le vecteur { pixel ; point de fuite } et le vecteur du gradient.
4. $\text{dir}(n,1)$, $\text{dir}(n,2)$, $\text{dir}(n,3)$ et $\text{dir}(n,4)$ sont les directions privilégiées du neurone n .

– Puis, une simple compétition WTA permet de sélectionner la cellule gagnante, qui représente la meilleure candidate.

Le contrôle moteur du robot est ensuite directement inspiré par les théories du contrôle utilisées par les véhicules de Braitenberg [Braitenberg, 1984]. Ce contrôle est assez simple : lorsque le point de fuite est détecté à droite (resp. à gauche), le robot tourne à droite (resp. à gauche). Un comportement convergent cohérent émerge de la boucle sensorimotrice entre le robot et son environnement, sans besoin de recourir à une quelconque représentation interne de l'environnement, ni aucune inférence. Par conséquent, la précision de discrimination angulaire est beaucoup moins importante que la fréquence d'exécution de la boucle dans un tel système. Nous avons testé cette algorithmes sur plusieurs images réelles de route (voir Fig.A.2) dans différentes situations.

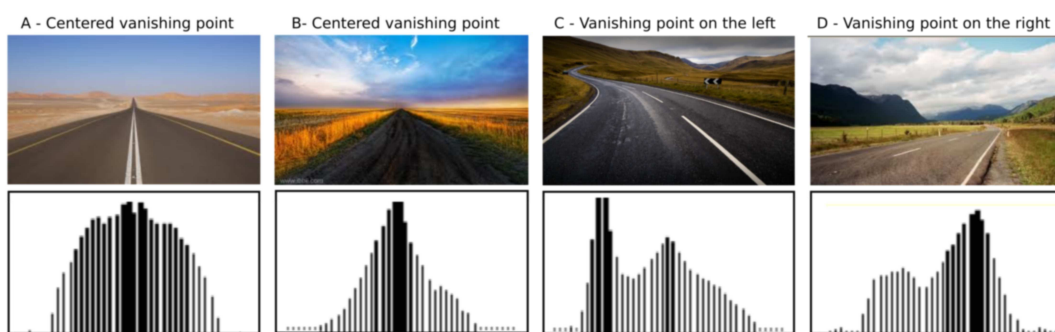


FIGURE A.2 – Détection du point de fuite dans plusieurs images de test. Pour chaque cas : **Haut**– Paysages réelles de route. **Bas**– Niveau d'activité de 41 cellules de point de fuite régulièrement espacées sur l'horizon. Chaque cellule répond à la détection d'un point de fuite pour un angle horizontal particulier. **A**– Dans un cas simple (route avec bordures significatives), le point de fuite est bien détecté au centre de l'image et l'activité se généralise bien sur le voisinage. **B**– L'activité est beaucoup plus saillante dans cet exemple. **C**– Pour une route sinueuse, 2 points de fuite sont détectés, l'un au centre, l'autre à gauche. Le plus actif des 2 est toutefois bien celui de gauche. **D**– Exemple d'un point de fuite détecté à droite.

Grâce à cette méthode, notre système est pleinement capable de suivre n'importe quel type de point de fuite incluant routes, couloirs, chemin ou voies ferrées. De plus, l'algorithme s'exécute à une fréquence très satisfaisante de 20 images par seconde (pour 41 cellules de point de fuite) et cette fréquence peut augmenter si l'on considère moins de neurones. Une telle fréquence est possible grâce au fait que l'algorithme ne considère que les valeurs les plus fortes du gradient. L'intensité du gradient a été normalisée en utilisant le cosinus de l'angle. La figure B A.2 ne présente qu'un faible gradient alors que le point de fuite reste correctement détecté.

Un inconvénient de notre méthode est l'ajustement manuel et empirique de l'élévation de la ligne d'horizon. Ce paramètre pourrait être adapté automatiquement par extraction de la ligne de l'horizon dans l'image en temps réel. Une autre méthode, plus plausible, serait de considérer

une matrice de points de fuite répartis sur l'ensemble de l'image, plutôt que de ne considérer qu'un vecteur sur l'horizon. Ce système permettrait également à un robot de suivre un point de fuite en 3 dimensions, à l'instar d'une mouche. Cela peut être intéressant si l'on considère le contrôle de drones volants plutôt que des plate-formes robotiques à roues.

Publications personnelles

- Jauffret, A., Grand, C., Cuperlier, N., Tarroux, P., and Gaussier, P. (2013d). How Can a Robot Evaluate its own Behavior ? A Neural Model for Self-Assessment. pages 1–8, International Joint Conference on Neural Networks (IJCNN)

Outils de simulation de réseaux de neurones

Les travaux présentés dans cette thèse ont été réalisés en majorité par le biais de 2 outils développés par le laboratoire ETIS : *Coeos* un environnement graphique de conception d'architectures neuronales ainsi que *Promethe* un simulateur temps-réel et distribué de réseaux de neurones [Lagarde et al., 2008; Quoy et al., 2000].

B.1 Coeos : une interface graphique de conception d'architectures neuronales

Coeos est un environnement graphique de conception de réseaux de neurones artificiels. Cet outil permet la construction et la compilation de fichier, incluant la structure du réseau de neurones, afin d'être par la suite interprété par le simulateur Promethe. Son fonctionnement repose sur l'édition de groupes neuronaux connectés entre eux via des liens à poids variables (symbolisant les connections synaptiques). Le logiciel autorise le choix du type de connectivité (connexion plastique, réflexe/inconditionnel, un vers tous, un vers un, un vers voisinage). La visualisation est celle d'un graphe orienté dont les groupes neuronaux représentent les nœuds et les liens synaptiques, les arêtes (voir fig. B.1).

Une architecture neuronale peut être scindée en plusieurs parties (ou scripts) afin de rendre la modélisation modulaire (e.g. un script peut être dédié aux traitements visuels tandis qu'un autre s'occupe du contrôle moteur du robot). Ces scripts communiquent entre eux de manière synchrone ou asynchrone par le biais d'un protocole réseau. Ils peuvent être exécutés sur des machines différentes, permettant ainsi de répartir la charge de calcul.

Les groupes peuvent être neuronaux, c'est à dire suivre des règles d'apprentissage propre aux réseaux de neurones artificiels, ou bien de type algorithmique, correspondant alors à des fonctions ad-hoc d'une tâche ou d'un système matériel particulier. Les groupes algorithmiques n'apprennent pas et proposent généralement une optimisation d'un calcul qui serait largement sous-optimal s'il était modélisé sous forme neuronale. Certains groupes algorithmiques permettent également de s'interfacer avec le matériel embarqué du robot (caméra, carte de contrôle des servomoteurs, capteurs, etc...).

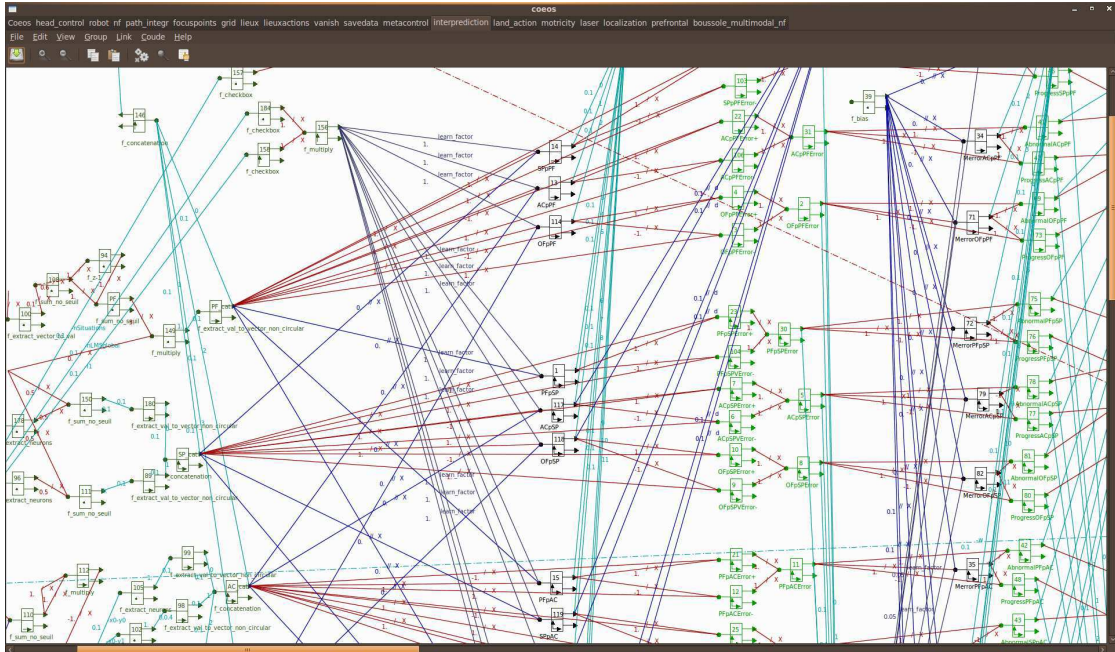


FIGURE B.1 – Capture d'écran de l'environnement graphique de conception d'architecture neuronale : Coeos. Une fenêtre d'interactions offre la possibilité d'éditer de nouveaux groupes neuronaux ou algorithmiques et de les interconnecter par le biais de liens de différentes sortes.

B.2 Promethe : un simulateur temps-réel et distribué de réseaux de neurones

Promethe est le simulateur de réseaux de neurones artificiels permettant l'exécution des scripts conçus puis compilés par Coeos. L'interface de debug de Promethe permet d'afficher une représentation de l'évolution de chaque groupe de neurones en temps réel, les ressources utilisées par tel ou tel groupe, ainsi que d'éventuelles informations complémentaires (images provenant d'une caméra embarquée, graphe temps-réel de l'évolution de l'activité d'un neurone). L'interface de debug est utilisée pour le contrôle du bon fonctionnement du système en renseignant l'utilisateur sur l'évolution des variables internes et des poids synaptiques du système. Chaque carré noir représente un groupe neuronal du script (voir fig. B.2). A l'intérieur de celui-ci, l'état des neurones du groupe est affiché sous la forme d'un carré blanc, dont la dimension est proportionnelle à l'activité du neurone correspondant. Un carré rouge dénote une activité "négative". Promethe peut également être lancé « en aveugle », c'est à dire sans aucun outil graphique de debug afin d'améliorer sa vitesse d'exécution.

Un séquenceur s'occupe de l'ordonnancement et de l'optimisation de l'exécution des groupes en fonction de la vitesse de simulation. L'exécution des groupes est réalisée suivant un système de propagation de jetons, à l'instar d'un réseau de Petri. Un groupe n'est exécuté que lorsqu'il reçoit un jeton de la part des groupes connectés en amont. Un fois son exécution terminée, le groupe transmet un jeton aux groupes connectés en aval. La mise à jour du réseau est donc réalisée par vagues de propagation des jetons. Un tel système de jetons permet de mettre en place des aspects d'exécution temps-réel si les jetons sont envoyés à une fréquence définie.

Les groupes peuvent également être exécutés à des vitesses différentes par l'utilisation d'échelles de temps fonctionnant comme des boucles imbriquées.

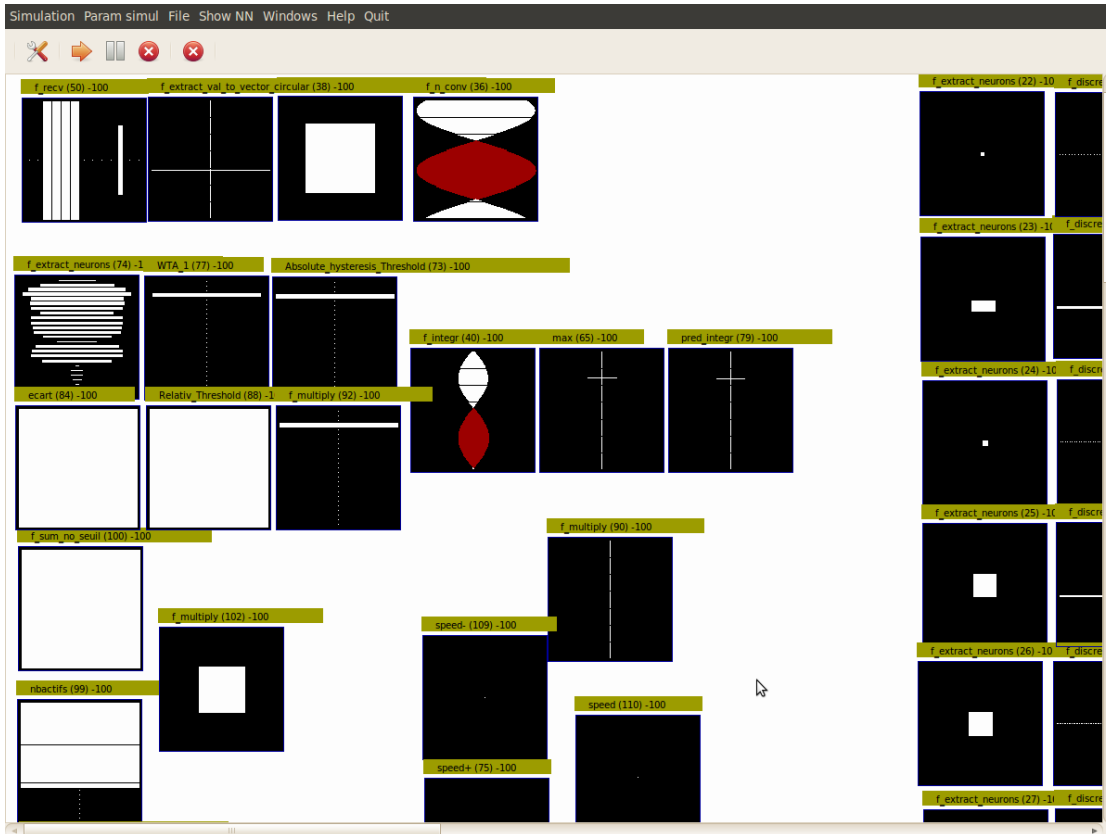


FIGURE B.2 – Interface de debug graphique du simulateur de réseau de neurones artificiels : Prometheus. L'activité des neurones de différents groupes est visible. Chaque neurone est représenté par un rectangle blanc (positif) ou rouge (négatif) dont la taille est proportionnelle à son activité. Les groupes peuvent contenir une matrice ou un vecteur de neurones ou encore un seul neurone.

Plates-formes robotiques

C.1 Plate-forme d'intérieur

La plate-forme Robulab de Robosoft est équipée comme suit :

- 2 roues motorisées de part et d'autre de la base mobile. Chaque roue possède un codeur incrémental donnant un retour odométrique des déplacements du robot.
- 9 capteurs ultra-son de proximité disposés sur le pourtour de la base mobile. Ils permettent d'estimer la distance aux obstacles sur une plage allant de 1cm à 2.5m. Chaque capteurs ultra-son possède une ouverture angulaire de 30°.
- 1 cou artificiel (un simple joystick disposé à l'avant du robot) auquel est attaché une laisse permettant à l'utilisateur d'interagir avec le robot en le tirant dans la direction souhaitée.
- 1 routeur wifi utilisé pour la communication sans fil avec un poste fixe. Cette communication sans-fil permet le déport de certains calculs si nécessaire ainsi qu'un poste fixe de supervision.
- 1 pc embarqué lui permettant de réaliser une partie des calculs de manière autonome. (processeur Intel Core I7 4 coeurs, 3,40 GH; mémoire RAM 2x8 Go DDR3, disque de stockage SSD 80 Go).
- 1 écran prodiguant un retour d'information visuel (Des expressions faciales caricaturales peuvent être représentées par le biais d'émoticônes pour la représentation de l'état émotionnel interne du système.)
- 2 haut-parleurs prodiguant un retour d'information sonore (Des phrases pré-enregistrées sont émises par un synthétiseur vocal et déclenchées par l'activation de certains neurones).
- 1 caméra FireWire, montée sur un servomoteur pan-tilt lui permet une prise d'image sur un panorama à 360°.
- 1 boussole magnétique renseigne le robot sur la direction du nord à tout moment.
- 1 système de localisation StarGazer Hagisonic. Ce système permet de localiser précisément le robot dans une pièce par le biais de repères passifs fixés au plafond. Attention, ce système de localisation n'est accessible qu'à l'utilisateur seulement. Il permet d'enregistrer précisément la position du robot, au cours d'une expérience, dans le but de pouvoir effectuer un tracé de trajectoires à postériori. En aucun cas il n'est utilisé par le robot pour se localiser.

C.2 Plate-forme d'extérieur

La plate-forme Roburoc de Robosoft est cette fois équipée pour faire face aux environnements extérieurs :

- 4 roues motrices, chacune possédant un codeur incrémental comme retour odométrique.
- 1 revêtement par film plastique ainsi qu'un parapluie afin de rendre le robot résistant à l'humidité.
- 1 sac à dos dans lequel se trouve une bâche étanche utile en cas de pluie passagère.
- 1 cou artificiel auquel est attaché une laisse pour guider le robot.
- 1 télémètre laser SICK pour l'évitement d'obstacle à 270° (Portée : 0m à 80m).
- 1 routeur wifi utilisé pour la communication sans fil.
- 1 pc embarqué.
- 1 pc portable permettant un contrôle du robot.
- 2 haut-parleurs pour le retour d'information par synthèse vocale.
- 1 caméra FireWire, montée sur servomoteur panoramique.
- 1 plate-forme stabilisatrice soutenant la caméra. Celle-ci s'aide d'un accéléromètre pour maintenir la caméra à l'horizontale lors du passage de dénivelés extérieurs.
- 1 boussole magnétique.
- 1 système de géo-localisation en extérieur (GPS). Comme pour le système de positionnement d'intérieur StarGazer, ce gps n'est accessible qu'à l'utilisateur seulement. Il permet d'enregistrer précisément les trajectoires du robot au cours des expériences en extérieur. Il n'est jamais utilisé par le robot pour se localiser.

C.3 Communication réseau

La masse de calcul requise par le robot est bien souvent trop importante pour être exécutée en temps réel sur la seule machine embarquée au sein de celui-ci. Bien souvent, il est nécessaire de déporter une part des calculs sur d'autres machines et donc de recourir à une architecture distribuée. Une telle architecture nécessite la mise en place d'un réseau sans fil (wifi) pour assurer la communication entre la partie mobile réactive (le robot) et la station de calcul, de sauvegarde et d'apprentissage (un ordinateur classique) (voir fig. C.1).

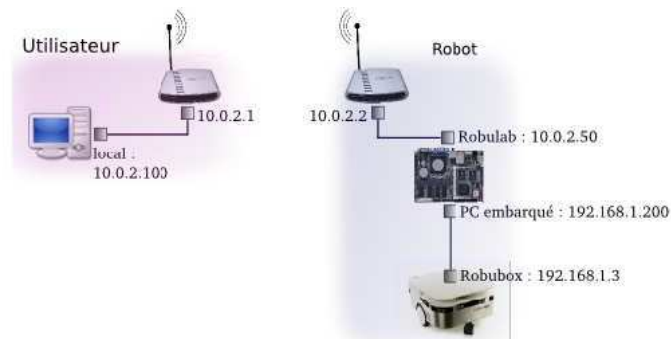


FIGURE C.1 – Schéma de fonctionnement réseau de l'architecture répartie entre machine locale et machine embarquée.

Ces dernières années, la puissance des cartes mères embarquées (processeur i7) nous permet même de ne plus dépendre d'une station de calcul déportée, puisque la totalité des calculs requis peut être exécuter par le robot en temps réel. Néanmoins, une connexion wifi avec un poste déporté est souvent nécessaire pour le contrôle et le debug distant. Chaque carte mère est connecté à un routeur wifi. Les 2 routeurs communiquent directement (connexion de type "bridge"). Enfin, le pc embarqué envoi des ordres moteurs haut-niveau à la base motrice (Robubox) du robot, laquelle se charge de contrôler les moteurs en fonction.