

# Un modèle neuronal pour expliquer la reconnaissance d'un lieu ou d'une vue.

C. Joulain & P. Gaussier, ENSEA ETIS  
6 Av du Ponceau, 95014 Cergy Pontoise Cedex  
e-mail: gaussier or joulain@ensea.fr

## 1 Introduction

Dans cet article nous présentons un modèle d'architecture neuronale permettant à un robot autonome de naviguer dans un environnement ouvert en utilisant uniquement des informations visuelles et une boussole. Le modèle s'inspire des "cellules de lieux" observées dans l'hippocampe du rat (une structure sous-corticale). Ces cellules semblent avoir une activation maximale lorsque le rat est à un endroit précis[6]. En apprenant à associer quelques lieux à un mouvement notre robot est en mesure de rejoindre un objectif particulier quelque soit son point de départ dans le même environnement. Notre modèle peut aussi rendre compte de l'existence de "cellules de vues" détectées dans l'hippocampe des primates[8]. Ces cellules réagissent lorsqu'un primate perçoit dans son champ de vision une vue particulière (alors qu'aucune "cellule de lieu" n'a été observée chez le primate). Dans cet article nous résumerons d'abord le principe de l'architecture neuronale utilisée pour la navigation ainsi que la manière dont les indices visuels sont utilisés. Des résultats expérimentaux de navigation dans un environnement ouvert seront ensuite présentés. Enfin nous montrerons comment une modification très simple des données utilisées permet au même système de reconnaître des "vues" plutôt que des "lieux".

## 2 Principe de Navigation

Lorsque le robot a trouvé quelque chose dans son environnement (nourriture ou eau dans le

cas d'une simulation d'un animal) il apprend quelques positions autour de l'objectif et le mouvement à effectuer pour le rejoindre (Figure 1). Un minimum de 3 positions doit être appris pour rejoindre l'objectif.

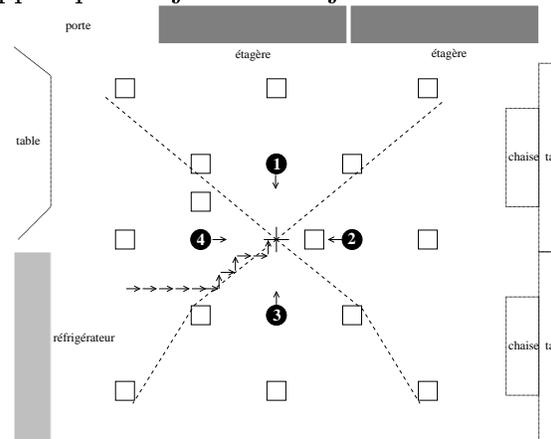


Figure 1: Quatres vues (cercles) sont apprises (i.e. associées à un mouvement élémentaire), les autres sont mises en correspondance avec une des vues apprises en fonction de leur similarité (rectangles). Si le robot apprend à rejoindre le but depuis chaque vue apprise, il est en mesure de le rejoindre depuis n'importe où. La suite de flèches représente un chemin possible. La zone représente une surface de  $1,2m \times 1,2m$ , les vues apprises sont à 30 cm du centre. L'échelle n'est pas respectée pour les positions des différents meubles (la distance au centre est de 1,5 m).

Chacun de ces endroits est appris par un neurone qui est notre équivalent de la "cellule de lieu". Le robot n'apprend pas le "but" en tant que tel. Il apprend par une approche dynamique de la perception et de l'action comment rejoindre un lieu grâce à de simples associations sensori-motrices. Quand le robot est à une position non apprise de l'environnement les "cellules de lieux" ont une activité qui est fonction de la distance entre la position

neurone le plus actif indique donc le mouvement à effectuer pour se rapprocher du but (figure 1 et 2).

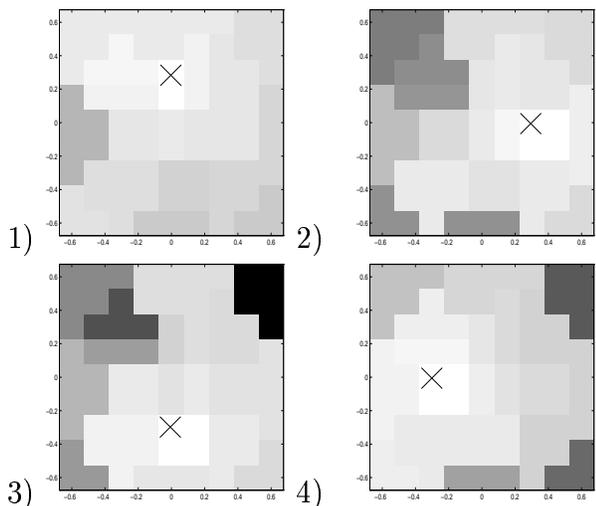


Figure 2: Réponse de chacune des “cellules de lieux” sur une surface de  $1.2m \times 1.2m$ , le niveau de gris représentant l’activité. Les croix correspondent aux position du but étant  $0,0$ .

La compétition entre ces activités crée une tessellation de l’environnement. L’architecture PerAc (Perception Action [4]) pour l’apprentissage des lieux permet donc de créer un bassin d’attraction autour du but (une approximation d’une fonction de champ de potentiel, voir figure 3).

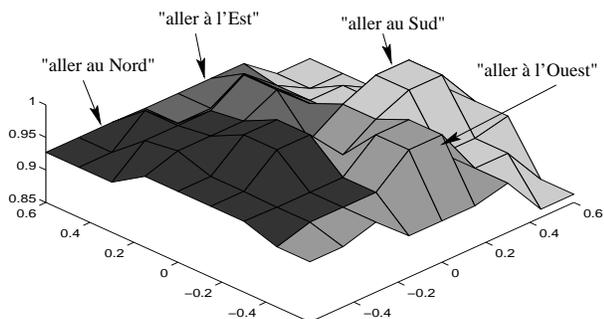


Figure 3: Activités des neurones ayant appris un lieu en fonction de la position du robot dans la pièce.

### 3 La vision pour naviguer

Notre système utilise une image panoramique ( $250^\circ$  de champ de vision voir figure 4) fabriquée en réunissant les parties centrales de

de comparer directement l’image panoramique vue à un endroit donné avec celles apprises car le changement de perspective déforme trop les images. Il faut donc trouver des points aux propriétés invariantes par changement de point de vue. Les angles sont de bons candidats. En effet un angle reste un angle par rotation, translation et changement d’échelle. Ainsi en séparant les informations relatives au “où” et au “quoi” comme le fait notre système visuel (voies temporelle et paritale) on obtient une représentation plus facilement exploitable. Pour trouver des amers potentiels dans l’image, le système fait une moyenne pondérée colonne par colonne de l’image panoramique pour obtenir une ligne. Ce signal est ensuite dérivé pour permettre de centrer la prise de vues locales sur les zones ayant un fort contraste horizontal (figure 4). C’est un mécanisme de focalisation de l’attention simple et rapide, nous travaillons aussi sur des mécanismes plus complexes[7] consistant à rechercher des zones angulaires. L’information de position d’un amer est donnée par son azimuth par rapport à un référentiel relatif ou absolu (i.e. le Nord avec une boussole électronique). On pourrait se passer de la boussole en utilisant un amer comme référentiel angulaire ou en ne travaillant que sur les différences d’angles. Chaque vue locale est réduite en effectuant un moyennage linéaire sur l’axe  $y$  et logarithmique sur l’axe  $x$  (figure 5).

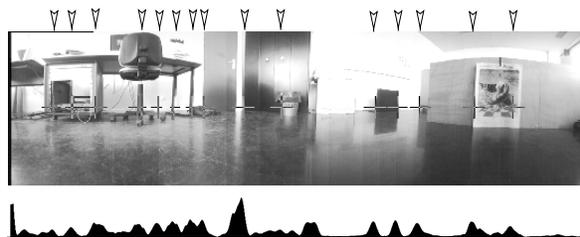


Figure 4: 15 vues locales apprises pour une position et la valeur absolue de la dérivée du contraste.

Notre modèle fusionne les informations représentant le “quoi” (groupe Amer Figure 6) avec les informations relatives à l’angle sous lequel est vu l’amer : le “où” (groupe Azimut).

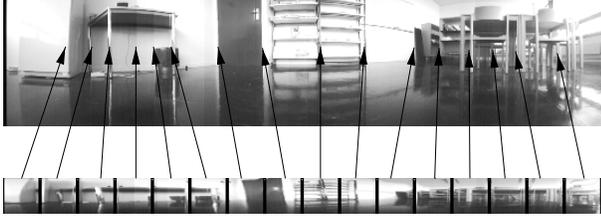


Figure 5: 16 exemples de vues locales  $32 \times 32$  utilisées comme amer et leur azimut dans l'image panoramique.

En neurobiologie, cette fusion “plastique”, par opposition à la reconnaissance statique d'une configuration multi-sensorielle, semble être effectuée par l'hippocampe [1]. Cette recombinaison (groupe Amers  $\times$  Azimuts) peut alors être apprise par des neurones appartenant à un groupe du type Winner Take All. Ces neurones simulés réagissent alors exactement comme les “cellules de lieux” enregistrées dans l'hippocampe des rats.

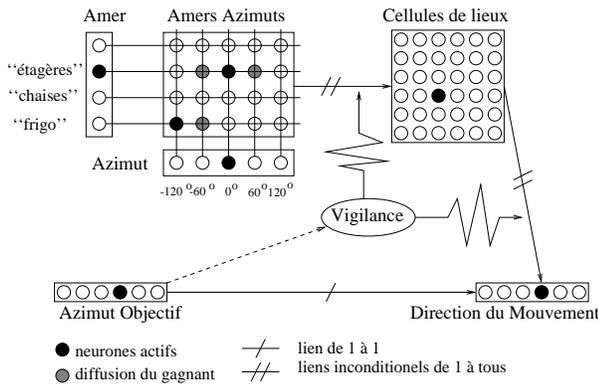


Figure 6: Réseau de neurones pour la navigation. Amers Azimuts fusionne les voies “où” et “quoi”.

Nos expériences sur le robot nous ont montrées que la reconnaissance des amers ne doit pas être trop sélective. Plusieurs vues locales apprennent au cours du temps peuvent correspondre à un même amer physique. Vouloir sélectionner forcément une seule des reconnaissances visuelles risque donc de favoriser de manière imprévisible une “cellule de lieu” particulière. Pour éviter d'avoir une activité réduite lorsqu'un amer est mal reconnu le groupe Amer n'a pas une seule sortie mais un ensemble de sorties plus ou moins actives en fonction de la reconnaissance.

Pour tester la robustesse et la capacité de généralisation de notre système nous lui avons manuellement fait apprendre une dizaine de positions dans toute une pièce pour atteindre un objectif distant. La figure 7 permet de voir qu'en ayant appris à associer 10 “lieux” de l'environnement à 10 mouvements, le robot est en mesure de rejoindre l'objectif depuis n'importe quel point dans une surface d'à peu près  $20m^2$ . Ce simple mécanisme de conditionnement crée une sorte de puit de potentiel vers le but. La figure 7 montre différents chemins empruntés par le robot en fonction de sa position et de son orientation de départ. L'introduction d'éléments perturbateurs dans le champ de vision n'empêche pas le robot d'atteindre son but.

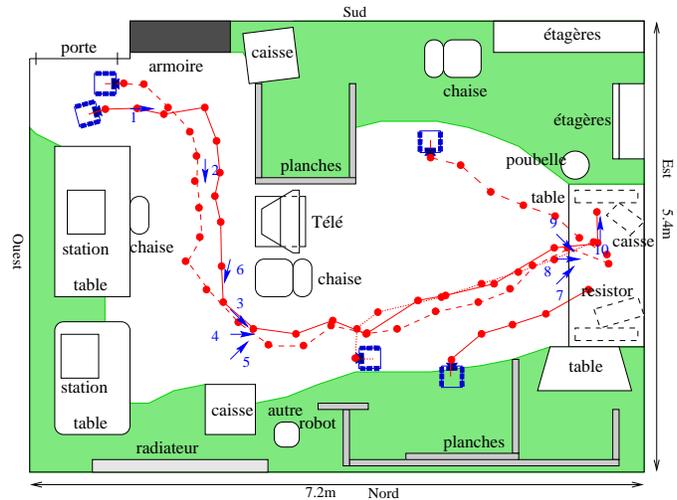


Figure 7: Les flèches numérotées représentent les positions apprises. Le but est sous la table (près de 10). Le robot ne peut rejoindre le but depuis les zones grisées. Le début d'une trajectoire et son orientation initiale sont représentés par les robots, les points reliés composent le reste de la trajectoire.

En biaisant l'activité des neurones grâce à un signal lié à une motivation la même architecture a été utilisée pour gérer plusieurs buts[5].

## 5 “Cellule de vue”

Nos expériences suggèrent qu'avec un champ visuel réduit notre robot montre le même type

statées chez les primates (figure 8). Si la “tête” est trop tournée alors il n’y a plus de relation entre les informations visuelles extraites même si le robot reste à la même place.

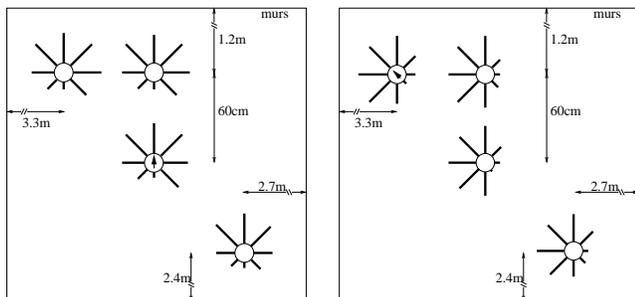


Figure 8: Lorsque le champ de vision du robot est réduit l’activité de type “cellule de lieu” devient une activité de type “cellule de vue”.

Ceci peut apporter une explication très simple au fait de ne pas avoir trouvé de “cellule de lieu” [8] chez les primates. Ainsi, pour reconnaître une position entourée d’un grand nombre d’amers, il est plus intéressant de prendre des amers tout autour de la scène. Le rat, grâce à la disposition de ses yeux, dispose d’un champ de vision très large. Par contre les primates ont un champ visuel plus réduit et les enregistrements de leurs cellules hippocampiques montre seulement des activités de type “cellule de vue”.

## 6 Conclusion

Notre algorithme est efficace même dans des situations délicates, la disparition d’amers ou l’introduction d’éléments perturbateurs [3]. Notre robot est capable d’apprendre à reconnaître des scènes relativement complexes et de naviguer correctement avec une architecture neuronale et un mécanisme d’apprentissage non supervisé très simple. Contrairement à la majorité des systèmes de navigation visuelle, ce système ne nécessite pas la présence d’un ingénieur pour identifier les vues à apprendre. Lorsque le robot sort d’une zone pour entrer dans une autre, il peut se tromper de mouvement, mais alors la somme des activités de ses “cellules de lieux” diminue très fortement. Ceci nous

de renforcement déclenchant l’apprentissage l’apprentissage d’un nouveau lieu et son association à un mouvement correct. L’aller retour entre la biologie et les simulations avec des signaux réels sur un robot à un intérêt double. Il nous permet à la fois de créer des architectures efficaces pour les robots autonomes et de mieux comprendre certains mécanismes biologiques.

## References

- [1] J.P. Banquet, P. Gaussier, J.C. Dreher, C. Joulain, and A. Revel. *Cognitive Science Perspectives on Personality and Emotion*, chapter Space-Time, Order and Hierarchy in Fronto-Hippocamal System: A Neural Basis of Personality. Elsevier Science BV Amsterdam, 1997.
- [2] P. Gaussier, C. Joulain, S. Zrehen, J.P. Banquet, and A. Revel. Navigation visuelle dans un environnement ouvert: reconnaissance de vues panoramiques. In *Actes du GRETSI’97*, pages 861–864, Grenoble, France, Septembre 1997.
- [3] P. Gaussier, C. Joulain, S. Zrehen, J.P. Banquet, and A. Revel. Visual navigation in an open environment without map. In *International Conference on Intelligent Robots and Systems - IROS’97*, pages 545–550, Grenoble, France, September 1997. IEEE/RSJ.
- [4] P. Gaussier and S. Zrehen. Perac: A neural architecture to control artificial animals. *Robotics and Autonomous System*, 16(2-4):291–320, December 1995.
- [5] S. Lepretre, P. Gaussier, and A. Revel. Utilisation d’une carte “cognitive” pour le problème de la sélection de l’action dans un cadre multi-motivationnel. In *Soumis à NSI’98*, Munster, France, Mai 1998.
- [6] J. O’Keefe. *Neural Connections, Mental Computation*, chapter Computations the hippocampus might perform, pages 225–284. MIT Press, 1989.
- [7] M. Quoy, O. Gallet, and P. Gaussier. Implémentation parallèle d’un système de focalisation de l’attention. In *Actes du GRETSI’97*, pages 1205–1208, Grenoble, France, Septembre 1997.
- [8] E.T. Rolls and S.M.O’Mara. View-responsive neurons in the primate hippocampal complex. *Hippocampus*, (5):409–424, 1995.